Winter 12-15-2021

Cancer Epigenome Reprogramming

Jennifer Ann Karlow
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Division of Biology and Biomedical Sciences
Computational and Systems Biology

Dissertation Examination Committee:
Ting Wang, Chair
Barak Cohen
Ramaswamy Govindan
Robi Mitra
Nancy Saccone
Mark Watson

Cancer Epigenome Reprogramming

by
Jennifer Karlow

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2021
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

The work put forth in this thesis is a product of the loving support of numerous family, friends, collogues, and mentors. I must start out by first extending my deepest appreciation for my thesis mentor, Ting Wang. Growing up in St. Louis, it had always been my dream to attend Washington University, so when I received the phone call of my acceptance to the Computational and Systems Biology program, you can imagine my excitement. Little did I know at the time, that call came from my future thesis advisor! Ting had also been one of my interviewers, and the work he presented during the interview lead me to pursue a rotation in his lab. His enthusiasm for cancer research, hypothesis-driven projects, and support for his students to explore both wet and dry lab techniques made the decision to join his lab a relatively easy one on my end. The only question was whether or not he was willing to take a fourth graduate student in the same year; luckily for me, he did! Over the years, Ting has supported me in countless ways, through helping me formulate new project ideas, assisting me in applying to fellowships and grants, coaching me through paper revisions, sponsoring my dive into the wet lab, and pushing for Friday lab volleyball. He has been there to guide me through the tough reality of paper rejections and failed experiments, and to celebrate milestones, most notably by throwing a lab party when I was engaged. I am forever grateful for all that he has done to help shape me into the scientist I am today. My sincerest thank you, Ting!

In a similar way, I'd like to thank my thesis committee, Barak Cohen, Rob Mitra, Govindan Ramaswamy, Nancy Saccone, and Mark Watson. Each member of my committee has gone out of there way and devoted countless hours to helping me shape my projects. As my thesis chair, Barak has always offered support and valuable insights regarding any questions I

bring his way. In the initial stages of project formation, Barak, Rob, and Nancy all met individually with me several times to hash out various aspects of the projects. As my thesis work began to focus further on the metastasis work, Govindan and Mark both provided valuable feedback, and continue to meet regularly with me as we wrap up our manuscripts.

While I am extremely grateful to Ting for supporting me in my wet lab endeavors, none of the work would have been possible without the continued mentorship from Josh Jang and Xiaoyun Xing. Both were patient with me while I learned techniques, supported me when my experiments failed, and helped me troubleshoot my failures in order to improve. It is because of all three of them that I was able to make experimental work a component of my thesis. Similarly, many of the project that I have been involved with over the years in the lab have been a joint effort involving many other talented graduate students, staffed scientists, and faculty. This sense of comradery was felt most while working on the WashU Virus Genome Browser. One day in lab meeting, Ting proposed the idea of extending our current WashU Epigenome Browser infrastructure to accommodate the ever-growing number of virus strains made publicly available, and 10 days later, our browser team (Mayank Choudhary, Changxu Fan, Gavriel Matt, Daofeng Li, Deepak Purushotham, Xiaoyu Zhuo, and Ting Wang) had completed the first iteration. From day one of this project to continued browser updates, the work presented in Chapter 6 would not have been possible without any member of this fantastic team!

In addition, I must thank the entirety of the Wang Lab for their continued support throughout the years. Because of this group of people, the lab has truly been my home-away-from-home. Each member of the lab has made a positive impact on science presented here, through thoughtful discussions, mentoring, and simply being there to play ping pong when hitting mental roadblocks. Over the years, in addition to time spent in the lab, annual float trips,

lab parties, Friday basketball (and occasionally volleyball), and happy hours have allowed me to really get to know many lab members and form long-lasting friendships.

The graduate school journey would not have been possible without the constant support of my partners-in-crime, Liron Ganel, Nick Jacobs, Rachel Lyman, and Dennis Wu. I consider myself extremely lucky to have entered into graduate school with this cohort. Whether it be surviving an unprepared float trip, sleepless nights working on machine learning assignments, game nights, or morning coffee, these friends have always been there when I've needed them the most. Our first year in graduate school, we took a picture, which I've had in a frame with a caption "I get by with a little help from my friends". The Beatles said it best.

During orientation week, Jim Skeath gave wonderful advice that stuck with me throughout graduate school. He emphasized that while you will undoubtably be busy in the lab, many other life milestones will occur during graduate school, some exciting, some difficult, necessitating a strong support group. In addition to my friends, words cannot express how indebted I am to my family for their unconditional love and support throughout my life, especially during my PhD work. I have always been close with my family, so when it came time to select a PhD program, it was the icing on the cake to be in St. Louis. Not only has my family been there for me in the form of phone calls and zoom chats, living in the same city has allowed for Catan games and water volleyball with brothers, Starbucks runs with mom, and volleyball and running with dad. Mom, dad, Eddie, Andrew, Jessie, and Snoopy, from the bottom of my heart, I sincerely appreciate all that you have done for me.

Lastly, I would not have gotten through these past several years without my best friend by my side, my husband, Nick. Although we met when we were in middle school, Nick and I did not start dating until college. Even so, by the time graduate school started we had been together

for a couple of years, so Nick has been there through it all! In graduate school, life can sometimes be a rollercoaster of emotions – you're excited about your science, afraid you might fail, sad you haven't slept, happy your paper was accepted, ect. Through every phase, Nick has not only celebrated and cried with me, he's also supported me and pushed me to be a better scientist and person because of it. Nick and I became engaged in the summer of 2019, and even though I had always envisioned a large wedding, the onset of the pandemic complicated our plans. When it came down to whether we would have a small wedding (less than 15 people) in late 2020 or wait indefinitely for the larger gathering, there was no contest. After being married now for over one year, I can say that I am so incredibly thankful for him every day and the life we have built together (which includes adopting our loving puppy, Leo, for whom I am also extremely grateful). Nick's continuous support has had a tremendous impact on my life and has shaped me into the person I am today, scientist and all.

Jennifer Karlow

Washington University in St. Louis

December 2021

Dedicated to my wonderful family.

ABSTRACT OF THE DISSERTATION

Cancer Epigenome Reprogramming

for Arts & Sciences Graduate Students

by

Jennifer Karlow

Doctor of Philosophy in Biology and Biomedical Sciences

Computational and Systems Biology

Washington University in St. Louis, 2021

Professor Ting Wang, Chair

The identification of recurrent genetic mutations in cancer and their functional characterization has provided a strong foundation for our understanding of tumorigenesis. The more recent observation of recurrent and specific epigenetic changes also present in cancer has widened this view, now establishing cancer as a disease of both genetic and epigenetic misregulation. Enhancers, genomic regions primarily responsible for tissue-specific gene expression, have been shown to be frequent targets of both genetic and epigenetic abnormalities. The observation that DNA methylation within regulatory regions has traditionally correlated with reduced gene expression, coupled with the known role of enhancers in regulating tissue-specific gene expression, suggests that cancer cells may lose their original cellular programming while also gaining expression for genes related to foreign cell types through altered DNA methylation within enhancers. Technological advancements allowing for genome-wide methylation profiling have vastly expanded our capacity to identify DNA methylation alterations present in enhancers, previously undetected in large comparative cancer studies. In Chapter 2, we compared global methylation alterations in two distinct cancer types, endometrioid

adenocarcinoma (EAC) and glioblastoma multiforme (GBM). We found that both cancer types displayed an increase in methylation over enhancers related to their cell type of origin. Recurrently hypermethylated enhancers in EAC clustered based on the presence of enriched transcription factor binding motifs. Enriched disease ontology pertaining to the majority of differentially methylated enhancer groups largely encompassed uterine-specific terms, suggesting that enhancers gaining methylation in EAC might contribute to the regulation of normal uterine function. Similarly, GBM hyperDMRs were shown to encompass a significantly greater number enhancers active in adult brain compared to those of developing brain. Further analysis of brain enhancer hyperDMRs revealed increased H3K27ac and H3K4me1 signal in adult brain compared to fetal brain, suggesting that the majority of enhancers being methylation were active in adult but not developing brain.

In a second series of studies (Chapters 3 – 5), we sought to better understand the properties governing metastasis organotropism, specifically non-small cell lung cancer metastasis to brain. Reasoning that genomic, transcriptomic, and epigenomic changes might functionally contribute similar alterations, we profiled all three in a subset of 45 paired primary and metastasis samples. While few significantly mutated genes were private to metastasis, 75 genes displayed recurrent metastasis enriched variants (MEVs), largely implicated in focal adhesion and extracellular matrix receptor interactions. A similar analysis revealed increased variant allele frequencies (VAFs) in metastases over a wide range of epigenetic regulators, suggesting that epigenetic misregulation may be selected for, and possibly functionally contribute to, NSCLC metastasis to brain. Consistent with these observations, we observed widespread changes in DNA methylation as a function of disease progression, many losses of which were found within annotated brain-specific active enhancers and correlated with increased

nearby gene expression. Transcriptional analyses corroborated these findings, revealing an overall increase in brain gene expression. Together these findings suggest that a potential contributor to NSCLC's successful colonization of brain is the aberrant activation of brain transcriptional programs.

To further understand how methylation changes might be contributing to metastasis, we examined DNA methylation differences across patients and found that the most recurrent and largest changes between primary and matched metastatic samples were within DNA methylation valleys (DMVs), large regions devoid of methylation in normal lung that primarily contain developmental genes. Further examination revealed that only a subset of DMVs, marked by high H3K9me3, H3K4me1, and H3K27me3 signal in normal lung, were undergoing progressive methylation gain. The known mutual exclusivity of H3K27me3 and DNA methylation within CpG islands prompted us to ask whether there was a loss of polycomb repressive complex 2 (PRC2), a multi-protein complex responsible for depositing H3K27me3, occupancy within DMVs accompanying the increased DNA methylation. The binding patterns of the PRC2 catalytic subunit, EZH2, in the non-small cell lung cancer lymph node-derived cell line, H1299, displayed a pervasive loss of EZH2 occupancy within DMVs, accompanying a large increase in DNA methylation. The mechanisms governing this epigenetic switch and its implications in metastasis are the topic of ongoing research, but may lead to the activation of encompassed developmental genes, shifting the tumor cells to a more stem-like state.

Chapter 1: Introduction

1.1 Cancer: a disease of both genetic and epigenetic alterations

The effects of cancer on human life have been felt for centuries, with earliest documented descriptions dating back to 3000 BC[1], and the first recorded, observed associations in the 1700s[2]. Technological advancement has vastly increased our understanding of the fundamental molecular changes underpinning tumorigenesis; however, treatment options remain limited and vary according to tumor origin as well as stage. Traditionally, cancer treatments have been dictated according to location of origin; however recent advances in personalized medicine have indicated that survivability is improved when treating a cancer according to molecular profiling, regardless of location, exemplifying the genetic underpinnings of cancer.

1.1.1 Genomic drivers of cancer

In the early 1900s, Peyton Rous observed that injecting chicks with the serum of tumors from closely related chickens could lead to the formation of tumors, later determined to be due to the transmission of a virus (*Rous sarcoma virus* (RSV))[3,4]. Later, in the 1970s, high similarity to a gene endogenous to many animals, c-*src*, provided some of the first evidence that cancer could be caused by genes within one's own genome, thus initiating the hunt for additional oncogenes[5]. Around the same time, Alfred Knudson observed that the frequency with which patients developed retinoblastoma was consistent with the accumulation of two mutations[6], soon thereafter discovered to be mutations within the *RB* gene, the first identified tumor suppressor.

Since the identification of the first oncogenes and tumor suppressor genes, a compendium of studies have illuminated the enormous degree to which genetic mutations impact cancer. In

2001, the compellation of a human reference genome provided a context in which to compare genetic alterations to a non-diseased, normal profile[7]. Shortly thereafter, large sequencing studies were conducted to identify genetic mutations in tumors across several cancer types, generating many testable hypotheses[8–44]. Integrative studies comparing the frequency of genomic mutations within and across cancer types have revealed recurrent pan-cancer genetic mutation signatures encompassing hundreds of genes, most prominently including genes involved in DNA mismatch repair (*MSH*, *PRKDC*, and *TP53*), cell cycle, growth, proliferation, and survival (*ATM, BRAF, CASP8, CDKN2A, ERBB3, HRAS, KRAS, NF1, NRAS, PIK3CA, PIK3R1, PTEN, RB1*, and *STAG2*), protection from oxidative stress (*NFE2L2*), protein degradation (*FBXW7*), immune response (*HLA-A*), and chromatin maintenance (*MLL2, MLL3, MLL4, ARID1A, CTCF, IDH1*)[45–47]. In addition, recurrent chromosomal aneuploidy[48], copy number alterations[49], and structural variations[50] have been identified. Accompanying analyses also revealed considerable cancer type-specific enriched mutations, highlighting the unique susceptibility of different cell types to genetic perturbations[51].

To date, functional characterization of recurrent genomic mutations continue to be widespread[52–55] and contribute to altered treatments[56–59]. In addition, continued work on large-scale studies to improve statistical power to detect potentially important, functional mutations contributing to pan-cancer and type-specific signatures, drug resistance, and metastatic dissemination, are the priority of many groups and consortia, such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC). Improved statistical power to detect genetic mutations that contribute to tumorigenesis at lower population frequencies or in a redundant fashion will be essential to providing further candidate mutations for functional validation.

1.1.2 Epigenetic alterations in cancer

An unexpected observation at the time of the first expansive cancer sequencing projects was the identification of recurrent genetic mutations within chromatin modulators, present in roughly half of all tumors[60], suggesting that consequential downstream epigenetic perturbations could have a selective advantage in cancer. Prior to this observation, the first discovery of epigenetic changes correlating with cancer was made by Andy Feinberg and Bert Vogelstein in their 1983 publication comparing DNA methylation in cancer and normal tissues, revealing substantial loss of methylation over specific genomic regions in cancer samples[61].

1.1.3 Role of DNA methylation

In addition to genetic modifications, altered profiles of normally occurring epigenetic states have also been identified and appear to be prevalent in cancer. One such epigenetic marker is DNA methylation, the covalent bond of a methyl group to the 5th position of a cytosine. DNA methylation is a broadly occurring modification, observed in the genomes of many different prokaryotes and eukaryotes. In mammalian genomes, DNA methylation is typically present in a CpG (cytosine adjacent to guanine) context[62]; however increased levels of CpH (non-CpG methylation contexts) have been detected in the brain as well as in developing tissue[63]. The modification is conferred by DNA methyltransferases, two of which are responsible for the *de novo* addition of methylation (*DNMT3A* and *DNMT3B*)[64], and another is required for proper maintenance of DNA methylation through replication (*DNMT1*)[65]. Mutations in *DNMT1* and *DNMT3B* have demonstrated embryonic lethality in mice, and mutations in *DNMT3A* have resulted in severe developmental abnormalities and shortened lifespans, implying an important functional role of DNA methylation in development[64,66]. DNA methylation can also be enzymatically removed by a family of three DNA demethylases known as ten-eleven

translocation (TET) proteins, *TET1*, *TET2*, and *TET3*, through oxidation[67], whose combined loss also impedes embryonic differentiation[68].

As embryonic knock-out studies of DNA methylation writers and erasers have implied, DNA methylation serves multiple functions in non-diseased tissue. One such purpose is silencing transposable elements, repetitive sequences with potential to move about the genome. Indeed, loss of *Dnmt* in mouse embryos leads to an abundant increase in retroviral transcription[69]. In a related manner, DNA methylation also functions within gene bodies to prevent non-canonical transcription initiation[70]. Additionally, chromosomal X-inactivation, necessary for proper dosage compensation between males and females, is maintained at least in part through DNA methylation[71], where the addition of a DNA demethylation drug (5-azacytidine) has been show to activate previously inactive X-chromosome genes[72]. DNA methylation has also demonstrated widespread reprogramming in the germline and preimplantation embryos in the form of global methylation reduction[62]. Roughly 20% of CpGs are protected from this loss of methylation, likely though KRAB domain-containing proteins, maintaining their parental methylation status as "imprinting control regions"[62] which results in transcription initiating from a single parent allele. At the blastocyst stage, a new wave of DNA methylation is initiated via *DNMT3A* and *DNMT3B*. Lineage restricting methylation patterns, largely within tissue-specific enhancers, are gradually accumulated and have been shown to be necessary for cellular differentiation[73].

DNA methylation within promoters has generally been correlated with decreased gene expression[74], whereas methylation within gene bodies shows a positive correlation with expression[75]. While this observation has been longstanding, questions still remain surrounding the causal role DNA methylation has in regulating specific gene expression, owing in large part to its complex relationships with other epigenetic marks, environmental exposures, and

transcription factors[76]. For example, DNA methylation has been shown to provide steric hinderance precluding methylation-sensitive transcription factor binding, and can recruit a slew of proteins with methylation-binding domains to further repress transcription through additional steric hindrance as well as through deposition of repressive histone modifications[76]. On the other hand, other "pioneering" transcription factors have the ability to bind to condensed chromatin, modify the surrounding DNA methylation concentration, and recruit additional transcription factors[76]. In addition, transcription factor binding at enhancer regions has demonstrated the ability to protect the genomic region from DNA methylation accumulation in certain cases[77]. Nevertheless, functional roles for DNA methylation-mediated transcriptional silencing have been identified, many of which in the context of cancer.

DNA methylation profiling in tumor samples has revealed recurrent and specific changes. Globally, cancer samples exhibit a reduction in DNA methylation, with focal points of increased methylation[78]. Loss of genome-wide methylation has been linked with several abnormalities that are likely to propagate the malignant phenotype. Removal of methylation within gene bodies or over transposable elements can expose sequences with regulatory potential, which have been linked to spurious transcriptional initiation of oncogenes[79]. In addition, global methylation loss has been linked to genomic instability through the rearrangement of large pericentromeric regions[80]. Methylation loss within specific oncogenes and testis-specific antigens has also been shown to correlate with increased expression and disease severity, and recent advances in technology are now allowing for the direct causal mechanisms for this relationship to be elucidated[81,82]. Gain of methylation in cancer has been identified in concentrated promoter regions of tumor suppressor genes, and several studies have probed DNA methylation's contribution to the repression of its target gene's expression. Early attempts to determine DNA

methylation's impact on tumor suppressor gene expression explored the effect of introducing a DNA de-methylating agent (usually 5-aza-2'-deoxycytidine, 5-deoxyazacytidine, or 5-azacytidine) in cultured cells mimicking the hypermethylation status observed in cancer, on methylation status of the region in question as well as altered expression. Several landmark observations of restored expression upon demethylation include *p16*[83], *VHL*[84], *E-cadherin*[85], *hMLH1*[86], and *LKB1*[87,88]. While these studies show promising evidence that DNA methylation helps mediate silencing within these specific promoters, the non-specific nature of these demethylating agents makes causal inference not entirely clear. Therefore, targeted methylation perturbations and resulting expression changes, made possible through recent advances in technology[81,82,89–91], will further solidify DNA methylation's causal role in epigenetic silencing. Finally, recurrent methylation changes are frequently identified in enhancers across different cancers, correlating with altered transcriptional profiles[92], and have demonstrated causal effects when probed via targeted DNA methylation alterations[93].

1.1.4 Role of histone modifications

Like DNA methylation, various post translational modifications to histone protein side chains and tails are typical in normal tissues, and help maintain the degree of chromatin rigidity. Broadly speaking, histone modifications are usually associated with either transcriptional activation or repression. Promoters of actively transcribed genes usually contain active histone 3 modifications in the form of tri-methylated lysine residue 4 (H3K4me3) and acylated lysine residue 27 (H3K27ac). Active enhancers also contain H3K27ac, but are distinct from promoters in containing mono-(as opposed to tri-) methylated histone 3 lysine residue 4 (H3K4me1). Repressive histone modifications H3K9me3 and H3K27me3 largely characterize transcriptionally silent heterochromatic regions, whereas the repressive mark H3K36me3

maintains transcriptional repression of gene bodies. Biochemically speaking, acylation has been shown to contribute to the openness of chromatin due to its elimination of the nucleosome's positive charge, lessening the histone's attraction to the negatively charged DNA phosphate backbone[94]. Mechanisms explaining the contribution of histone methylation to chromatin conformation are more nuanced since some methylation is associated with activation and some with repression, but are likely intimately linked to the different binding specificities of different histone modification "readers" (see below)[95].

Similar to DNA methylation, various proteins are responsible for histone modification deposition ("writers"), interactions ("readers"), and removal ("erasers"). While many different histone methyltransferases have been identified[96], their enzymatic activity is typically conferred by a SET domain[97]. Histone methylation removal, on the other hand, can be catalyzed via one of two histone demethylase families, LSD demethylases and JMJC demethylases, which function to remove methylation via oxidation and dioxygenase reactions, respectively[98]. Chromodomain-containing proteins, as well as a variety of others, are able to interact with methylated histone residues and induce chromatin conformation changes either in favor of or opposing transcription (more on this in 1.1.2.3), depending on the genomic context[99]. In a similar manner, histone acylation is conferred by acyltransferases, removed by deacetylases, and provides binding sites for a specific group of bromodomain-containing proteins (and more recently discovered YEATS domain-containing proteins[100]). Bromodomain-containing proteins can then exert a variety of functions, including direct transcriptional initiation, recruitment of transcription factors, or directly or indirectly act to modify the surrounding epigenome[101].

In cancer, several recurrent alterations have been identified in histone modifying enzymes. Two of the most pervasive abnormalities observed are within the histone "writers"

EZH2 and MLL, responsible for the deposition of H3K27me3 and H3K4me, respectively. *EZH2* has been shown to be frequently overexpressed or display activating mutations across many different cancer types[102], whereas *MLL* is often duplicated in blood cancers or fused with one of several target genes[102,103]. Similarly, translocations resulting in fusion proteins containing histone "readers" have also been identified and found to disrupt the normal progression of histone modification patterning[102,104]. Finally, an increased mutational frequency within the H3K27-specific demethylase gene, *UTX*, has been identified in several cancers[105], as well as overexpression of histone deacetylases[106], demonstrating selection of epigenetic alterations correlated with aberrant silencing and activation in cancer.

1.1.5 Role of polycomb repressive complexes

Originally identified in drosophila, polycomb complexes consist of several protein members which broadly assemble into one of two complexes, Polycomb Repressive Complex 1 (PRC1) or Polycomb Repressive Complex 2 (PRC2). Both PRC1 and PRC2 contain a set of core proteins whose presence is vital to the enzymatic activity of the complex, and can also contain a subset of auxiliary proteins, the composition of which results in canonical and non-canonical complex assemblies[107].

Core proteins of the PRC1 complex consist of RING1A and RING1B (one of which is necessary for PRC1 formation) whose membership is responsible for carrying out the complex's catalytic activity as an E3 ubiquitin ligase[107] which results in the addition of a single ubiquitin to Lysines 118 and 119 of histone 2A. In addition to genomic sequence[108], components of the PRC1 complex largely dictate how it is targeted to different genomic regions. Canonical PRC1 has shown binding affinity for H3K27me3 (deposited by PRC2), whereas other PRC1 complex

forms are targeted to unmethylated DNA[109], specific regions guided by lncRNAs or transcription factor interactions[108], or may bind more spuriously throughout the genome[107].

Core complex components belonging to PRC2 include EED, SUZ12, RBBP4, RBBP7, EZH1, and EZH2[107], of which, EZH1 or EZH2 (only one of which is required for PRC2 formation) carries out the enzymatic activity of depositing methylation on H3K27 (H3K27me1, H3K27me2, and H3K27me3)[107]. Like some PRC1 complexes, one of the PRC2 complexes is also recruited to unmethylated DNA[108], and has been hypothesized to be recruited to specific locations via transcription factor interactions. Specific PRC2 complexes are also recruited to H2AK119ub1 (deposited by PRC1), circularizing the connection between PRC1 and PRC2 and providing a possible explanation for the concordant accumulation of both H3K27me3 and H2AK119ub1 in large domains[108]. In addition, both PRC1 and PRC2 are able to bind to their respective histone modification deposits, resulting in a positive feedback loop of polycomb repression[108].

The catalytic activity (as well as non-enzymatic properties) of PRC1 and PRC2 function by silencing their target genes, most of which are implicated in development[110]. It should be noted that polycomb complexes are balanced by the presence of trithorax protein complexes, also originally discovered in drosophila as mediators of chromatin changes activating transcription[111]. Collectively, polycomb group and trithroax proteins demonstrate recurrent abnormalities in cancer[112]. As noted above, *EZH2* is frequently overexpressed and exhibits gain of function mutations broadly across cancer types[113]. Increased EZH2 activity has been shown to correlate with, and directly induce, silencing of a number of tumor suppressor genes[114]. This observation has provided motivation to determine if EZH2 inhibitors might provide a therapeutic benefit[114]. However, additional mutations have been identified in H3K27 as well as in the PRC2 agonist

(*CATACOMB*), both resulting in a widespread loss of H3K27me2/3[107], suggesting that polycomb may function as a tumor suppressor in certain cases. In light of this current paradox, it should be noted that a full mechanistic understanding of polycomb involvement in cancer is far from complete, and the topic of much continued research (including mine!).

1.2 Metastatic disease

Although many advances have been made in cancer treatments over the past century, roughly 90% of all cancer-related deaths are the result of metastatic dissemination[115], necessitating further understanding of malignant progression and intervention. Metastasis occurs when primary tumor cells acquires the potential to invade and survive in a foreign location. For a primary tumor cell to successfully colonize a distant site in the body, it must undergo several steps, beginning with detachment from the primary tumor and surrounding tissue and traversing the extracellular matrix[116,117]. Once the invading cell contacts a blood vessel, it must intravasate into the blood stream where it must avoid attack by natural killer cells and macrophages, shearing under the stress of blood flow, and must survive without access to its original microenvironment[116,117]. Although millions of tumor cells are predicted to disseminate into the bloodstream, very few of them can survive under these harsh conditions and consequently do not reach a secondary location. Of the cells that do manage to seed a distant site, the vast majority are not equipped to survive in the new tissue's microenvironment and therefore do not end up forming lesions[116,117]. The series of obstacles a tumor cell must face ensures a selective bottleneck such that the cells successful in seeding the metastasis should be enriched for functional properties that allowed the cell to survive this journey as compared to their primary cell relatives.

Research comparing primary tumors with different known outcomes (eventual growth of metastasis or confined to local growth)[118], and primary tumors with distant metastases[119] has shed some light onto several key players and pathways that may mediate metastasis formation. In addition to maintaining several aspects of primary tumor formation, such as increased proliferation and evasion of immune and death signals, the functional gain or loss of additional genes, either resulting from expression changes or genetic mutations, can help propagate the metastatic phenotype without affecting primary tumor growth[116,117,120]. Genes whose loss of function contributes to this process separately from primary tumor formation are collectively called metastasis suppressor genes[120,121]. Several of these genes have been identified in various cancer types, including *CDH1* which is responsible for maintaining cell-to-cell junctions[116], and tissue inhibitors of metalloproteinase (TIMPs) which repress matrix metalloproteinases from degrading the extracellular matrix[122]. However, expression of many metastasis-related genes have been found to correlate with varying physiologies depending on the cancer type; for example, reduced expression of *NM23* is associated with poor prognosis in several cancers including melanoma, breast, colon, and gastric cancers, but high levels correlate with poor prognosis in thyroid cancers, neuroblastomas, and osteosarcomas.

1.2.1 Organotropism of Non-small cell lung cancer metastasis

Lung cancer is the second most commonly diagnosed cancer type world-wide and the type responsible for the most cancer-related deaths (18%)[123]. Lung cancer can be broadly separated into two classes: small cell lung cancer and non-small cell lung cancer (NSCLC), representing 20% and 80% of lung cancer cases, respectively[124]. If caught in the earliest stage (stage I), the five-year survival rate is between 49%-45%. However, NSCLC often goes undetected until a late stage when most symptoms arise, resulting in roughly 40%-50% of

patients presenting with stage IV lung cancer (distant metastatic spread) at the time of diagnosis[125].

When NSCLC metastasizes, it non-randomly targets certain organs, most often going to the brain, then in decreasing order, to the bones, liver, adrenal glands, thoracic cavity, and distant lymph nodes[116,117,126]. Research has determined that several factors are involved in determining the tendencies of specific cancer types to colonize specific secondary sites, involving proximity and patterns of blood flow, tumor-intrinsic properties, and interactions between the primary tumor and foreign microenvironments[127]. Despite these resent advances, there is still a great deal of obscurity surrounding the specific changes required for NSCLC to successfully colonize the brain.

1.3 Outstanding questions in the field and motivation for thesis

1.3.1 Do methylation changes within enhancers exhibit shared characteristics across cancer types?

While the field of cancer epigenomics has been fast-evolving, at the time we began the work for Chapter 2 the majority of largescale DNA methylation comparisons across cancer types had been primarily restricted to assessing a limited portion of the genome – primarily targeting promoters[14,51]. Advances in methylation-assessing technologies have now allowed for the comprehensive profiling of tumor samples, which have demonstrated a large degree of methylation abnormalities outside of promoters, within enhancers[128,129]. Due to their abundant role in reregulating cell type-specific gene expression, we hypothesized that DNA methylation abnormalities within enhancers might contribute to loss of original cell type identity. While this is a largescale question with much additional work needed to solidify mechanisms (see Chapter 7), the first step needed to address this question was to perform a global comparison of enhancer

methylation changes across different cancer types, providing the motivation for the work presented in Chapter 2.

1.3.2 What allows non-small cell lung cancer to routinely, successfully colonize the brain?

Although correlations have been found between the expression patterns of various genes and the likelihood of metastasis[118], far fewer studies have drilled down to answer mechanistically whether the associated genes play a role in the process, and the processes responsible for how these genes are losing/gaining function. To date, few specific genetic mutations that confer a selective advantage in metastatic cells as compared to primary tumor cells have been found[116,119,130], and even fewer that contribute to tissue-specific dissemination. However, this does not mean that such mutations do not exist. It may be that these studies were underpowered to detect such mutations (particularly if mutations confer a selective metastatic advantage specific to the tumor of origin/metastasis site pair), and that many individual mutations may not have a large effect sizes on their own. Nevertheless, this observation lead us to hypothesize that additional mechanisms contribute to potentiating metastasis, including expression alterations and epigenetic modifications, each potentially contributing similar functional changes.

To increase power over previous studies to detect changes that may help propagate NSCLC metastasize to brain, we reasoned that the use of paired primary and metastasis samples from the same patients could be leveraged. Therefore, samples pertaining to a cohort of 45 patients were isolated, and genomic, transcriptiomic, and epigenomic profiling was completed. The initial phase of characterizing recurrent changes (Chapter 3 and Chapter 5) revealed some interesting trends that might help explain NSCLC brain organotropism, including an overall activation of brain-specific transcriptional networks, accompanied by methylation changes within brain-specific active enhancers and increased variant allele frequencies of many

epigenetic regulators in the metastases relative to the primary samples. Large, recurrent methylation increase in the metastasis samples over DNA methylation valleys provided the motivation for Chapter 4's exploration of EZH2 binding dynamics in metastasis and their relationship with methylation changes.

1.4 References

1. Understanding what cancer is: ancient times to present. *American Cancer Society* Available at: https://www.cancer.org/cancer/cancer-basics/history-of-cancer/what-is-cancer.html. (Accessed: 30th September 2021)
2. Weinstein, I. B. & Case, K. The history of cancer research: Introducing an AACR centennial series. *Cancer Res.* **68**, 6861–6862 (2008).
3. Rous, P. A sarcoma of the fowl transmissible by an agent separable from the tumor cells. *J. Exp. Med.* **13**, 397–411 (1911).
4. Rous, P. A transmissible avian neoplasm. (sarcoma of the common fowl.). *J. Exp. Med.* **12**, 696–705 (1910).
5. Stehelin, D., Varmus, H. E., Bishop, J. M. & Vogt, P. K. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* **260**, 170–173 (1976).
6. Knudson, A. G. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc. Natl. Acad. Sci.* **68**, 820–823 (1971).
7. Lander, S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
8. Sjöblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science (80-.).* **314**, 268–274 (2006).
9. McLendon, R. *et al.* Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
10. Weinstein, J. N. *et al.* Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315–322 (2014).
11. Collisson, E. A. *et al.* Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
12. Bass, A. J. *et al.* Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).
13. Davis, C. F. *et al.* The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* **26**, 319–330 (2014).
14. Hoadley, K. A. *et al.* Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158**, 929–944 (2014).
15. Agrawal, N. *et al.* Integrated Genomic Characterization of Papillary Thyroid Carcinoma. *Cell* **159**, 676–690 (2014).
16. Lawrence, M. S. *et al.* Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576–582 (2015).
17. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
18. Akbani, R. *et al.* Genomic Classification of Cutaneous Melanoma. *Cell* **161**, 1681–1696 (2015).
19. Ciriello, G. *et al.* Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* **163**, 506–519 (2015).
20. Bell, D. *et al.* Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
21. Abeshouse, A. *et al.* The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**,

1011–1025 (2015).
22. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N. Engl. J. Med.* **374**, 135–145 (2016).
23. Zheng, S. *et al.* Comprehensive pan-genomic characterization of adrenocortical carcinoma. *Cancer Cell* **29**, 723–736 (2016).
24. Campbell, J. D. *et al.* Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* **48**, 607–616 (2016).
25. Kim, J. *et al.* Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169–174 (2017).
26. Burk, R. D. *et al.* Integrated genomic and molecular characterization of cervical cancer. *Nature* **543**, 378–384 (2017).
27. Fishbein, L. *et al.* Comprehensive Molecular Characterization of Pheochromocytoma and Paraganglioma. *Cancer Cell* **31**, 181–193 (2017).
28. Cherniack, A. D. *et al.* Integrated Molecular Characterization of Uterine Carcinosarcoma. *Cancer Cell* **31**, 411–423 (2017).
29. Farshidfar, F. *et al.* Integrative Genomic Analysis of Cholangiocarcinoma Identifies Distinct IDH-Mutant Molecular Profiles. *Cell Rep.* **18**, 2780–2794 (2017).
30. Ally, A. *et al.* Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* **169**, 1327-1341.e23 (2017).
31. Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
32. Raphael, B. J. *et al.* Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **32**, 185-203.e13 (2017).
33. Robertson, A. G. *et al.* Integrative Analysis Identifies Four Molecular and Clinical Subsets in Uveal Melanoma. *Cancer Cell* **32**, 204-220.e15 (2017).
34. Robertson, A. G. *et al.* Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell* **171**, 540-556.e25 (2017).
35. Abeshouse, A. *et al.* Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell* **171**, 950-965.e28 (2017).
36. Radovich, M. *et al.* The Integrated Genomic Landscape of Thymic Epithelial Tumors. *Cancer Cell* **33**, 244-258.e10 (2018).
37. Shen, H. *et al.* Integrated Molecular Characterization of Testicular Germ Cell Tumors. *Cell Rep.* **23**, 3392–3406 (2018).
38. Hmeljak, J. *et al.* Integrative molecular characterization of malignant pleural mesothelioma. *Cancer Discov.* **8**, 1549–1565 (2018).
39. Hammerman, P. S. *et al.* Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
40. Koboldt, D. C. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
41. Ley, T. J. *et al.* Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
42. The Cancer Genome Atlas Research Network *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).
43. Creighton, C. J. *et al.* Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
44. Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155**, 462

(2013).
45. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
46. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
47. Campbell, P. J. *et al.* Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
48. Taylor, A. M. *et al.* Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* **33**, 676-689.e3 (2018).
49. Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
50. Li, Y. *et al.* Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112–121 (2020).
51. Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e6 (2018).
52. Kastan, M. B., Onyekwere, O., Sidransky, D., Vogelstein, B. & Craig, R. W. Participation of p53 protein in the cellular response to DNA damage. *Cancer Res.* **51**, 6304–6311 (1991).
53. Kang, S., Bader, A. G. & Vogt, P. K. Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 802–807 (2005).
54. Boehm, J. S. & Hahn, W. C. Towards systematic functional characterization of cancer genomes. *Nat. Rev. Genet.* **12**, 487–498 (2011).
55. Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).
56. Pirozzi, C. J. & Yan, H. The implications of IDH mutations for cancer development and therapy. *Nat. Rev. Clin. Oncol.* **18**, 645–661 (2021).
57. Schmitt, M. W., Loeb, L. A. & Salk, J. J. The influence of subclonal resistance mutations on targeted cancer therapy. *Nat. Rev. Clin. Oncol.* **13**, 335–347 (2016).
58. Dowell, J. E. & Minna, J. D. EGFR mutations and molecularly targeted therapy: A new era in the treatment of lung cancer. *Nat. Clin. Pract. Oncol.* **3**, 170–171 (2006).
59. Castel, P., Toska, E., Engelman, J. A. & Scaltriti, M. The present and future of PI3K inhibitors for cancer therapy. *Nat. Cancer* **2**, 587–597 (2021).
60. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science (80-.).* **357**, (2017).
61. Feinberg, A. P. & Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301**, 89–92 (1983).
62. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).
63. He, Y. & Ecker, J. R. Non-CG Methylation in the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **16**, 55–77 (2015).
64. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).
65. Bestor, T., Laudano, A., Mattaliano, R. & Ingram, V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. Mol. Biol.* **203**, 971–983 (1988).

66. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
67. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: Mechanism, function and beyond. *Nat. Rev. Genet.* **18**, 517–534 (2017).
68. Dawlaty, M. M. *et al.* Loss of tet enzymes compromises proper differentiation of embryonic stem cells. *Dev. Cell* **29**, 102–111 (2014).
69. En, L., Beard, C. & Jaenisch, R. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature* **366**, 362–365 (1993).
70. Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72–77 (2017).
71. Wutz, A. Gene silencing in X-chromosome inactivation: Advances in understanding facultative heterochromatin formation. *Nat. Rev. Genet.* **12**, 542–553 (2011).
72. Mohandas, T., Sparkes, R. S. & Shapiro, L. J. Reactivation of an Inactive Human X Chromosome: Evidence for X Inactivation by DNA Methylation. *Science (80-.).* **211**, 393–396 (1981).
73. Jackson, M. *et al.* Severe Global DNA Hypomethylation Blocks Differentiation and Induces Histone Hyperacetylation in Embryonic Stem Cells. *Mol. Cell. Biol.* **24**, 8862–8871 (2004).
74. Jones, P. A. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492 (2012).
75. Jones, P. A. The DNA methylation paradox. *Trends Genet.* **15**, 34–37 (1999).
76. Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
77. Xu, J. *et al.* Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12377–12382 (2007).
78. Jones, P. A. & Baylin, S. B. The Epigenomics of Cancer. *Cell* **128**, 683–692 (2007).
79. Jang, H. S. *et al.* Transposable elements drive widespread expression of oncogenes in human cancers. *Nat. Genet.* **51**, 611–617 (2019).
80. Robertson, K. D. DNA methylation and human disease. *Nat. Rev. Genet.* **6**, 597–610 (2005).
81. Morita, S. *et al.* Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat. Biotechnol.* **34**, 1060–1065 (2016).
82. Sapozhnikov, D. M. & Szyf, M. Unraveling the functional role of DNA demethylation at specific promoters by targeted steric blockage of DNA methylatransferase with CRISPR/dCas9. *Nat. Commun.* **12**, (2021).
83. A, M. *et al.* 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat. Med.* **1**, 686–692 (1995).
84. Herman, J. G. *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 9700–9704 (1994).
85. Yoshiura, K. *et al.* Silencing of the E-cadherin invasion-suppressor gene by CpG methylation in human carcinomas. *Proc. Natl. Acad. Sci.* **92**, 7416–7419 (1995).
86. Herman, J. G. *et al.* Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6870–6875 (1998).
87. Esteller, M. *et al.* Epigenetic inactivation of LKB1 in primary tumors associated with the

Peutz-Jeghers syndrome. *Oncogene* **19**, 164–168 (2000).
88. Baylin, S. B. & Herman, J. G. DNA hypermethylation in tumorigenesis: Epigenetics joins genetics. *Trends Genet.* **16**, 168–174 (2000).
89. Lei, Y. *et al.* Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. *Nat. Commun.* **8**, 1–10 (2017).
90. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).
91. Pflueger, C. *et al.* A modular dCas9-SunTag DNMT3A epigenome editing system overcomes pervasive off-target activity of direct fusion dCas9-DNMT3A constructs. *Genome Res.* **28**, 1193–1206 (2018).
92. Aran, D., Sabato, S. & Hellman, A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol.* **14**, (2013).
93. Liu, X. S. *et al.* Editing DNA Methylation in the Mammalian Genome. *Cell* **167**, 233-247.e17 (2016).
94. Hong, L., Schroth, G. P., Matthews, H. R., Yau, P. & Bradbury, E. M. Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 'tail' to DNA. *J. Biol. Chem.* **268**, 305–314 (1993).
95. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).
96. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
97. Jambhekar, A., Dhall, A. & Shi, Y. Roles and regulation of histone methylation in animal development. *Nat. Rev. Mol. Cell Biol.* **20**, 625–641 (2019).
98. Kooistra, S. M. & Helin, K. Molecular mechanisms and potential functions of histone demethylases. *Nat. Rev. Mol. Cell Biol.* **13**, 297–311 (2012).
99. Musselman, C. A., Lalonde, M. E., Côté, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* **19**, 1218–1227 (2012).
100. Li, Y. *et al.* AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell* **159**, 558–571 (2014).
101. Fujisawa, T. & Filippakopoulos, P. Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nat. Rev. Mol. Cell Biol.* **18**, 246–262 (2017).
102. Chi, P., Allis, C. D. & Wang, G. G. Covalent histone modifications-miswritten, misinterpreted and mis-erased in human cancers. *Nat. Rev. Cancer* **10**, 457–469 (2010).
103. Krivtsov, A. V. & Armstrong, S. A. MLL translocations, histone modifications and leukaemia stem-cell development. *Nat. Rev. Cancer* **7**, 823–833 (2007).
104. Wang, G. G. *et al.* Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **459**, 847–851 (2009).
105. Van Haaften, G. *et al.* Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nat. Genet.* **41**, 521–523 (2009).
106. Johnstone, R. W. Histone-deacetylase inhibitors: Novel drugs for the treatment of cancer. *Nat. Rev. Drug Discov.* **1**, 287–299 (2002).
107. Piunti, A. & Shilatifard, A. The roles of Polycomb repressive complexes in mammalian development and cancer. *Nat. Rev. Mol. Cell Biol.* **22**, 326–345 (2021).
108. Blackledge, N. P. & Klose, R. J. The molecular principles of gene regulation by Polycomb

repressive complexes. *Nat. Rev. Mol. Cell Biol.* **0123456789**, (2021).
109. Wu, X., Johansen, J. V. & Helin, K. Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. *Mol. Cell* **49**, 1134–1146 (2013).
110. Schuettengruber, B., Bourbon, H. M., Di Croce, L. & Cavalli, G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **171**, 34–57 (2017).
111. Schuettengruber, B., Martinez, A. M., Iovino, N. & Cavalli, G. Trithorax group proteins: Switching genes on and keeping them active. *Nat. Rev. Mol. Cell Biol.* **12**, 799–814 (2011).
112. Morgan, M. A. & Shilatifard, A. Chromatin signatures of cancer. *Genes Dev.* **29**, 238–249 (2015).
113. Varambally, S. *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624–629 (2002).
114. Kim, K. H. & Roberts, C. W. M. Targeting EZH2 in cancer. *Nat. Med.* **22**, 128–134 (2016).
115. Hanahan, D. & Weinberg A, R. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
116. Nguyen, D. X. & Massagué, J. Genetic determinants of cancer metastasis. *Nat. Rev. Genet.* **8**, 341–352 (2007).
117. Vanharanta, S. & Massagué, J. Origins of Metastatic Traits. *Cancer Cell* **24**, 410–421 (2013).
118. Veer, L. J. Van *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
119. Brastianos, P. K. *et al.* Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov.* **5**, 1164–1177 (2015).
120. Steeg, P. S. Metastasis suppressors alter the signal transduction of cancer cells. *Nat. Rev. Cancer* **3**, 55–63 (2003).
121. Rinker-Schaeffer, C. W., O'Keefe, J. P., Welch, D. R. & Theodorescu, D. Metastasis suppressor proteins: Discovery, molecular mechanisms, and clinical application. *Clin. Cancer Res.* **12**, 3882–3889 (2006).
122. Brew, K., Dinakarpandian, D. & Nagase, H. Tissue inhibitors of metalloproteinases: Evolution, structure and function. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1477**, (2000).
123. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
124. Duma, N., Santana-Davila, R. & Molina, J. R. Non–Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. *Mayo Clin. Proc.* **94**, 1623–1640 (2019).
125. Goldstraw, P. *et al.* The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM Classification for lung cancer. *J. Thorac. Oncol.* **11**, 39–51 (2016).
126. Niu, F.-Y. *et al.* Distribution and prognosis of uncommon metastases from non-small cell lung cancer. *BMC Cancer* **16**, 149 (2016).
127. Gao, Y. *et al.* Metastasis Organotropism: Redefining the Congenial Soil. *Dev. Cell* **49**, 375–391 (2019).
128. Zhang, B. *et al.* Comparative DNA methylome analysis of endometrial carcinoma reveals

complex and distinct deregulation of cancer promoters and enhancers. *BMC Genomics* **15**, (2014).
129. Nagarajan, R. P. *et al.* Recurrent epimutations activate gene body promoters in primary glioblastoma. *Genome Res.* **24**, 761–774 (2014).
130. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 4283–4288 (2008).

Chapter 2: Common DNA methylation dynamics in endometrioid adenocarcinoma and glioblastoma suggest universal epigenomic alterations in tumorigenesis

The contents of this chapter have been adapted from the published manuscript:

> "Karlow, J.A., Miao, B., Xing, X. et al. Common DNA methylation dynamics in endometrioid adenocarcinoma and glioblastoma suggest universal epigenomic alterations in tumorigenesis. Commun Biol 4, 607 (2021). https://doi.org/10.1038/s42003-021-02094-1".

2.1 Abstract

Trends in altered DNA methylation have been defined across human cancers, revealing global loss of methylation (hypomethylation) and focal gain of methylation (hypermethylation) as frequent cancer hallmarks. Although many cancers share these trends, little is known about the specific differences in DNA methylation changes across cancer types, particularly outside of promoters. Here, we present a comprehensive comparison of DNA methylation changes between two distinct cancers, endometrioid adenocarcinoma (EAC) and glioblastoma multiforme (GBM), to elucidate common rules of methylation dysregulation and changes unique to cancers derived from specific cells. Both cancers exhibit significant changes in methylation over regulatory elements. Notably, hypermethylated enhancers within EAC samples contain several transcription factor binding site clusters with enriched disease ontology terms highlighting uterine function, while hypermethylated enhancers in GBM are found to overlap active enhancer marks in adult

brain. These findings suggest that loss of original cellular identity may be a shared step in tumorigenesis.

2.2 Introduction

Since the advent of the first reference human genome in 2001[1], the identification of genetic mutations in cancer has largely established cancer as a genetic disease. Many groups have compared the mutational landscape across different cancer types to identify functional genomic mutations and pathways mechanistically linked to cancer-type-specific and pan-cancer tumorigenesis[2–8]. The more recent identification of epigenetic alterations in cancer has revealed increased complexity of cancer gene regulation, extending the view of cancer abnormalities beyond simply genetic alterations[5,9,10].

One epigenetic modification in particular, DNA methylation, has long been associated with gene expression, where promoter methylation absence and high gene body methylation positively correlate with gene expression[11–13]. Studies regarding DNA methylation changes and additional epigenetic modifications have highlighted important roles these alterations play in cancer, leading the field to now recognize cancer as both a genetic and epigenetic disease[9,10,14–19]. DNA methylation alterations, in particular genome-wide loss and local gains within promoters, are considered hallmarks of many cancers[9,15–19] and could possibly be causal[20–22]. As DNA methylation can impact gene expression, methylation alterations in cancer likely impact the tumor phenotype by modulating the regulatory landscape, thereby helping shape the cancer's cell fate[10,23]. Although epigenetic abnormalities have been observed in many cancers, their specific functions and possible roles in tumorigenesis remain unclear. Moreover, commonalities in the locations of abnormal DNA methylation residing outside promoters and their functional roles

across different cancer types remain understudied, as the majority of pan-cancer DNA methylation analyses to date have included only array-based methylation data[2,5,10,24].

We have previously generated global DNA methylation profiles for endometrial cancers[25] as well as for glioblastoma multiforme (GBM)[26], revealing altered DNA methylation over regulatory enhancers and promoters, as well as hypomethylated gene body promoters, respectively. In addition, studies of colon cancer[27–29] and of multiple cancer cell lines[30,31] suggest that enhancer methylation is drastically altered in cancers and is closely related to altered transcriptional profiles. Together, these results suggest that the regulatory landscape in cancer may be altered directly, through methylation changes within regulatory elements, or indirectly, through reassignment of regulatory element target genes.

The compilation of reference human epigenomes for a variety of cell types and tissues allows for a comprehensive, cell-type-specific annotation of epigenetic abnormalities[32]. We hypothesize that by comparing the global epigenetic abnormalities of two highly distinct cancer types in the context of non-malignant cell-type epigenomes, we can unveil both shared and unique epigenetic mechanisms contributing to cancer. To better understand how epigenetic abnormalities differ between cancers in both location and function, we directly compared deeply profiled DNA methylomes of two distinct cancer types, GBM and endometrioid adenocarcinoma (EAC), whose DNA methylation abnormalities have been previously identified[25,26]. GBM, also known as grade IV astrocytoma, originates from astrocytes and quickly develops into highly heterogeneous malignancies[33] . Roughly 90% of GBM cases are classified as primary GBM and are associated with a poor clinical outcome[33], contributing to an overall 5-year survival rate of about 5.5%[34]. Uterine corpus cancer, of which 90% of cases originate in the endometrium[35], is associated with a much better prognosis[36]. Endometrial cancer is broadly classified into two

categories. Type 1 tumors, which include EAC, constitute the majority of cases, are typically low grade, only moderately differentiated, and are hormone-sensitive[37–39]. If detected in an early, localized stage, the 5-year survival rate for uterine corpus cancers can be as high as 85–96%[36]. GBM and EAC appear to have little in common, suggesting they are good candidates for identifying shared and unique epigenetic abnormalities with predicted functional impacts on cancer phenotype.

In this study, we directly compare the DNA methylation abnormalities of these two cancer types and annotate their likely impacts on gene regulation using normal reference epigenomes. Our results indicate that both cancer types exhibit thousands of local, recurrent DNA methylation abnormalities, in the form of both increased (hyper) and decreased (hypo) methylation, which are significantly enriched in genomic regions annotated as regulatory elements, such as promoters and enhancers. Only ~50% of these DNA methylation abnormalities fall within regions targeted by the Infinium 450k array, the most common platform for DNA methylation profiling chosen by projects including The Cancer Genome Atlas (TCGA), highlighting the importance of whole-genome, unbiased approaches for profiling cancer DNA methylomes. Despite being very distinct diseases, EAC and GBM share a significant number of differentially methylated regions (DMRs). Notably, both cancers demonstrate an enrichment of hyperDMRs within the apoptosis pathway and hypoDMRs within the long terminal repeat (LTR) retrotransposon subfamily MER52A, alluding to potential pan-cancer signatures. We further report that clusters of enhancer hyperDMRs in EAC defined by the presence of binding sites for enriched TFs most notably enrich for uterine disease ontology. In addition, enhancer hyperDMRs in GBM largely overlap active enhancer histone modifications in adult brain tissue. These results suggest that regulatory regions of genes involved in functions related to the cancer's original cell

type often become methylated during tumor progression, perhaps contributing to the loss of a phenotypic cellular identity experienced by cancer cells.

Taken together, our results support findings that cancer is a complex disease with a large epigenetic component. Although the locations of DNA methylation abnormalities are often specific to a particular cancer type, many appear to functionally contribute in similar ways by potentially silencing cell-type-of-origin enhancers. Our results begin to shed light on the mechanistic principles that drive both common and cancer-type-specific DNA methylation abnormalities and their functional consequences.

2.3 Results

2.3.1 Comparative analysis of EAC and GBM DNA methylation abnormalities

We previously profiled the DNA methylomes of EAC[25] and GBM[26]. A combined technique of methylated DNA immunoprecipitation sequencing (MeDIP-seq) and methylation-sensitive restriction enzyme sequencing (MRE-seq), which detect methylated CpGs and unmethylated CpGs, respectively, was used for methylome profiling[40]. Complete DNA methylomes were generated for a total of five GBM samples[26], two normal frontal cortex brain samples[26], three EAC samples[25], and one normal endometrial sample pooled from ten healthy individuals[25]. DMRs between EAC samples and pooled normal endometrium (EAC DMRs), and between GBM samples and normal brain (GBM DMRs) were identified by integrating the MeDIP-seq and MRE-seq data using M&M[41], allowing for the comparison of two distinct cancer-type DMR sets in the context of their normal, tissue-specific DNA methylomes. Examination of DMRs called across tumors within a specific cancer type revealed considerable intertumoral heterogeneity. However, roughly 60% of DMRs identified in one sample were also discovered in an additional sample (Figure 2.1), suggesting that the requirement for a DMR to be

identified in at least two tumors to be considered, as in previous analyses[25,26], allowed for the identification of the majority of common epigenetic changes for the given cancer type. DMRs were further classified according to their direction of methylation change, a hypermethylated DMR (hyperDMR) being more highly methylated in the cancer than in the normal, and a hypomethylated DMR (hypoDMR) being less methylated. DMRs were also classified as either cancer-type unique (DMR present in only one of the two cancers) or shared between EAC and GBM, resulting in 6 categories: EAC-unique hyperDMRs, EAC-unique hypoDMRs, GBM-unique hyperDMRs, GBM-unique hypoDMRs, shared hyperDMRs, and shared hypoDMRs.

We identified 26,990 DMRs in EAC (10,414 (38.6%) of which were shared across all 3 samples) and 14,672 in GBM (426 (2.9%) of which were shared across all 5 GBM samples). The ratio of hyperDMRs to hypoDMRs was remarkably similar between GBM (ratio = 2.265 : 1) and EAC (ratio = 2.098 : 1). Although the number of shared hyperDMRs ($n = 2760$) drastically outnumbered the shared hypoDMRs ($n = 195$), both were highly significant ($p < 2.2\text{E}-308$ and $p < 6.0\text{E}-199$, respectively, hypergeometric tests, using as background 500 bp genomic regions with at least one CpG and MeDIP and/or MRE signal in at least one sample) (2.5 Methods and Figure 2.2a). Examination of the methylation levels of the identified shared EAC and GBM DMR regions within TCGA samples spanning 24 cancer types revealed similar trends (2.5 Methods). Specifically, 22/24 cancer types profiled in TCGA with both tumor and normal methylation data showed a significant increase in methylation from normal to tumor over shared EAC and GBM hyperDMRs (Figure 2.3). Similarly, 21/24 cancer types profiled in TCGA exhibited a significant decrease in methylation over shared EAC and GBM hypoDMRs (Figure 2.4), supporting that the methylation changes observed over these shared regions likely extend beyond EAC and GBM, possibly corresponding to a pan-cancer methylation signature.

To better understand the possible functional contribution of altered DNA methylation within EAC and GBM, we identified the fraction as well as enrichment of DMRs that fell into genomic regions annotated as promoters (1 kb and 2.5 kb), CpG islands, gene bodies, enhancers defined either by Fantom 5[42,43], or by the VISTA enhancer project[44], super enhancers[45], and intergenic regions. We found that the three hyperDMR groups (EAC-unique hyperDMRs, GBM-unique hyperDMRs, and shared hyperDMRs) exhibited a much higher percentage overlap to promoter regions, CpG islands, gene bodies, and super enhancers than did the three hypoDMR groups (EAC-unique hypoDMRs, GBM-unique hypoDMRs, and shared hypoDMRs), also reflected by increased relative enrichment (Figure 2.2b). Conversely, the hypoDMR groups exhibited a higher percentage overlap to intergenic regions, where the hyperDMR groups and EAC-unique hypoDMRs were depleted within intergenic regions (Figure 2.2b). The high enrichment of both hyperDMRs and hypoDMRs in Fantom and VISTA enhancers suggests that *cis*-regulatory enhancer elements may contain both activating and inactivating DNA methylation abnormality hotspots in cancer.

DMRs were further characterized according to their chromatin-state annotations, defined based on various histone modifications[32]. Taking advantage of the 18-state chromatin models (chromHMM) generated from complete human epigenome references for various cell and tissue types[46], we calculated the fraction and enrichment of DMRs overlapping each chromatin state across many different cell and tissue types ($n = 80$) (2.5 Methods). HyperDMR groups were generally enriched within regulatory regions annotated as active transcription start sites (TSSs), regions flanking TSSs (both upstream and downstream), genic enhancers, and active enhancers across different tissues (Figure 2.2c and Figure 2.5). In contrast, hypoDMRs exhibited a range of enrichment values in active chromatin states, varying from depletion to weak enrichment (Figure

2.2c and Figure 2.5). Although hyperDMRs were most highly enriched within bivalent/poised TSSs across tissues, hypoDMRs were generally depleted or had little enrichment over these regions. Conversely, hypoDMRs were enriched within weakly repressed polycomb regions across tissues, whereas hyperDMRs were generally depleted or showed no enrichment (Figure 2.2c and Figure 2.5). HyperDMRs and hypoDMRs were both enriched within repressed polycomb regions and bivalent/poised enhancers, although the magnitude of enrichment was much higher for hyperDMRs in both cases. In concordance with the most severe gain of methylation in both cancer types residing over polycomb and bivalent regions, we found that EZH2 binding in normal human astrocytes[47] was highly enriched within GBM hyperDMRs, suggesting that these tumors undergo an epigenetic switch involving gain of DNA methylation over regions normally bound by EZH2 in non-malignant astrocytes (Table 2.1). These results demonstrate that DMRs are enriched within regulatory regions in both cancers, both within and outside promoters, highlighting the complexity of DNA methylation alterations within cancer[27–31].

2.3.2 EAC and GBM DMRs exhibit similar characteristics on a pathway level

To better understand the shared functionality of DMRs between EAC and GBM, we compared the frequency at which they contained potential regulatory regions. We found that although only about half of all EAC and GBM hypoDMRs overlapped regions with regulatory annotations, this percentage increased to roughly 90% for EAC and GBM hyperDMRs, reflecting a large enrichment within both cancers. We found that EAC and GBM hyperDMRs largely overlapped regions annotated as both promoters and enhancers (44.09% and 41.12%, respectively), whereas an additional 27.49% and 33.32% overlapped promoters only, and an additional 17.63% and 12.54% overlapped active enhancers only (2.5 Methods and Figure 2.6a).

As the silencing of tumor suppressor genes (TSGs) by DNA methylation is a common mechanism of gene inactivation in cancer[20–22,48–50], we examined the frequency at which DMRs overlapped TSG core promoters. We found that the number of TSGs suffering hypermethylation within their core promoters in EAC and GBM, as well as those overlapping shared hyperDMRs, were statistically significant ($p < 2.22\text{E}-08$, $p < 5.27\text{E}-05$, and $p < 1.94\text{E}-03$, respectively; hypergeometric tests) (Appendix A.1). The number of TSGs suffering hypomethylation within their core promoters was not significant for either EAC or GBM individually, or for those overlapping shared hypoDMRs (Appendix A.1). In addition, the number of TSGs undergoing hypermethylation in both EAC and GBM was higher than expected by chance ($p < 8.39\text{E}-14$; hypergeometric test), suggesting that methylation of some of the same TSGs might be a shared attribute of these cancers (Figure 2.6b and Appendix A.2).

As enhancers frequently undergo DNA methylation changes during tumorigenesis, we further examined distal DMRs near TSGs. Hypermethylation in promoters and/or active enhancers was found in 350 TSGs in EAC and 245 TSGs in GBM (Figure 2.6c and Appendix A.3). Hypermethylation of regulatory elements around TSGs was generally associated with decreased expression of TSGs in these two cancer types.

We next examined the commonalities in distal DNA methylation changes between these two cancer types on a pathway level. In our study, the majority of EAC and GBM DMRs fell outside gene promoters; however, many of these non-promoter DMRs (EAC hyperDMRs: 72%, GBM hyperDMRs: 63%, EAC hypoDMRs: 46%, GBM hypoDMRs: 45%) exhibited an active enhancer signature in at least one of the 80 different tissues or cell types provided by Roadmap Epigenomics Consortium[32]. Therefore, we defined non-promoter DMRs that fell within regions annotated as an active enhancer (state “9_EnhA1” or “10_EnhA2”) in at least 1 of 80

epigenomes as cancer-enhancer DMRs (ceDMRs). By comparing enriched biological processes associated with merged ceDMRs (2.5 Methods), we found that both cancer-type cancer-enhancer hyperDMRs (ce-hyperDMRs) were highly enriched for terms related to apoptosis, a process commonly deregulated in cancer cells[51], sometimes through DNA methylation alterations[52] (Figure 2.6d and Appendix A.4). In EAC, 36 of the 140 apoptosis-associated genes contained hyperDMRs within regulatory elements, 4 of which contained hyperDMRs over both promoters and enhancers. Similarly, in GBM, 30 of the 140 apoptosis-associated genes contained regulatory element hyperDMRs, 3 of which exhibited both promoter and enhancer hyperDMRs (Figure 2.6e). One gene in particular, *BCL2L11*, contained a single hyperDMR that spanned both the gene's promoter and an active enhancer in both cancer types (Figure 2.7). Although both EAC and GBM hyperDMRs were enriched within apoptosis pathway genes, only 11 genes gained regulatory region methylation in both cancer types (Figure 2.6e), highlighting the complex strategies different cancers might take to reach the same functional consequences.

2.3.3 Abnormally methylated enhancer-potential regions in cancer are associated with deregulated transcription factors

DNA cytosine methylation status has been shown to be associated with transcription factor (TF)-binding events[53,54], although the mechanism linking the two remains unclear. Motif discovery within ceDMRs identified highly enriched sets of TF-binding sites (TFBSs; Figure 2.8a), most of which were exclusive to one cancer type. Expression of TFs with motifs enriched in EAC ce-hyperDMRs were downregulated in EAC compared to normal endometrium (Figure 2.8b), whereas TFs whose bindings motifs were enriched in EAC cancer-enhancer hypoDMRs (EAC ce-hypoDMRs) were generally upregulated in EAC (Figure 2.8c), suggesting that changes in TF expression may help dictate changes in DNA methylation at target motifs in EAC.

In support of this directionality, we observed cases of hypomethylation in GBM over non-brain enhancers, accompanied by gain in expression of both the TFs with predicted binding motifs and their target genes, illustrated in two examples (Figure 2.8d and Figure 2.8e). Neuropipin-2 (NRP2), a non-tyrosine kinase receptor frequently overexpressed in various malignancies, including GBM, regulates endosome maturation and EGFR trafficking, supporting the growth and replication of cancer cells[55]**.** We identified a GBM hypoDMR in the *NRP2* gene body, located 70.5 kb downstream of the *NRP2* TSS (Figure 2.8d). RNA polymerase II ChIA-PET data in HeLa cells generated by the ENCODE consortium indicated a direct physical interaction between the *NRP2* TSS and the hypoDMR in this particular cell line, arguing that these two genomic regions have the potential to interact. In addition, this region also contained strong H3K27ac and H3K4me1 signal in adipose tissue, indicative of an active enhancer in this cell type. Motif analysis revealed the presence of a TCF12-binding site within this hypoDMR and chromatin immunoprecipitation sequencing (ChIP-seq) data in a cancer cell line (A549) supported strong binding of TCF12 in this enhancer region. *TCF12* and *NRP2* were found to be relatively highly expressed in adipose tissue, lowly expressed in the brain frontal cortex, and highly expressed again in the GBM. Therefore, this site reflects a possible co-opted adipose enhancer that lost DNA methylation in GBM cells, possibly as a result of abnormal upregulation of *TCF12* in GBM, resulting in an upregulation of *NRP2*. Similarly, CD248 (Endosialin) marks tumor-associated pericytes in high-grade glioma[56], where blocking *CD248* can inhibit the growth and differentiation of perivascular cells[57]. We identified a GBM hypoDMR located ~9 kb upstream of *CD248*, which contained an enhancer with strong H3K27ac and H3K4me1 signal in adipocyte tissue (Figure 2.8e). K562 RNA polymerase II ChIA-PET data suggested this enhancer has the capability to physically interact with the *CD248* TSS. An EGR1-binding motif was

identified within this hypoDMR and both *EGR1* and *CD248* were more highly expressed in adipose and GBM than in normal brain. This suggests another example where GBM is adopting the potential regulation of *CD248* by *EGR1* seen in adipose tissue.

2.3.4 Gain of methylation over original cell-type enhancers may contribute to loss of cellular identity during cancer progression

Although DNA methylation alterations in GBM and EAC exhibited many commonalties, the two cancers also displayed distinct signatures, reflecting tissue type specificity[5]. To better understand these unique differences, we first calculated the enriched vertebrate TF-binding motifs within the ceDMRs using Homer[58] and opposite ceDMR groups as background (2.5 Methods). Enriched motifs were then filtered to only include those TFs with significant expression changes in TCGA, and ceDMRs were then clustered based on presence or absence of these enriched motifs, using a distance method of "Euclidean" and a clustering method of "complete" (2.5 Methods). Clusters of ceDMRs were identified by cutting the dendrogram at various heights, which resulted in robust groupings and enriched Gene Ontology (GO) terms. Clustering ceDMRs by the presence of enriched TF-binding motifs revealed several clusters of DMRs with similar subsets of TF-binding motifs, likely reflecting the high similarity among binding motifs of related TFs, e.g., GATA family (Cluster 3) and SMAD family (Cluster 10) in EAC ce-hyperDMRs (Figure 2.9a) and the FOX family (Cluster 4) in GBM ce-hyperDMRs (Figure 2.9b). However, despite a propensity for clusters of DMRs to harbor motifs for a handful of TFs, many different, sometimes largely non-overlapping clusters of EAC ce-hyperDMRs (clusters 1, 3, 4, 6, 7, 8, 9, and 10) enriched for similar disease ontology terms, most commonly centered around uterine neoplasia (Figure 2.9a), suggesting that targeted silencing of intrinsic cell-identity pathways may contribute to cancer progression. GBM ce-hyperDMR clusters of

enriched TF-binding motifs, on the other hand, enriched for terms most notably related to heart functions (Figure 2.9b).

To better understand whether brain functionality was being targeted by hypermethylation in GBM, we classified whether distal GBM hyperDMRs (those located outside promoters) overlapped the enhancer-defined chromatin state ("7_Enh" based on the chromHMM 15-state model) for 13 adult and fetal brain-related tissues using the brain epigenome references generated by the Roadmap Epigenomics Consortium[32]. We found that a significantly greater proportion of GBM hyperDMRs overlapped annotated enhancers in adult brain tissues as opposed to in fetal astrocyte and progenitor cells, with the exception of the male fetal brain sample (Figure 2.10a and Figure 2.10b). To characterize the potential activity of brain enhancers that gained methylation in GBM, we determined whether they overlapped ChIP-seq peaks for H3K27ac and H3K4me1 in fetal and adult brain tissues (2.5 Methods). We found that the majority of GBM hyperDMR brain enhancers overlapped H3K27ac peaks in adult brain (60.96–72.50%), whereas very few GBM hyperDMR brain enhancers overlapped H3K27ac peaks in fetal brain (22.04%) (Figure 2.10c and Figure 2.10d). The percentage of H3K27ac peaks overlapping GBM hyperDMR brain enhancers was much greater in the adult brain samples (0.725–0.877%) compared to the fetal brain sample (0.314%), suggesting that the increased number of GBM hyperDMR brain enhancers overlapping H3K27ac peaks in adult brain is not due to increased global H3K27ac signal, but is in fact specific. Likewise, the majority of GBM hyperDMR brain enhancers overlapped H3K4me1 peaks in adult brain (66.87–80.24%), whereas a significantly smaller proportion overlapped H3K4me1 peaks in fetal brain (31.38–66.20%) (Figure 2.10c and Figure 2.10e). In addition, although the percentage of H3K4me1 peak base pairs within GBM hyperDMR brain enhancers was similar between adult and fetal brain samples

(0.68–1.01% vs. 0.463–0.803%, respectively), a higher percentage overlap in the adult brain samples indicates the increase in H3K4me1 signal in adult brain is not simply due to increased background H3K4me1 signal. These results suggest that although a portion of GBM hyperDMR brain enhancers bear the mark of active enhancers (H3K4me1) in the developing brain, a significantly greater percentage contain these marks in adult brain tissues. In addition, most of these enhancers are not located in open chromatin (H3K27ac peaks) in developing tissue, suggesting they might be more active in adult tissue and, therefore, important for the maintenance of normal functionality of mature glia cells as opposed to brain and glial cell development.

When considering TFs with motifs enriched within EAC ce-hypoDMRs, 30 exhibited a corresponding increase in expression in TCGA EAC samples (a subset of the TCGA uterine corpus endometrial carcinoma (UCEC) cohort). Clustering these 30 TFs based on enriched motif locations within the ce-hypoDMRs revealed 6 clusters of more than 1 TF. In contrast to targeting enhancers related to the cell-type-of-origin function, EAC ce-hypoDMR clusters that contained enriched TF-binding motifs often enriched for various cancer types, most notably hemangiomas (Figure 2.11a). Similarly, 67 TFs with enriched motifs within GBM ce-hypoDMRs demonstrated an increased expression in TCGA GBM samples, comprising 9 TF clusters. The majority of enrichment terms across all clusters of associated ce-hypoDMRs were related to various tumors, most notably gastrointestinal (Figure 2.11b). Taken together, we observe that enhancer-potential regions with loss of methylation, containing enriched motifs for TFs that exhibit increased expression, are primarily tied to a variety of cancers, suggesting that the loss of methylation over and potential activation of aberrant enhancers may be a common trend among distinct cancer types.

2.3.5 Distinct spectrum of epigenetic abnormalities within transposable elements in cancer

Transposable elements (TEs) are hotspots of epigenetic abnormalities during carcinogenesis and were generally believed to be globally hypomethylated in cancer cells[59]. We observed that 39–62% of DMRs contained TEs (GBM: 44% of hyperDMRs and 49% of hypoDMRs; EAC: 39% of hyperDMRs and 62% of hypoDMRs). Although 23.77% and 12.02% of GBM and EAC hyperDMR-overlapped TEs, respectively, fell within RefSeq-defined promoters, an additional 39.32% and 46.54%, which were located outside canonical promoters, were annotated as TSSs in at least 1 of the 80 Roadmap cell/tissue types, based on chromHMM 18-state chromatin predictions (Figure 2.12a). In addition, 18.53–27.85% of DMR-overlapped TEs outside promoters fell within predicted active enhancer regions based on Roadmap Epigenomics data chromHMM 18-state chromatin predictions (Figure 2.12a). We estimated the enrichment of epigenetically altered TE subfamilies and discovered distinct patterns within GBM and EAC (Figure 2.12b). A small number of TE subfamilies, including LTR16A1, MLT1C, and MER52A, were highly enriched in both GBM and EAC hypoDMRs, where the primate-specific LTR retrotransposon MER52A exhibited the highest enrichment (Figure 2.12b). We further examined the DNA methylation level of MER52A copies in normal tissues and cancer, and found that the majority of MER52A subfamily copies were highly methylated in various normal tissues, but became demethylated in cancer (Figure 2.12c). A small number of MER52A copies were lowly methylated in normal breast myoepithelial cells, liver, and pancreas tissues, suggesting that some MER52A copies maintain an active epigenetic state and may provide regulatory functions in normal cells. Finally, we found that several TEs overlapping hypoDMRs in GBM encompassed H3K4me3 signal, a mark of active promoters, in the GBM cell line U87,

consistent with previous findings of hypomethylated TEs providing cryptic promoters during tumorigenesis[60] and embryonic development[61] (Figure 2.12d).

2.4 Discussion

In the present study, we sought to expand the current view of methylation alteration comparisons across distinct cancer types. By utilizing DNA methylation data derived from MeDIP-seq and MRE-seq, we were able to comprehensively explore the propensity for methylation alterations in cancer to be specific, based on their cell type of origin, and the ways in which alterations were shared physically or functionally between two distinct cancer types: GBM and EAC. We identified thousands of DMRs in both cancers, highly enriched within regulatory regions including promoters and enhancers.

The Infinium 450k array platform and more recent 850k array platform have been the standard of practice for measuring methylation and the method of choice for many TCGA studies[2,5,62–64]. When validating our DMRs using cancer DNA methylation data generated by TCGA with the Infinium 450K platform, we found that roughly half of our DMRs could not be detected, due to the platform's relatively low coverage (52.55% and 46.55%, EAC and GBM DMRs, respectively) (Table 2.2). Even when comparing to the more recent 850K platform, we still found that 41.85% (EAC) and 38.13% (GBM) of our DMRs were missed (Table 2.2). HypoDMRs often did not contain a probe (75.87% (450k) and 62.44% (850k) for EAC, and 81.55% (450k) and 67.00% (850k) for GBM), whereas more than half of all hyperDMRs contained probes (58.57% (450k) and 67.97% (850k) for EAC, and 68.90% (450k) and 74.62% (850k) for GBM. Forty-one to 49% of the CpGs within DMRs profiled via our method but missed by the Infinium 450K array were located within regions annotated as having active enhancer potential based on Roadmap's chromHMM 18-state annotations. As we and others

have demonstrated that methylation at enhancer regions can play an important role in cancer[25,26,65], the inclusion of this set of CpGs not covered by the Infinium array, but covered by our method, is instrumental in understanding the impact DNA methylation abnormalities have on cancer.

Although the combined use of MeDIP-seq and MRE-seq to interrogate genome-wide methylation levels has many advantages over array-based techniques, it should be noted that this approach is not without limitations. For example, MeDIP-seq enriches for genomic regions with a methylated CpG; however, the exact CpG that is methylated within a given read captured is unknown. In addition, MRE-seq is limited to interrogate restriction sites for enzymes used in the protocol, which only cover a small fraction of all genomic CpG sites[66]. Although these methods do not produce a quantitative measurement of the methylation status at each CpG, as is generated using whole-genome bisulfite sequencing (WGBS), together these methods provide complementary data that can be used to computationally predict the methylation status of individual CpGs, which recapitulate WGBS results well and at a fraction of the cost[41].

Upon discovering that there were many more shared hyperDMRs between the two cancer types than expected by chance, we sought to elucidate possible functions within these regions. As expected, we found that TSG core promoters were often enriched within hyperDMRs. Examination of expression changes over TSGs with enhancer and/or promoter hypermethylation revealed general expression loss within tumors compared to normals, suggesting that different cancers jointly alter the methylation status of TSG regulatory regions, possibly contributing to their loss of expression in cancer.

GBM and EAC hyperDMRs also enriched for active enhancer regions, which displayed the unique enrichment of several biological processes and pathways, as well as shared

enrichment regarding processes such as apoptosis. Examination of 140 apoptosis-related genes revealed several with increased promoter methylation in both cancer types, whereas many more accrued methylation changes within the surrounding enhancers. These results suggest that silencing genes involved in the apoptotic signaling pathway via methylation at enhancers and promoters may be a common mechanism shared across cancer types.

The extent to which DNA methylation alterations shape transcriptional activity remains unclear in cancer. To better understand the relationship between altered DNA methylation and phenotypic impact via gene expression changes, we identified enriched TF-binding motifs in both EAC and GBM ceDMRs. We found that TFs associated with EAC ce-hyperDMRs exhibited reduced gene expression in cancer when compared to normal endometrium, whereas TFs associated with EAC ce-hypoDMRs exhibited increased expression. Similarly, in GBM, we observed specific incidences of methylation loss over regions bearing marks of active enhancers in alternative cell types, accompanying an increased expression of encompassed motif TFs. These results suggest that altered TF abundance may likely be driving the differential methylation patterns over regulatory regions in these cancers.

GO enrichment analyses of EAC ce-hyperDMR sets that contained clusters of enriched TF-binding motifs often enriched for GO terms associated with the original cell type—primarily uterine-specific disease terms. Although GBM ce-hyperDMRs did not show a similar enrichment of brain-specific disease terms, examination of their annotations across numerous normal brain and developing brain tissues revealed that hundreds more GBM hyperDMRs could be annotated as enhancers in adult brain tissues rather than in fetal brain. In addition, H3K4me1 and H3K27ac peaks within adult brain tissues were more commonly found in GBM hyperDMRs compared with peaks in developing brain tissues, suggesting that enhancers gaining methylation in GBM

are more active in maintaining adult brain function as opposed to developing brain function. In contrast, ce-hypoDMRs with enriched TF-binding motifs were primarily associated with cancer-related GO terms. These results are consistent with the "cancer cell-identity crisis" hypothesis[67]: during carcinogenesis, tissue-specific enhancers may become methylated and silenced in addition to the silencing of tissue-specific TFs, contributing to the loss of original cellular identity. Hypomethylated cancer enhancers and associated upregulated TFs may also contribute to carcinogenesis by activating pro-growth, pro-migration pathways, and genes specific to other cell types, resulting in a deregulated cell fate. This concept is further illustrated by two examples of loss of methylation in GBM over enhancers active in a distant tissue type accompanied by increased expression of the predicted TF binding the enhancer and the target gene (Figure 2.8d and Figure 2.8e).

As TEs have been shown to harbor regulatory elements and have been routinely filtered out in methylation array-based studies, we examined the methylation status across various TE subfamilies in both cancer types. A large proportion of both EAC and GBM hyper- and hypoDMR-overlapped TEs had either enhancer and/or promoter potential, as determined by Roadmap's reference human epigenome annotations. Distinct cancer-specific methylation abnormalities were found in GBM and EAC, possibly associated with tissue-specific activity, consistent with the observation that TEs can play tissue-specific enhancer roles[68]. However, enrichment of the retrotransposon subfamily MER52A was observed in both cancer-type hypoDMRs and across several additional cancer types, suggesting a potential role for this subfamily in carcinogenesis. Finally, several instances of hypomethylated TEs exhibited the active promoter histone mark, H3K4me3, in the U87 cell line (GBM), suggesting that a shared

cancer mechanism may include altering the gene regulatory landscape through the demethylation of regulatory elements harbored within TEs.

2.5 Methods

2.5.1 Statistics and reproducibility

A description of all statistical methods used for each test can be found in the sections below.

2.5.2 DMR calling guidelines

For a genomic region to be called an EAC DMR, the region must have been differentially methylated between the cancer and the normal endometrium in at least two of the three EAC samples[25]. For a genomic region to be called a GBM DMR, the region must have been differentially methylated between the cancer and both normal brain samples in at least two of the five GBM samples[26]. In both cases, DMRs were defined at a 500 base pair resolution using the M&M tool[41]. The M&M tool integrates MeDIP-seq and MRE-seq data from two different samples and determines regions where the methylation levels are significantly different. In both previous studies where DMRs were called[25,26], default parameters were set, which included "mreratio = 3/7" (the ratio of the percentage of unmethylated genome to the percentage that is methylated), "method = 'XXYY'" (specifying to use MeDIP and MRE in testing), "psd = 2" (pseudo count added to MeDIP and MRE reads), "mkadded = 1" (pseudo count added to the number of CpGs in total and MRE-CpGs), "a = 1e − 16" (*p*-value cutoff when sum of observations is smaller than "top"), "cut = 100" (*p*-value cutoff when less than the sum of observations), and "top = 500" (*p*-value cutoff when less than the sum of observations and *p*-values < "a"). Additional default parameters used for selecting significant DMRs included "up = 1.45" (minimum threshold for MeDIP1/MeDIP2 read ratio), "p.value.MM = 0.01" (*p*-value threshold), "p.value.SAGE = 0.01" (SAGE *p*-value threshold), "*q*-value = 0.00005" (*q*-value threshold), "cutoff = '*q*-value'" (measurement to use to call significance), and "quant = 0.6" (minimum threshold for the rank of the absolute value of the difference between MeDIP1 and MeDIP2).

2.5.3 Determining significance of DMR overlap

To calculate the significance of hyperDMRs and hypoDMRs shared by EAC and GBM, a hypergeometric test was performed using the phyper() function in R. More specifically, for calculating the significance of shared hyperDMRs, the values considered were as follows: 10,178 (total GBM hyperDMRs), 18,278 (total EAC hyperDMRs), 2760 (shared hyperDMRs), and 5,196,471 (number of 500 bp genomic regions with at least 1 CpG, and MeDIP and/or MRE signal in at least 1 sample). To calculate the significance of observing at least 2760 shared hyperDMRs, we used "lower.tail=FALSE," as well as subtracted 1 from our "*x*" value (2760 − >2759). For calculating the significance of shared hypoDMRs, the values considered were as follows: 4494 (total GBM hypoDMRs), 8712 (total EAC hypoDMRs), 195 (shared hypoDMRs), and 5,196,471 (number of 500 bp genomic regions with at least 1 CpG, and MeDIP and/or MRE signal in at least 1 sample). To calculate the significance of observing at least 195 shared DMRs, we used "lower.tail=FALSE," and subtracted 1 from our "*x*" value (195 − >194). To calculate the expected number of shared hyper/hypoDMRs, the following equation was used:

$$expected\ \#\ of\ shared\ DMRs = \left(\frac{\#\ of EAC\ DMRs}{\#\ of\ background\ 500bp\ bins}\right) * \#\ of\ GBM\ DMRs \quad (1)$$

2.5.4 Methylation validation using TCGA array data

Hg19-aligned TCGA methylation array-based datasets were downloaded from https://gdc.cancer.gov using the gdc-client (v1.6.0) for the following cancers: adrenocortical carcinoma, bladder urothelial carcinoma, breast invasive carcinoma, cervical squamous cell carcinoma and endocervical adenocarcinoma, cholangiocarcinoma, colon adenocarcinoma, lymphoid neoplasm diffuse large B-cell lymphoma, esophageal carcinoma, GBM, head and neck squamous cell carcinoma, kidney chromophobe, kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, acute myeloid leukemia, brain lower grade glioma, liver hepatocellular carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, mesothelioma, ovarian serous cystadenocarcinoma, pancreatic adenocarcinoma, pheochromocytoma and paraganglioma, prostate adenocarcinoma, rectum adenocarcinoma, sarcoma, skin cutaneous melanoma, stomach adenocarcinoma, testicular germ cell tumors, thyroid carcinoma, thymoma, UCEC, uterine carcinosarcoma, and uveal melanoma. Methylation files corresponding to the same patient ID were averaged at each probe location. The average methylation values over probes overlapping DMRs were calculated for each tumor sample and normal sample (when available).

2.5.5 Genomic characterization of DMRs

To determine the genomic landscape of each DMR class, we considered the following genomic regions:

- Promoters (1 kb core (500 bp upstream to 500 bp downstream the TSS) and 2.5 kb (2 kb upstream the TSS to 500 bp downstream the TSS), defined by refGene (last updated: 3 April 2016), downloaded from the UCSC Gene Annotation Database[69] (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/)).
- Unmasked CpG Islands (last updated: 1 June 2014), downloaded from the UCSC Gene Annotation Database[69] (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/).
- Gene bodies, defined by refGene (last updated: 3 April 2016), downloaded from the UCSC Gene Annotation Database[69] (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/).
- Fantom 5 Enhancers, human permissive enhancers phase 1 and 2 (http://fantom.gsc.riken.jp/5/datafiles/latest/extra/Enhancers/)[42,43].
- VISTA Enhancers (1745 human enhancers downloaded on 21 December 2015) human elements[44] (Appendix A.5).
- Super enhancers (defined by dbSUPER[45]).
- Intergenic regions, defined by refGene (last updated: 3 April 2016), downloaded from the UCSC Gene Annotation Database[69] (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/).

For each DMR class, we computed the fraction of DMR nucleotides that overlapped each genomic category. As these genomic categories are not mutually exclusive, DMR positions may have been counted more than once if they applied to multiple categories. Therefore, the percentages for each DMR group may sum to more than 1. Background regions considered when calculating enrichment were 500 kb genomic regions that excluded the ends of chromosomes, excluded bins overlapping blacklisted CpGs, excluded bins not containing at least one MeDIP

and/or MRE read for at least one sample, excluded bins not containing at least one CpG, and excluded those on chrY (and chrX for GBM). Enrichment was then calculated as:

$$\boldsymbol{enrichment} = \frac{\boldsymbol{\%\ of\ DMR\ nucleotides\ overlapping\ genomic\ feature}}{\boldsymbol{\%\ of\ background\ region\ base\ pairs\ considered\ overlapping\ genomic\ feature}} \quad (2)$$

2.5.6 Epigenetic characterization of DMRs

To determine the distribution of epigenomic annotations for each DMR group, we used chromHMM maps predefined for 80 cell and tissue types[32], downloaded from Roadmap Epigenomics Data Portal, https://egg2.wustl.edu/roadmap/web_portal/ (Appendix A.5). For each DMR group, we calculated the percentage of DMR bps that overlapped each feature (defined according to the 18-state chromHMM model) in each cell/tissue type. To calculate enrichment, we divided the percentage of DMR bps overlapping the feature in the cell/tissue type by the percentage of background bps overlapping the feature in the cell/tissue type. Genomic regions considered for background purposes were defined as described above in 2.5.5.

2.5.7 Polycomb binding enrichment within DMRs

To determine whether GBM hyperDMRs were enriched for polycomb binding in normal brain, we first downloaded control-normalized EZH2 ChIP-seq data for ENCODE's NH-A sample (GSM1003532) in BigWig format and converted the file to bed format[47]. We then calculated the average normalized signal in GBM-unique hyperDMRs and GBM/EAC shared hyperDMRs, as well as the average normalized signal in background regions (described above in 2.5.5). Enrichment was then calculated as the ratio of the average signal in the DMR group to the average signal in the genomic background.

2.5.8 DMR overlap to 450k and 850k array probes

If a DMR contained at least one probe found on the 450k array or 850k array, the DMR was considered identifiable via the 450k platform or 850k platform, respectively. Locations of Illumina HumanMethylation450 BeadChip probes (Infinium HumanMethylation450K v1.2) were downloaded from https://support.illumina.com/array/array_kits/infinium_humanmethylation450_beadchip_kit/downloads.html ("HumanMethylation450 v1.2 Manifest File (CSV Format)"). Locations of Illumina MethylationEPIC BeadChip probes (850K array) were downloaded from https://support.illumina.com/array/array_kits/infinium-methylationepic-beadchip-kit/downloads.html ("Infinium MethylationEPIC v1.0 B4 Manifest File (CSV Format)").

2.5.9 DMR overlap with potential regulatory regions

We identified regions of the genome with possible regulatory function as any 200 bp region that was annotated as one of the following chromatin states in at least one of the cell or tissue types listed above: 1_TssA (Active TSS), 2_TssFlnk (Flanking Active TSS), 3_TssFlnkU (Flanking TSS Upstream), 4_TssFlnkD (Flanking TSS Downstream), 7_EnhG1 (Genic Enhancers 1), 8_EnhG2 (Genic Enhancers 2), 9_EnhA1 (Active Enhancers 1), 10_EnhA2 (Active Enhancers 2), and 11_EnhWk (Weak enhancers). We then calculated the percentage of DMR base pairs that overlapped regions of regulatory potential. To calculate enrichment for EAC DMRs, the background was calculated as the percentage of hg19 base pairs with chromHMM 18-state annotations for the cell and tissue types above, excluding chrY and chrM, those that overlapped blacklisted CpG 500 bp bins, bins without a CpG, and bins without MeDIP and/or MRE signal in

at least one sample (2,170,900,000 bp) that met the above criteria (1,109,392,000 bp (51.10%)). The background GBM was calculated similarly, additionally excluding chrX (of 2,076,745,800 bp, 1,081,211,800 bp (52.06%) met the above criteria).
We then calculated the percentage of DMR bps that overlapped specific types of regulatory regions: promoters (defined according to refGene and Roadmap chromHMM 18-state reference epigenomes), active enhancers (defined according to Roadmap chromHMM 18-state reference epigenomes), and both promoters and active enhancers. Regions of active enhancer potential were defined as any region annotated as "9_EnhA1" or "10_EnhA2" in at least 1 of the 80 tissues or cell-type chromHMM 18-state models described above. Promoter regions were defined as any region annotated as "1_TssA," "2_TssFlnk," "3_TssFlnkU," or "4_TssFlnkD" in at least 1 of the 80 tissues or cell-type chromHMM 18-state models in addition to regions defined using refGene TSS annotations, where a promoter spanned 2 kb upstream the TSS to 500 bp downstream the TSS.

2.5.10 TSG core promoter DMR overlap analysis

A human TSG list containing 1217 genes was obtained from the TSGene Tumor Suppressor Gene Database[70,71] (http://bioinfo.mc.vanderbilt.edu/TSGene/Human_TSGs.txt); however, only 1216 of these genes could be identified with RefSeq (missing gene: TRP53COR), so we proceeded with the list of 1216 TSGs. The TSS of each TSG was identified using RefSeq and then TSG core promoter regions were defined as the region spanning 500 bp upstream to 500 bp downstream the TSS. All transcripts for each TSG were considered.

2.5.11 Assigning enhancer regions to genes

A list of all genomic positions that were annotated as either state "9_EnhA1" or "10_EnhA2" in at least 1 Roadmap reference epigenome listed above, based on the chromHMM 18-state model[32], was compiled. DMRs not overlapping core promoter regions (1 kb regions centered around TSSs) were overlapped to these enhancer-potential regions. DMRs overlapping enhancers were assigned to the gene with the nearest TSS. If the nearest TSS was >500,000 bp away, the DMR was not assigned to any gene.

2.5.12 Gene expression changes associated with TSG DNA methylation alterations

Normalized TCGA RNA-seq data (level-3, reads per kilobase of transcript, per million mapped reads (RPKM)) and clinical metadata of EAC, their matched-control samples, and GBM were downloaded from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/). Expression of normal brain samples ($n = 28$, frontal cortex) was downloaded from GTEx. TSGs with core promoters and/or active enhancers (assigned as described above) overlapping EAC hyperDMRs and/or GBM hyperDMRs were identified. This resulted in a list of 350 TSGs with an EAC hyperDMR overlapping the core promoter and/or an active enhancer, and 245 TSGs with a GBM hyperDMR overlapping the core promoter and/or an active enhancer. However, of the 350 and 245 TSGs, only 310 and 210 had available expression data (see Appendix A.3). In the case of EAC hyperDMRs, two TSGs with regulatory regions overlapping DMRs (BRINP1 and CCAR2) had the same alias: DBC1. As the expression data were associated with the alias, DBC1 was only counted once. For each TSG with available RNA-seq data, the mean RPKM value was calculated for both cancer (EAC or GBM) and normal (normal endometrium or normal brain), and the expression fold change was calculated as:

$$TSG\ Expression\ fold\ change = log2\left(\frac{TSG\ Mean\ RPKM\ in\ Cancer}{TSG\ Mean\ RPKM\ in\ Normal}\right) \quad (3)$$

2.5.13 GO enrichment for merged ceDMRs

Consecutive 500 bp DMRs were merged for each DMR group (EAC hyperDMRs, GBM hyperDMRs, EAC hypoDMRs, and GBM hypoDMRs) and any merged DMRs that overlapped promoter regions (refGene annotations, 2.5 kb) were discarded. The remaining DMRs were filtered to only include those that overlapped a region annotated as an active enhancer (chromHMM 18-states "9_EnhA1" or "10_EnhA2") in at least 1 of the 80 tissues/cell types mentioned above. GREAT[72] was then run with the remaining DMRs (each group separately), using version 3.0.0 and default parameters (including the whole genome as background). GO biological processes terms that were significant in both the binomial and hypergeometric tests (BinomFDRQ ≤ 0.05 and HyperFDRQ ≤ 0.05), were within the top 500 ranked Binomial test terms and had a region fold enrichment ≥ 2 were considered.

2.5.14 Apoptosis pathway genes with DMRs overlapping promoters and enhancers

Apoptosis-associated genes were obtained from KEGG[73] (Entry: hsa04210, $n = 140$). Genes with a DMR overlapping their promoter (refGene, 2.5 kb) were identified. A list of active enhancer locations (chromHMM 18-states "9_EnhA1" and "10_EnhA2" in at least 1 of the 80 tissue/cell types listed above) was obtained and active enhancers (unmerged) that overlapped promoters (refGene, 2.5 kb) were removed. Remaining active enhancer regions were then assigned to the nearest gene (shortest distance to the TSS) and those that were assigned to apoptosis genes and that overlapped DMRs were identified.

2.5.15 TF expression differences in EAC vs. normal endometrium

To determine whether there was a correlation between ceDMRs and changes in TF expression, we used publicly available RNA-seq data from TCGA (https://portal.gdc.cancer.gov/). TFs with enriched motifs (see method below in 2.5.16 with the exception: q-value ≤ 0.01) present in at least 20% of the DMRs were considered. Genes with an RPKM value ≤ 1 were excluded. RPKM values were then log2 transformed and *z*-scored.

2.5.16 Motif and GO analysis for ceDMRs

EAC ce-hyperDMRs

The Homer[58] (v4.9) function "findMotifsGenome.pl" was run using as input EAC ce-hyperDMRs: EAC hyperDMRs that did not overlap RefSeq promoters but did overlap an active enhancer annotation ("9_EnhA1" or "10_EnhA2") in at least 1 of 80 tissues/cell types from Roadmap's chromHMM 18-state model predictions (listed above). EAC ce-hypoDMRs were used as background. Aside from the pre-specified background and the flag "-size given," default parameters were used to detect enriched TFBSs using known vertebrate motifs ($n = 364$). Resulting enriched motifs were then filtered to include only those with a q-value ≤ 0.05. Remaining motifs were then matched with their most likely TF using Homer's Motif database and the expression of each TF was calculated in TCGA EAC samples and normal endometrium samples. TFs with a significant loss of expression in the EAC samples relative to the normal samples (*t*-test, Benjamini-Hochberg corrected) were retained. The Homer2 function "annotatePeaks.pl" was then run to identify the location of each enriched motif within the EAC ceDMRs, using default parameters. The binary matrix of EAC ce-hyperDMRs and enriched motifs (where 1 indicated the presence of the motif in the DMR and 0 indicated the absence) was

then clustered using R's heatmap.2 function with distance method "euclidean" and clustering method "complete." To identify clusters of TFs, the dendrogram was cut at a height of 42. All resulting groups that had more than one TF were called a cluster. For each cluster, DMRs that contained an enriched TF-binding motif were then checked for enriched Disease Ontology terms using GREAT[72] v3.0 and parameters: Species Assembly: Human GRCh37, Background regions: Whole genome; and default association rule settings. The top five enriched Disease Ontology terms are displayed for each cluster.

GBM ce-hyperDMRs

The same methods described for EAC ce-hyperDMRs were used here, with the following exceptions. GBM ce-hyperDMRs: GBM hyperDMRs that did not overlap RefSeq promoters but did overlap an active enhancer annotation ("9_EnhA1" or "10_EnhA2") in at least 1 of 80 tissues/cell types from Roadmap's chromHMM 18-state model predictions (Appendix A.5) were used as input to the Homer2 (v4.9) function "findMotifsGenome.pl." GBM ce-hypoDMRs were used as background. TFs with a significant loss of expression in TCGA GBM samples relative to normal brain were retained. To identify clusters, the dendrogram was cut at a height of 25.

EAC ce-hypoDMRs

The same methods described above were used here, with the following exceptions. EAC ce-hypoDMRs: EAC hypoDMRs that did not overlap RefSeq promoters but did overlap an active enhancer annotation ("9_EnhA1" or "10_EnhA2") in at least 1 of 80 tissues/cell types from Roadmap's chromHMM 18-state model predictions (listed above) were used as input to the Homer2 (v4.9) function "findMotifsGenome.pl." EAC ce-hyperDMRs were used as background. TFs with a significant gain of expression in TCGA EAC samples relative to normal endometrium were retained. To identify clusters, the dendrogram was cut at a height of 30.

GBM ce-hypoDMRs

The same methods described above were used here, with the following exceptions. GBM ce-hypoDMRs: GBM hypoDMRs that did not overlap RefSeq promoters but did overlap an active enhancer annotation ("9_EnhA1" or "10_EnhA2") in at least 1 of 80 tissues/cell types from Roadmap's chromHMM 18-state model predictions (listed above) were used as input to the Homer2 (v4.9) function "findMotifsGenome.pl." GBM ce-hyperDMRs were used as background. TFs with a significant gain of expression in TCGA GBM samples relative to normal brain were retained. To identify clusters, the dendrogram was cut at a height of 22.

2.5.17 Data sources for enhancer hypomethylation examples

ChIA-PET HeLa RNAPol2: ENCODE data portal (https://www.encodeproject.org/). Adipose H3K4me1, H3K4me3, H3K27ac: processed adipose ChIP-seq data (bigWig: H3K4me1, H3K4me3, and H3K27ac) were downloaded from the ENCODE data portal (https://www.encodeproject.org/). A549 TCF12 ChIP-seq: ENCODE data portal (https://www.encodeproject.org/). ChIA-PET K562 RNAPol2: ENCODE data portal (https://www.encodeproject.org/). Adipose (subcutaneous, visceral) and brain (frontal cortex) RNA-seq: GTEx. GBM expression: TCGA (https://portal.gdc.cancer.gov/).

2.5.18 Determining GBM hyperDMR overlap to H3K27ac and H3K4me1 peaks in adult and fetal brain samples

We began by starting with all 500 bp GBM hyperDMRs that overlapped state "7_Enh" (chromHMM 15-state model) in at least one fetal or adult Roadmap brain sample (E053, E054, E067, E068, E069, E070, E071, E072, E073, E074, E081, E082, and E125), and that did not overlap with refGene-defined 2.5 kb promoters. We then obtained H3K4me1 peak files from Roadmap (downloaded from Roadmap Epigenomics Data Portal

(https://egg2.wustl.edu/roadmap/web_portal/)) for the following samples: E053, E054, E070, E081, E082, E125 (fetal brain) and E067, E068, E069, E071, E072, E073, E074 (adult brain), as well as H3K27ac narrow peak files from Roadmap (downloaded from Roadmap Epigenomics Data Portal (https://egg2.wustl.edu/roadmap/web_portal/)) for the following samples: E125 (fetal brain) and E067, E068, E069, E071, E072, E073, and E074 (adult brain). All H3K4me1 adult brain peak files were merged to generate one adult brain H3K4me1 signal file and all H3K4me1 fetal brain peak files were merged to generate one fetal brain H3K4me1 signal file. All H3K27ac adult brain peak files were merged to generate one adult brain H3K27ac signal file and the E125 H3K27ac peak file was used as the fetal brain H3K27ac signal file.

2.5.19 TE DMR overlap and subfamily enrichment

RepeatMasker annotations were downloaded from the UCSC Genome Browser[69]. Simple repeats and low-complexity repeats were removed from annotation. The number of DMRs that overlapped TEs were determined using bedtools (v2.27.1). Regions of DMRs that overlapped TEs were extracted and the percentages of bps that overlapped genic promoters (RefSeq annotations, 2.5 kb) were determined. Regions not overlapping genic promoter annotations were then tested to see if they overlapped epigenomically defined promoter states (regions annotated as "1_TssA," "2_TssFlnk," "3_TssFlnkU," or "4_TssFlnkD" in at least 1 of the 80 tissues or cell-type chromHMM 18-state models). Finally, remaining regions were tested to see if they overlapped active enhancer regions (regions annotated as "9_EnhA1" or "10_EnhA2" in at least 1 of the 80 tissues or cell-type chromHMM 18-state models described above). Subfamily enrichment was calculated as:

$$Es = \frac{\frac{n_{TE}}{n_{DMR}}}{\frac{N_{TE}}{N_{all}}} \quad (4)$$

Where n_{TE} is the number of DMRs containing TEs, n_{DMR} is the total number of DMRs, N_{TE} is the number of genomic windows overlapped with TEs in the human genome, and N_{all} is the number of 500 bp windows in the human genome (hg19).

2.5.20 MER52A subfamily DNA methylation measurements across various normal and cancer tissues and cell types

WGBS data were downloaded from the Roadmap data portal (http://www.roadmapepigenomics.org/) and the ENCODE data portal (https://www.encodeproject.org/). The CpG sites were filtered to only include those that had at least 5× coverage. The average methylation level of each MER52A copy was calculated for generating boxplots and heat maps.

2.5.21 H3K4me3 signal over hypomethylated TEs in GBM

H3K4me3 ChIP-seq signal (bigWig) file for U87 (normalized reads per million) was downloaded from GEO (GSM2634761). The bigWig file was visualized on the WashU Epigenome Browser[74,75].

2.6 Data Availability

The data analyzed in the present study can be accessed as described below. Source data used to generate manuscript figures are available in the Github repository:

https://github.com/jaflynn5/EAC_GBM_comparative_epigenomics, which is also linked to the Zenodo repository with the identifier [DOI: 10.5281/zenodo.4637753]. Any additional source data can be obtained from the corresponding authors upon reasonable request.

- EAC MeDIP-seq + MRE-seq: [GEO: GSE51565].
- GBM MeDIP-seq + MRE-seq: [EGA: EGAS00001000685].
- Hg19 RefSeq annotations (last updated: 3 April 2016), downloaded from the UCSC Gene Annotation Database (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/).
- Hg19 unmasked CpG islands (last updated: 1 June 2014), downloaded from the UCSC Gene Annotation Database (http://hgdownload.cse.ucsc.edu/goldenpath/hg19/database/).
- Fantom 5 Enhancers, human permissive enhancers phase 1 and 2: (http://fantom.gsc.riken.jp/5/datafiles/latest/extra/Enhancers/).
- VISTA Enhancers (1745 human enhancers downloaded on 21 December 2015) human elements: https://enhancer.lbl.gov/.
- Super enhancers: https://asntech.org/dbsuper/.
- Blacklisted CpGs: http://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeMapability.
- chromHMM 18-state maps and 15-state maps: downloaded from Roadmap Epigenomics Data Portal, https://egg2.wustl.edu/roadmap/web_portal/.
- NH-A EZH2 ChIP-seq: [GEO: GSM1003532].
- 450k and 850k array probe locations: Locations of Illumina HumanMethylation450 BeadChip probes (Infinium HumanMethylation450K v1.2) were downloaded from https://support.illumina.com/array/array_kits/infinium_humanmethylation450_beadchip_kit/downloads.html ("HumanMethylation450 v1.2 Manifest File (CSV Format)"). Locations of Illumina MethylationEPIC BeadChip probes (850K array) were downloaded from https://support.illumina.com/array/array_kits/infinium-methylationepic-beadchip-kit/downloads.html ("Infinium MethylationEPIC v1.0 B4 Manifest File (CSV Format)").
- Tumor suppressor genes: https://bioinfo.uth.edu/TSGene/.
- TCGA EAC, normal endometrium, and GBM RNA-seq data (level-3, RPKM) and clinical metadata of EAC, GBM, and their matched-control samples were downloaded from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/).
- Adipose (subcutaneous, visceral) and brain (frontal cortex) RNA-seq: GTEx.
- Apoptosis gene list: KEGG[73] (Entry: hsa04210, $n = 140$).
- ChIA-PET HeLa RNAPol2: ENCODE data portal (https://www.encodeproject.org/).
- Processed adipose ChIP-seq data (bigWig, H3K4me1, H3K4me3, and H3K27ac) were downloaded from the ENCODE data portal (https://www.encodeproject.org/).
- A549 TCF12 ChIP: ENCODE data portal (https://www.encodeproject.org/).
- ChIA-PET K562 RNAPol2: ENCODE data portal (https://www.encodeproject.org/).
- H3K27ac and H3K4me1 Peaks in Adult and Fetal Brain Samples: We obtained H3K4me1 peak files from Roadmap (downloaded from the Roadmap Epigenomics Data Portal (https://egg2.wustl.edu/roadmap/web_portal/)) for the following samples: E053, E054, E070, E081, E082, E125 (fetal brain); and E067, E068, E069, E071, E072, E073, E074 (adult brain); as well as H3K27ac narrow peak files from Roadmap (downloaded from the Roadmap Epigenomics Data Portal (https://egg2.wustl.edu/roadmap/web_portal/)) for the following samples: E125 (fetal brain); and E067, E068, E069, E071, E072, E073, E074 (adult brain).

- RepeatMasker annotations were downloaded from the UCSC Genome Browser[69] https://hgdownload.soe.ucec.edu/download.html.
- WGBS data (thymus, ovary, pancreas, lung, mid-frontal cortex, brain germinal matrix, frontal cortex neuron, frontal cortex glia, atrium, sigmoid colon, colon tumor, colorectal cell line HCT116, breast myoepithelial, breast cancer HCC1954, liver, HepG2) were downloaded from the Roadmap data portal (http://www.roadmapepigenomics.org/) and the ENCODE data portal (https://www.encodeproject.org/).
- U87 H3K4me3 ChIP-seq: [GEO: GSM2634761].
- TCGA Infinium 450k array probe data for all available cancer types: https://portal.gdc.cancer.gov/.
- Normal glia RNA-seq: [GEO: GSE41826].
- TCGA methylation array data: https://gdc.cancer.gov/.

2.7 Code Availability

Custom scripts generated for use in this study are available in the Github repository: https://github.com/jaflynn5/EAC_GBM_comparative_epigenomics, which is also linked to the Zenodo repository with the identifier [DOI: 10.5281/zenodo.4637753]. Any additional scripts can be obtained from the corresponding authors upon reasonable request. Software utilized include: R (v3.3.0), Homer2 (v4.9), and GREAT (v3.0).

2.8 Acknowledgements

This work was supported by the National Institutes of Health [U24ES026699, U01HG009391, and R25DA027995], the American Cancer Society [RSG-14-049-01-DMC], and the Goldman Sachs Philanthropy Fund [Emerson Collective]. Funding for open access charge: National Institutes of Health.

2.9 Author Information

2.9.1 Affiliations

The Edison Family Center for Genome Sciences and Systems Biology, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA

- Jennifer A Karlow, Benpeng Miao, Xiaoyun Xing, Ting Wang, & Bo Zhang

Center of Regenerative Medicine, Department of Developmental Biology, Washington University School of Medicine, St. Louis, MO, USA

- Benpeng Miao & Bo Zhang

McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA

- Ting Wang

2.9.2 Contributions

Conceptualization: T.W. and B.Z. Performed experiments: X.X. Data analysis: J.A.K., B.Z., B.M., and T.W. Manuscript preparation: J.A.K., B.Z., B.M., and T.W.

2.9.3 Corresponding authors

Correspondence to Ting Wang or Bo Zhang.

2.9.4 Ethics declarations

The authors declare no competing interests.

2.10 Figures

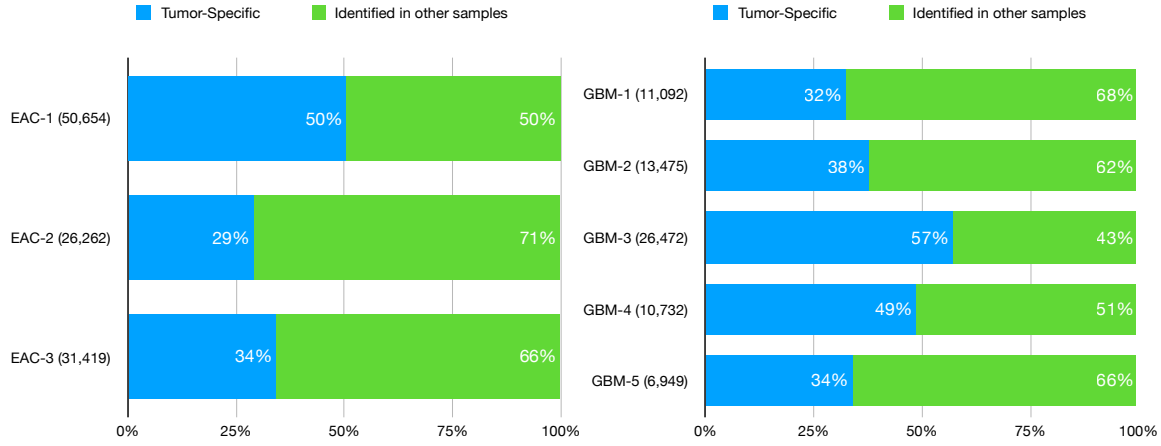


Figure 2.1. DMR variability across tumors within cancer types.
Number of DMRs identified for each tumor/normal comparison are depicted in parentheses adjacent the tumor ID. The percentage of each tumor's DMRs that were identified in at least one other tumor/normal comparison of the same cancer type are depicted in green ("Identified in other samples"). The percentage of each tumor's DMRs that were only identified in that tumor/normal pair for the particular cancer type are depicted in blue ("Tumor-Specific"). Left: EAC samples; Right: GBM samples.

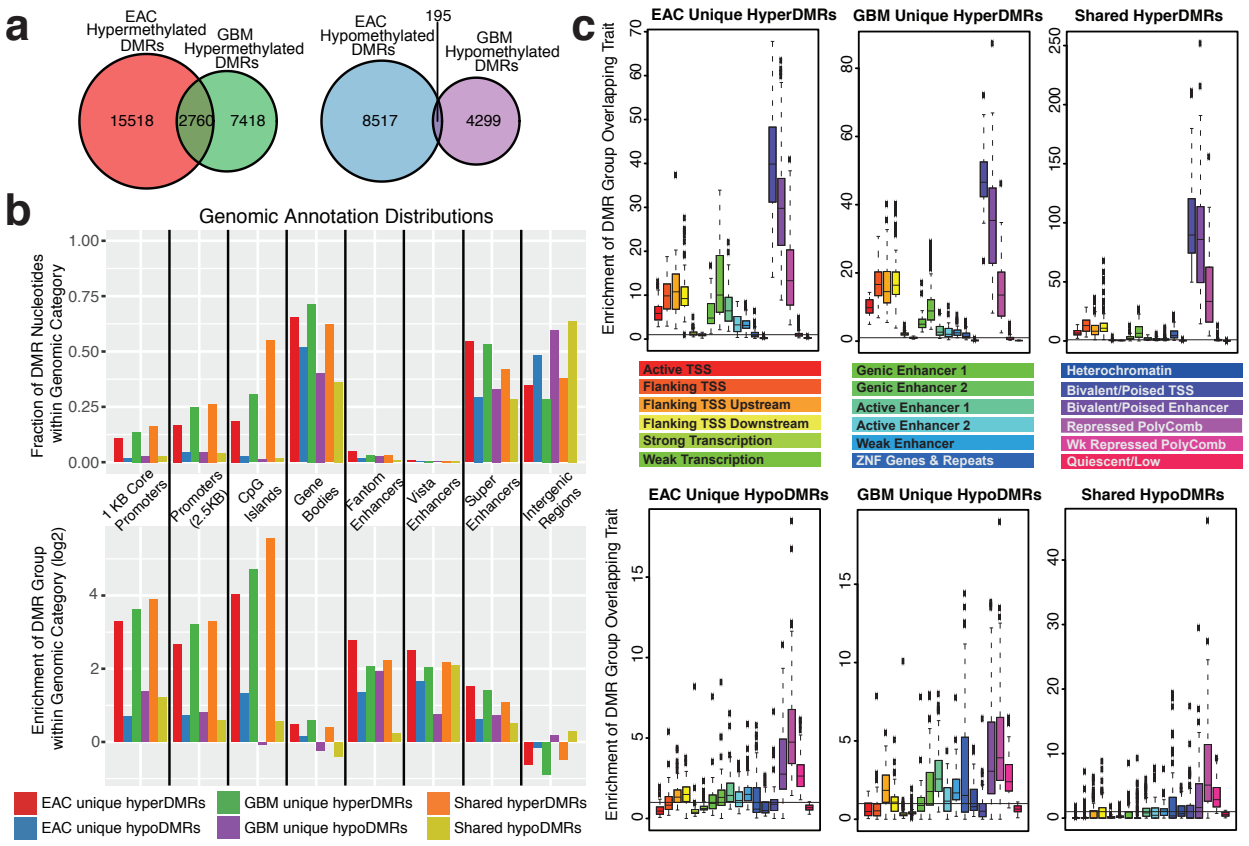


Figure 2.2. Comparative analysis of EAC and GBM DNA methylation abnormalities.
a. Overlap of 500 bp differentially methylated regions (DMRs) in EAC and GBM. Left: hypermethylated DMRs; right: hypomethylated DMRs. Both overlaps are significant ($p < 2.2E-308$ and $p < 6.0E-199$, respectively, hypergeometric tests). **b.** Genomic annotation distribution of EAC and GBM DMRs. Top: percentage of DMR group nucleotides within each genomic category. Bottom: enrichment of DMR group nucleotides within each genomic category. **c.** EAC and GBM DMR enrichment within epigenetic annotations (chromHMM 18-states) across a variety of cell and tissue types ($n = 80$, Human Roadmap Epigenome[32]). For each DMR group, the percentage of base pairs (bps) that overlapped each epigenetic state in each of the cell/tissue types was calculated, and enrichment scores were calculated by dividing those percent overlaps by the percentage of background bps that overlapped each epigenetic state in each cell/tissue type. Background regions considered when calculating enrichment in both **b** and **c** were 500 bp genomic regions that excluded chromosome ends, excluded bins overlapping blacklisted CpGs, excluded bins not containing at least one MeDIP and/or MRE read for at least one sample, excluded bins not containing at least one CpG, and excluded those on chrY (and chrX for GBM).

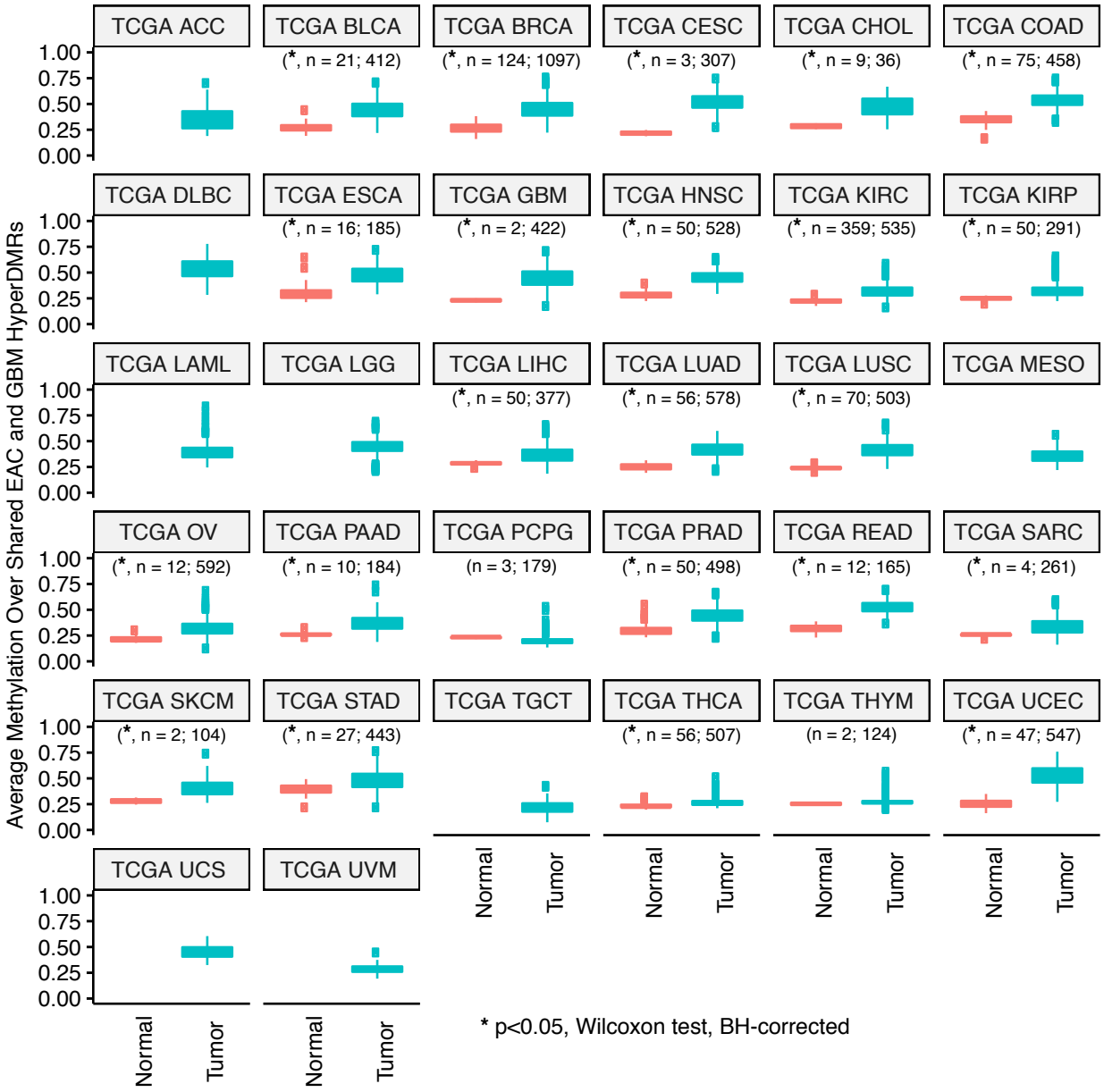


Figure 2.3. Methylation levels within TCGA samples over shared EAC and GBM hyperDMRs.

Average methylation over TCGA CpG probes for 32 cancer types, spanning shared EAC and GBM hyperDMRs in normal (red) and tumor (blue) samples, where data are available (*n*=number of normal cases; number of cancer type cases). Significant methylation changes (all in the form of methylation gain) were identified in 22/24 TCGA cancer types (t-tests, $p<0.05$, BH corrected).

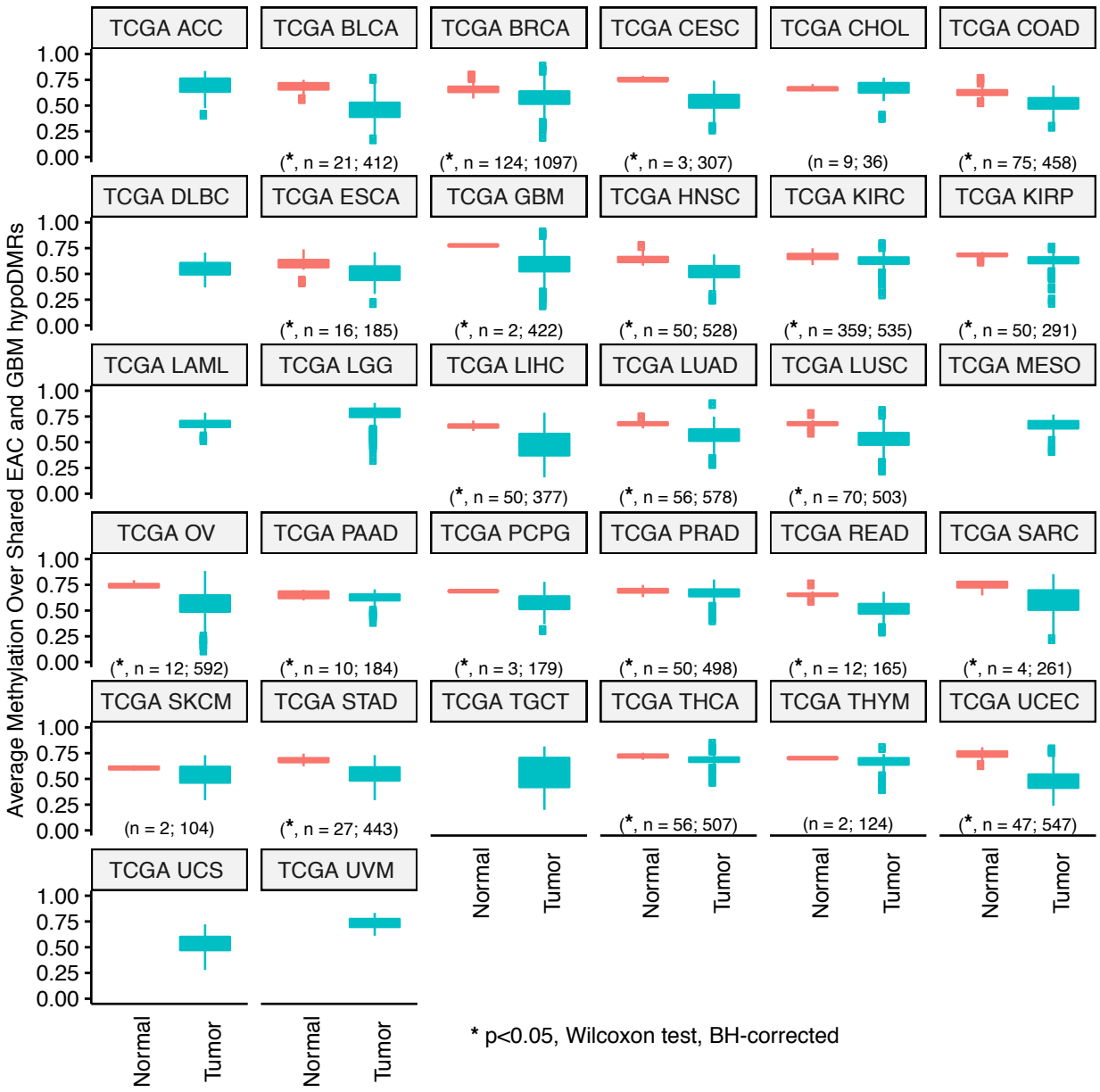


Figure 2.4. Methylation levels within TCGA samples over shared EAC and GBM hypoDMRs.

Average methylation over TCGA CpG probes for 32 cancer types, spanning shared EAC and GBM hypoDMRs in normal (red) and tumor (blue) samples, where data are available (*n*=number of normal cases; number of cancer type cases). Significant methylation changes (all in the form of methylation loss) were identified in 21/24 TCGA cancer types (t-tests, $p<0.05$, BH corrected).

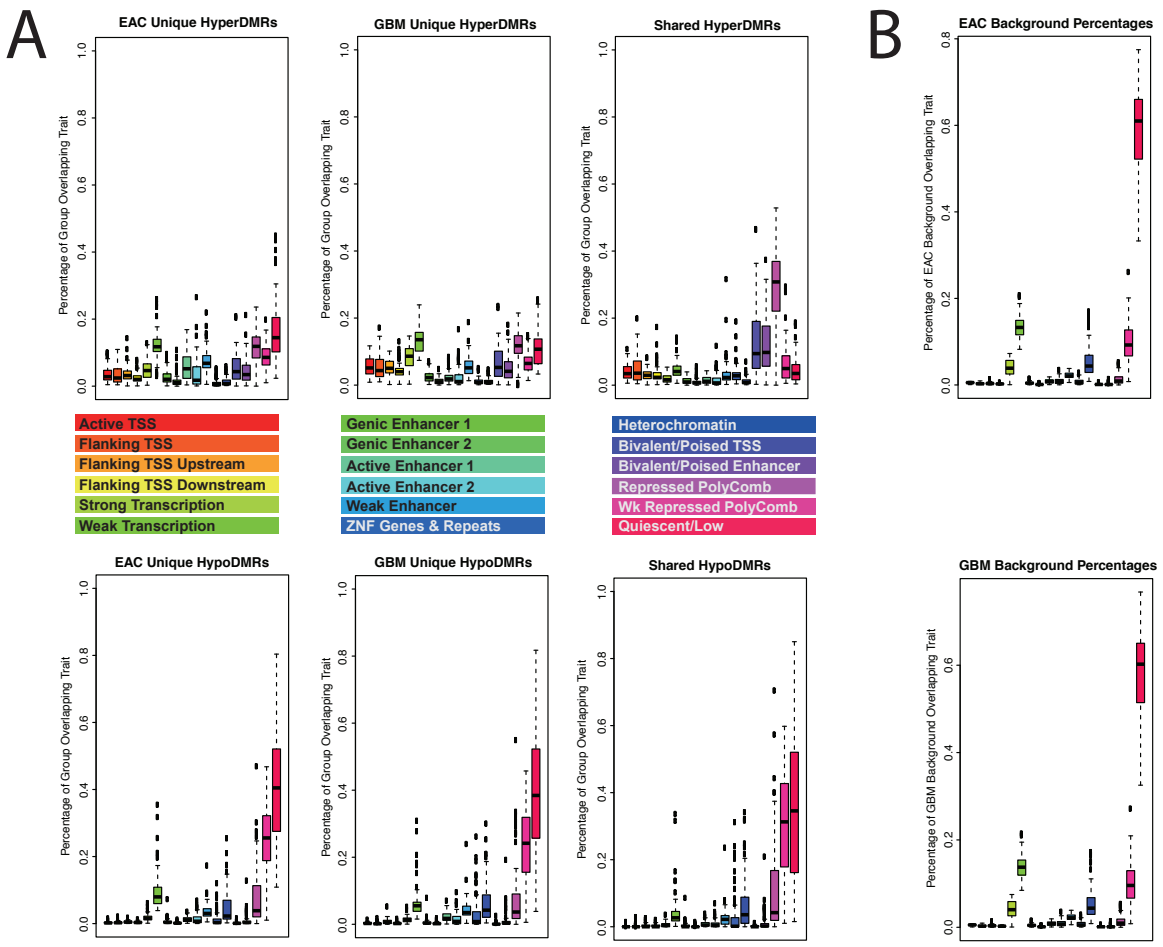


Figure 2.5. Comparative analysis of EAC and GBM DNA methylation abnormalities.
a. EAC and GBM DMR percentage overlaps with epigenetic annotations (chromHMM 18-states) across a variety of cell and tissue types (n=80, Human Roadmap Epigenome, see Appendix A.5 for a complete list). **b.** Genomic background percentage overlaps with epigenetic annotations (chromHMM 18-states) across a variety of cell and tissue types (n=80, Human Roadmap Epigenome, see Appendix A.5 for a complete list).

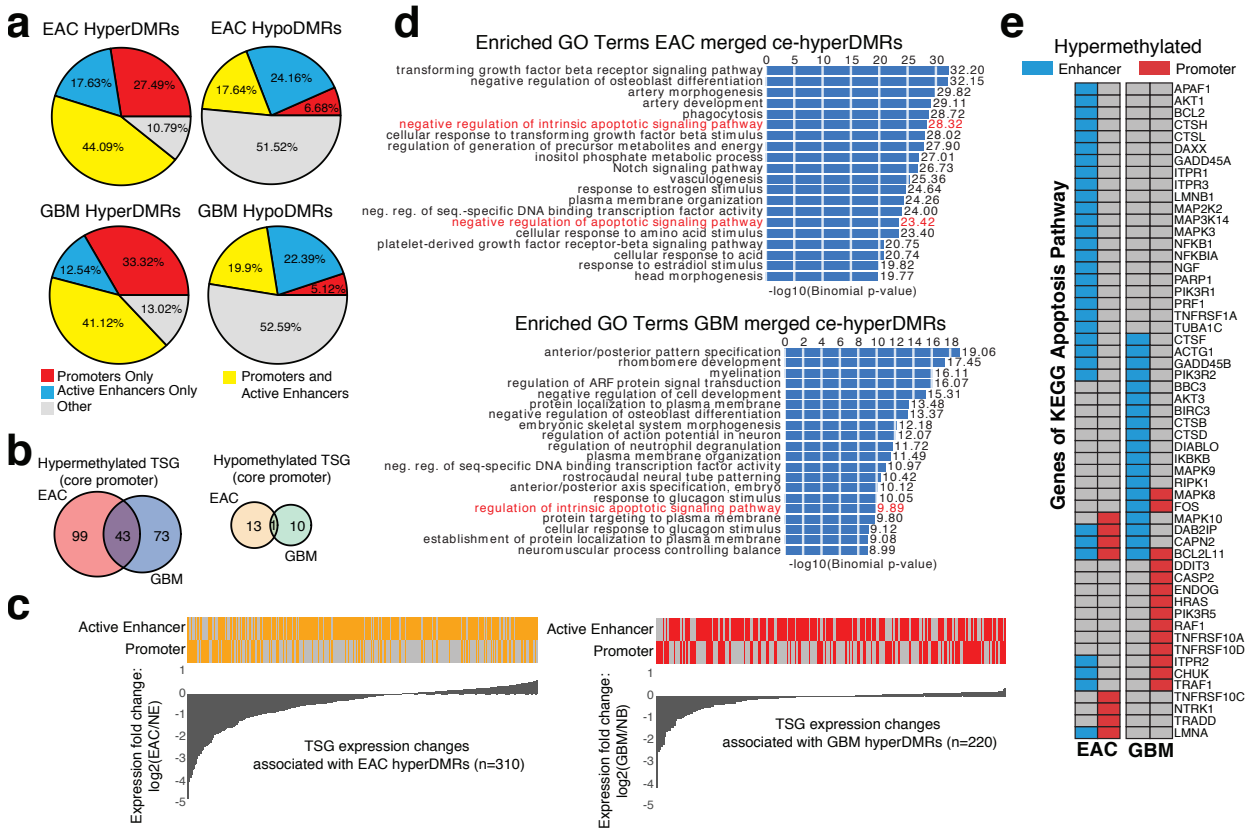


Figure 2.6. EAC and GBM DMRs show similar characteristics on a pathway level.
a. Percentage of DMRs with active regulatory annotations (promoter and/or active enhancer) in EAC and GBM. **b.** Overlap of tumor suppressor genes (TSGs) with abnormally methylated core promoters (1 kb, centered around TSS) in EAC and GBM. Left: hyperDMRs (overlap is significant: p<8.39E-14, hypergeometric test); right: hypoDMRs. **c.** TSGs with abnormally methylated promoters (1 kb, centered around the TSS) and/or active enhancers in EAC (left) and GBM (right), accompanied by fold change in expression of TSGs in EAC and GBM. NB: normal brain; NE: normal endometrium. Orange and red bars indicate the presence of a DMR within an active enhancer/promoter in EAC and GBM hyperDMRs, respectively. **d.** Top enriched GO terms (Biological Process) of ce-hyperDMRs in EAC (top) and GBM (bottom). Values represent the -log10(Binomial p-value). **e.** Hypermethylation status of apoptosis gene promoters (red) and enhancers (blue) in EAC (left) and GBM (right).

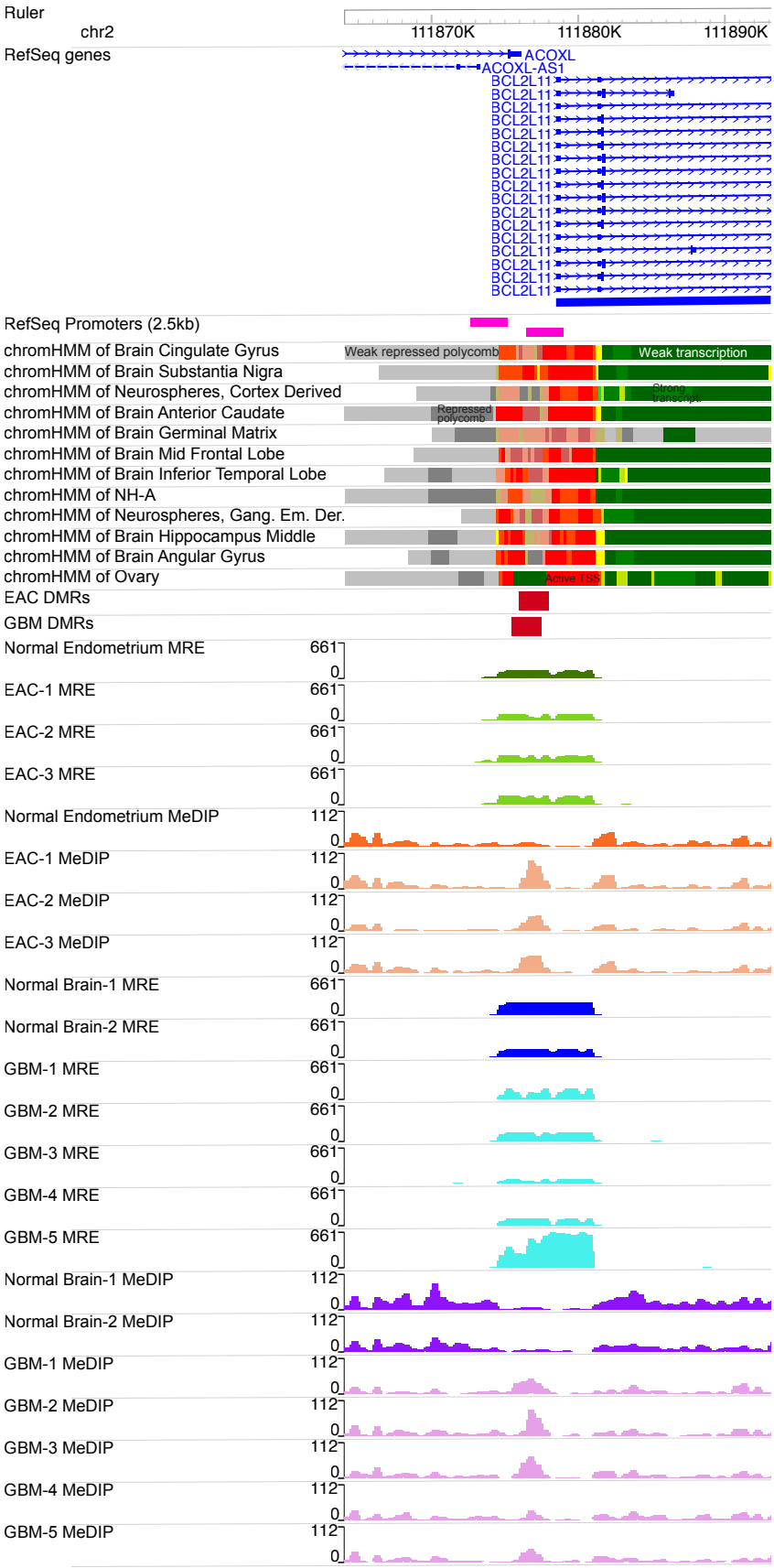

Ruler
chr2
111870K
111880K
111890K
RefSeq genes
ACOXL
ACOXL-AS1
BCL2L11
RefSeq Promoters (2.5kb)
chromHMM of Brain Cingulate Gyrus
Weak repressed polycomb
Weak transcription
chromHMM of Brain Substantia Nigra
chromHMM of Neurospheres, Cortex Derived
chromHMM of Brain Anterior Caudate
Repressed polycomb
chromHMM of Brain Germinal Matrix
chromHMM of Brain Mid Frontal Lobe
chromHMM of Brain Inferior Temporal Lobe
chromHMM of NH-A
chromHMM of Neurospheres, Gang. Em. Der.
chromHMM of Brain Hippocampus Middle
chromHMM of Brain Angular Gyrus
chromHMM of Ovary
Active TSS
EAC DMRs
GBM DMRs
Normal Endometrium MRE
EAC-1 MRE
EAC-2 MRE
EAC-3 MRE
Normal Endometrium MeDIP
EAC-1 MeDIP
EAC-2 MeDIP
EAC-3 MeDIP
Normal Brain-1 MRE
Normal Brain-2 MRE
GBM-1 MRE
GBM-2 MRE
GBM-3 MRE
GBM-4 MRE
GBM-5 MRE
Normal Brain-1 MeDIP
Normal Brain-2 MeDIP
GBM-1 MeDIP
GBM-2 MeDIP
GBM-3 MeDIP
GBM-4 MeDIP
GBM-5 MeDIP
661
112
0

Figure 2.7. Browser[74,75] view of EAC and GBM hyperDMRs overlapping *BCL2L11* regulatory regions.
First panel displays RefSeq gene locations across genomic coordinates (shown in ruler at the top), focused in on the *BCL2L11* promoter and surrounding regions. The second panel displays the locations of RefSeq promoters, defined as the regions spanning 2 kb upstream to 500 bp downstream each transcription start site (TSS). The proceeding 12 tracks display predicted chromatin states for various cell and tissue types surrounding the *BCL2L11* promoter, provided by the Roadmap Epigenomics Consortium[32]. A full catalogue of color IDs has been made available[32]. The next 2 tracks display the locations of EAC and GBM DMRs, respectively (red = hyperDMR; blue = hypoDMR). The following track displays normal endometrium MRE-seq data, and the next 3 tracks display MRE-seq data across 3 EAC samples. Similarly, next is a track displaying normal endometrium MeDIP-seq data followed by MeDIP-seq data tracks for the same 3 EAC samples. Finally, in a similar manner, 2 normal brain MRE-seq tracks are followed by 5 GBM MRE-seq tracks, and 2 normal brain MeDIP-seq tracks are followed by 5 GBM MeDIP-seq tracks. (MeDIP-seq and MRE-seq tracks depict raw counts; "Gang. Em. Der." = "Ganglion Eminence Derived").

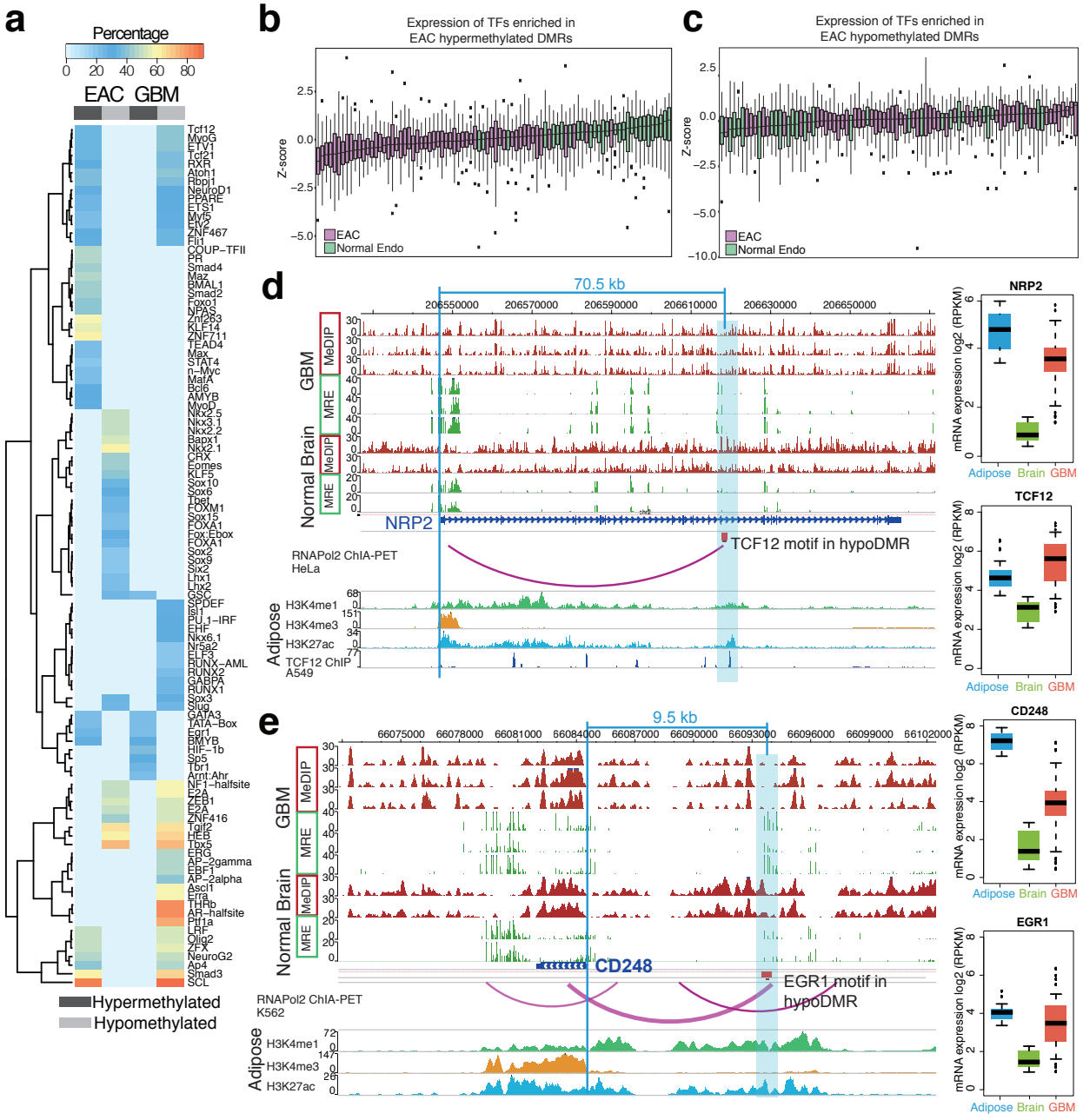


Figure 2.8. Abnormally methylated enhancer-potential regions in cancer are associated with deregulated transcription factors.

a. Percentage of EAC and GBM DMRs with enriched transcription factor-binding motifs. Only those with a *q*-value <= 0.01 and percentage >= 20% are shown. **b, c.** Normalized expression (RPKM, *z*-scores) of TFs enriched in EAC ce-hyperDMRs (**b**, number of EAC samples = 57, number of normal endometrial samples = 29, number of TFs = 30) and ce-hypoDMRs (**c**, number of EAC samples = 57, number of normal endometrial samples = 29, number of TFs = 21) in normal endometrium (green) and EAC (purple) samples. **d.** Epigenome Browser[74,75] view demonstrating a hypoDMR (blue highlighted region) across GBM samples within the *NRP2* gene. The DMR contains a TCF12-binding motif, as evidenced by a ChIP-seq peak for the TF in

A549 cells, which demonstrates interaction with the *NRP2* promoter in HeLa cells. *TCF12* is relatively unexpressed in brain (frontal cortex, $n = 28$), but more highly expressed in adipose tissue (subcutaneous, visceral, $n = 1204$) where it also bears the marks of an active enhancer (H3K4me1 and H3K27ac, lacking H3K4me3). Expression of *TCF12* increases in GBM ($n = 160$), accompanying an increase in the presumed target gene *NRP2*'s expression, as also seen in adipose tissue. MeDIP and MRE tracks depict raw read data. **e.** Epigenome Browser[74,75] view demonstrating a hypoDMR (blue highlighted region) across GBM samples upstream of the *CD248* gene. The DMR contains an EGR1-binding motif that demonstrates interaction with the *CD248* promoter in K562 cells. *EGR1* is relatively unexpressed in brain (frontal cortex, $n = 28$), but more highly expressed in adipose tissue (subcutaneous, visceral, $n = 1204$) where it also bears the marks of an active enhancer (H3K4me1 and H3K27ac, lacking H3K4me3). Expression of *EGR1* increases in GBM ($n = 160$), accompanying an increase in the presumed target gene *CD248*'s expression, as also seen in adipose tissue. MeDIP and MRE tracks depict raw read data.

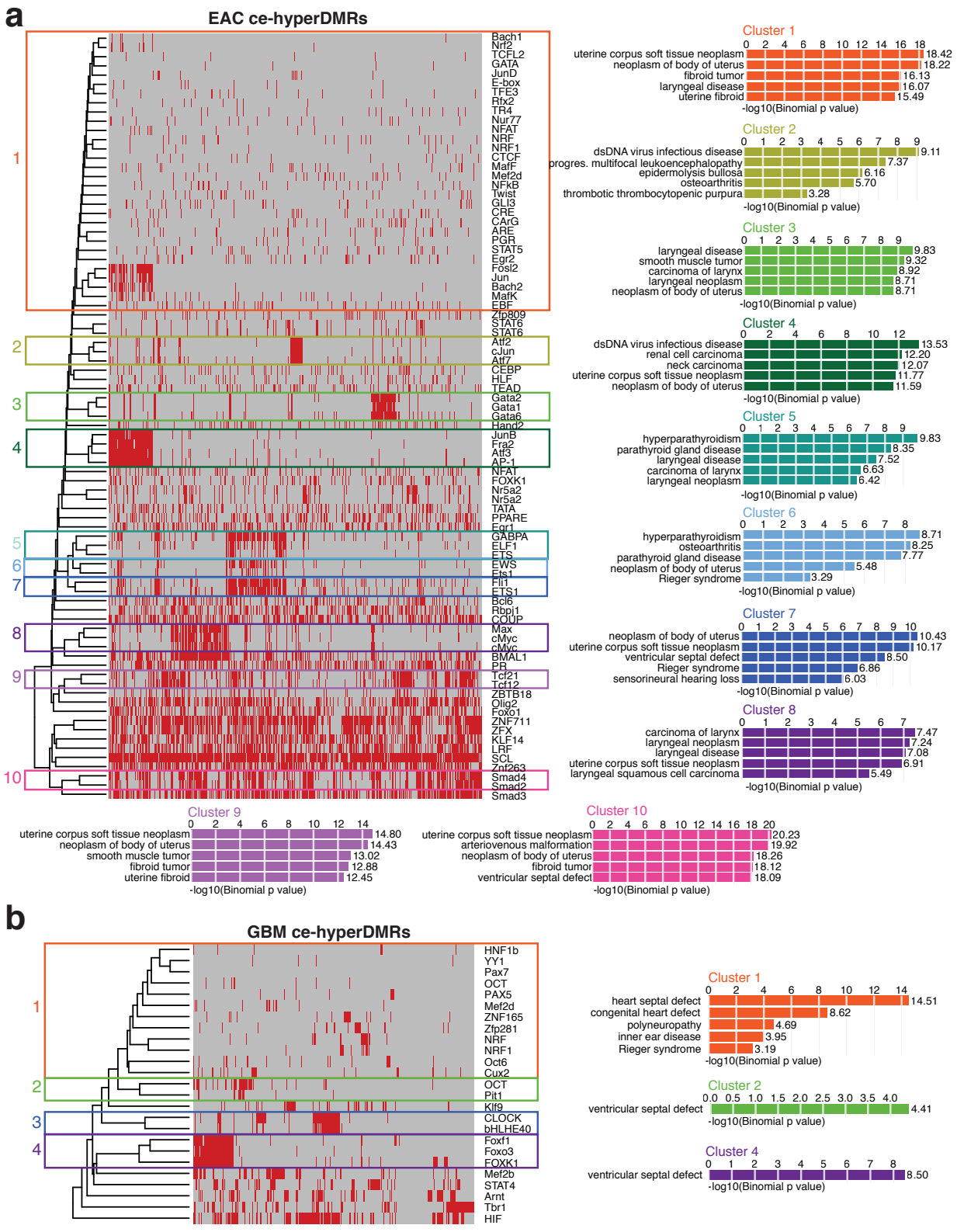


Figure 2.9. Gain of methylation over original cell-type enhancers may contribute to loss of cellular identity during cancer progression.

a. Heatmap indicating the presence of enriched TF-binding motifs in EAC ce-hyperDMRs and top GO disease ontology enrichment results for DMR clusters (red: present; gray: absent). **b.** Heatmap indicating the presence of enriched TF-binding motifs in GBM ce-hyperDMRs and top GO disease ontology enrichment result for DMR clusters (red: present; gray: absent).

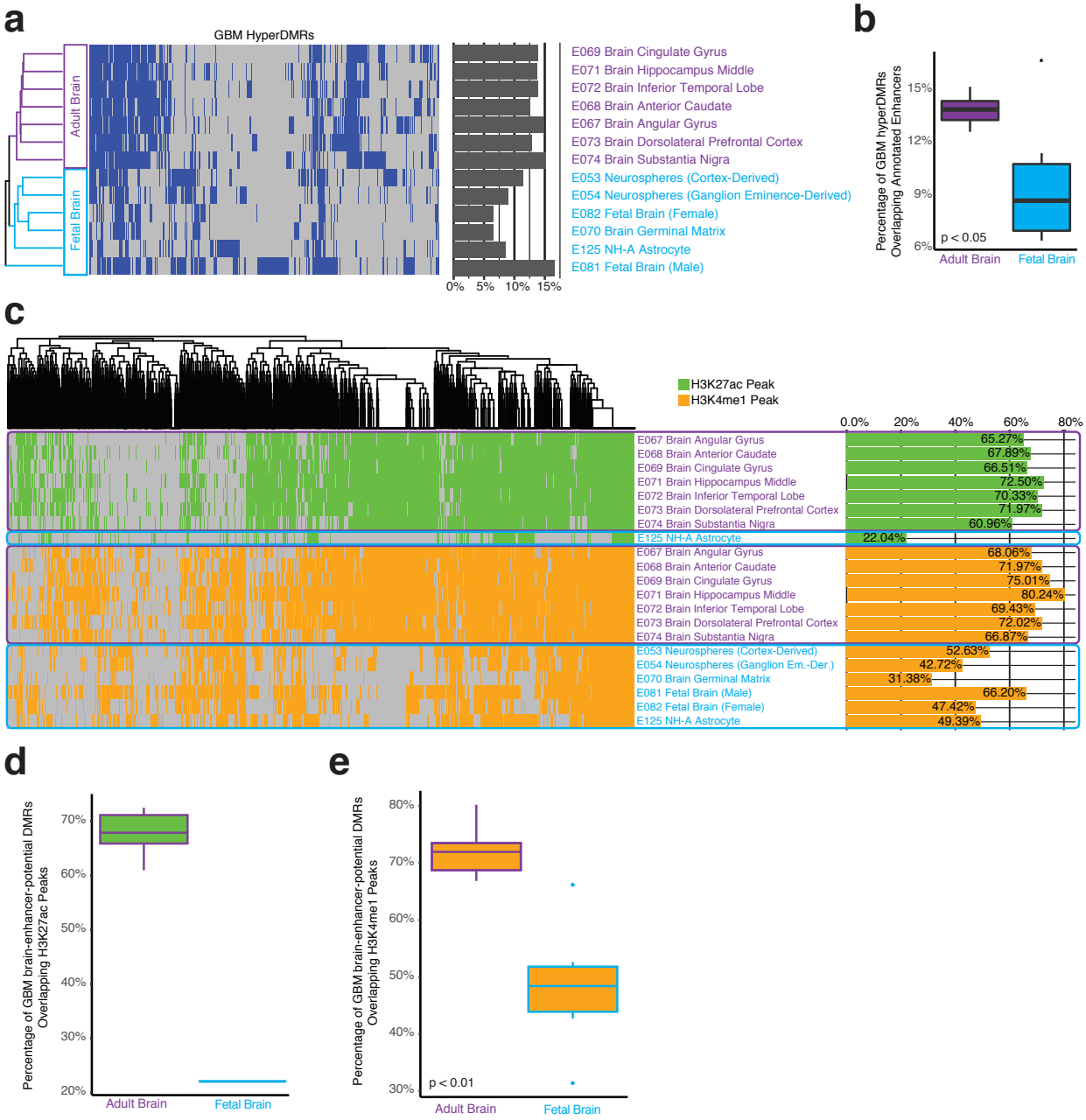


Figure 2.10. GBM hyperDMRs possess enhancer chromatin marks in adult brain tissues. **a.** Enhancer annotation (chromHMM 15-state model annotation "7_Enh"[32]) presence or absence in GBM hyperDMRs in fetal ($n = 6$, light blue) and adult brain ($n = 7$, purple) tissue/cells[32] (heatmap colors: blue: enhancer annotation present in DMR; gray: enhancer annotation absent in DMR). Bar plot indicates the fraction of all GBM hyperDMRs overlapping enhancer annotations in each cell/tissue type. **b.** Boxplot depicting the distribution of fractions of GBM hyperDMRs overlapping enhancer annotations, comparing adult ($n = 7$) and fetal ($n = 6$) brain tissues (t-test, $p < 0.05$). **c.** H3K27ac (green) and H3K4me1 (orange) peak occupancy within GBM enhancer-potential hyperDMRs (GBM hyperDMRs overlapping an enhancer annotation in at least one adult or fetal brain tissue) across fetal ($n = 1$ and 6, for H3K27ac and H3K4me1, respectively, light blue) and adult ($n = 7$ for both H3K27ac and H3K4me1, purple) brain tissues. Bar plot

indicates the fraction of all GBM enhancer-potential hyperDMRs that contained H3K27ac peaks or H3K4me1 peaks in each cell/tissue type. **d.** Boxplot depicting fractions of GBM enhancer-potential hyperDMRs overlapping H3K27ac peaks in adult brain tissues ($n = 7$, left) and horizontal line depicting the fraction of GBM enhancer-potential hyperDMRs overlapping H3K27ac peaks in fetal brain ($n = 1$, right). **e.** Comparison of the proportions of GBM enhancer-potential hyperDMRs overlapping H3K4me1 peaks in adult brain tissues ($n = 7$, left) to the proportions of GBM enhancer-potential hyperDMRs overlapping H3K4me1 peaks in fetal brain tissues ($n = 6$, right) (t-test, $p < 0.01$).

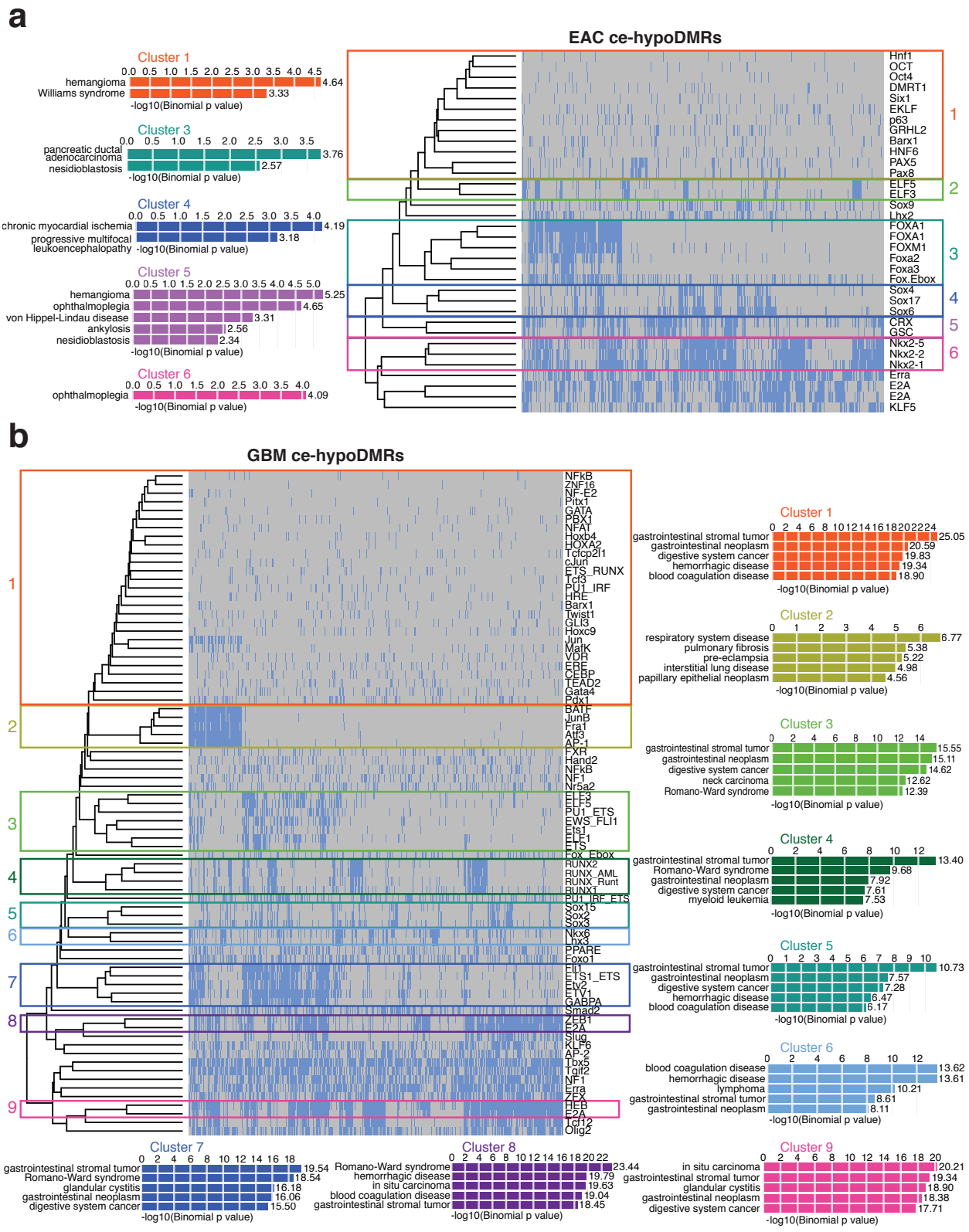


Figure 2.11. Shared loss of methylation in EAC and GBM over regions encompassing cancer-related enhancers and motifs for upregulated TFs.

a. Heatmap indicating the presence of enriched TF-binding motifs in EAC ce-hypoDMRs and top GO disease ontology enrichment results for clusters of DMRs (blue: present; gray: absent). **b.** Heatmap indicating the presence of enriched TF-binding motifs in GBM ce-hypoDMRs and top GO disease ontology enrichment results for clusters of DMRs (blue: present; gray: absent).

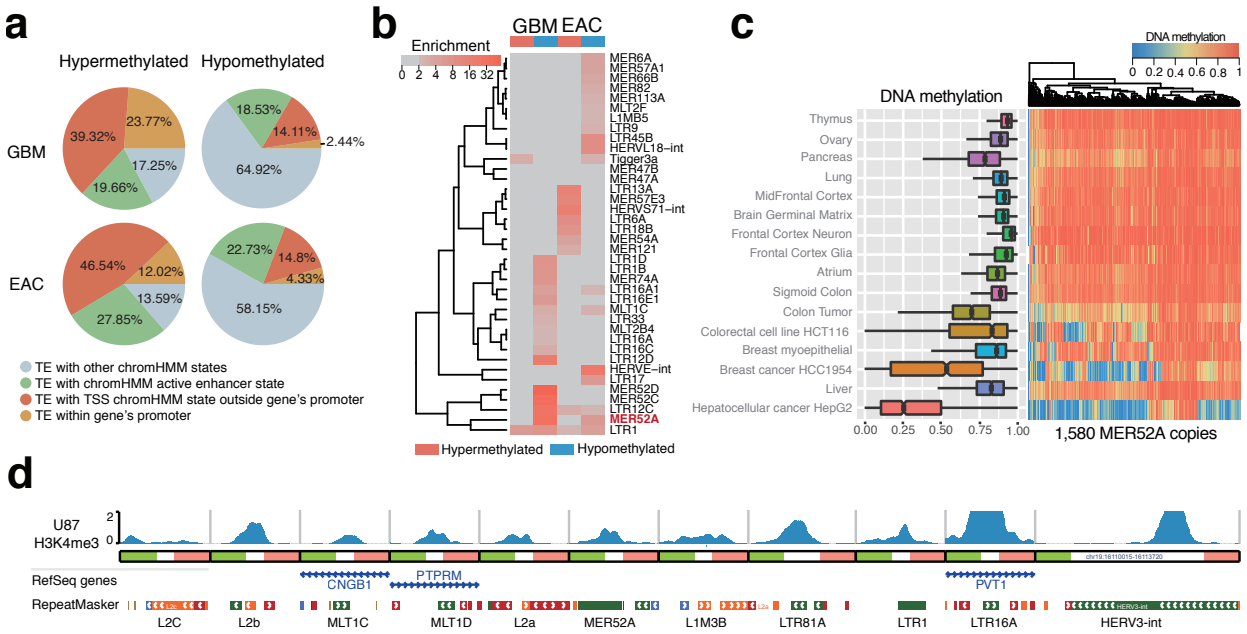


Figure 2.12. Distinct spectrum of epigenetic abnormalities within transposable elements in cancers.

a. Percentage of abnormally methylated transposable elements (TEs) within predicted regulatory regions (promoter and enhancer) in EAC and GBM. ChromHMM 18-state models[32] were used to define active enhancer states and TSS states outside genic promoters. **b.** Enrichment of abnormally methylated TEs in EAC and GBM at the subfamily level. **c.** DNA methylation level of MER52A copies across different tissues and cancer types. Left: boxplot showing DNA methylation across all MER52A copies ($n = 1580$). Right: heatmap showing DNA methylation across all MER52A copies. **d.** Epigenome Browser[74,75] view of the promoter-associated histone modification H3K4me3 in the U87 cell line (GBM cells, normalized reads per million) across 11 TEs.

2.11 Tables

Table 2.1. EZH2 binding enrichment within GBM hyperDMRs

	GBM-Unique HyperDMRs	Shared GBM/EAC HyperDMRs
Average EZH2 Signal Over DMRs	16.6406	35.8355
Average EZH2 Signal Over Background	3.51367	3.51367
Enrichment of EZH2 Signal in DMRs Over Background	4.73596	10.1989

ChIP-Seq data from ENCODE's NH-A sample (GSM1003532)

Table 2.2 Summary of DMRs containing CpG probes in the Infinium 450K and 850K platforms

	DMRs in cancer	Total	Infinium 450K		Infinium 850K	
			DMRs Covered	Percentage	DMRs Covered	Percentage
GBM	Hypermethylated	10178	7013	68.90%	7595	74.62%
	Hypomethylated	4494	829	18.45%	1483	33.00%
	Total	14672	7842	53.45%	9078	61.87%
EAC	Hypermethylated	18278	10706	58.57%	12423	67.97%
	Hypomethylated	8712	2102	24.13%	3272	37.56%
	Total	26990	12808	47.45%	15695	58.15%

2.12 References

1. Lander S, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(February):860-921. doi:10.1038/35057062
2. Hoadley KA, Yau C, Wolf DM, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*. 2014;158(4):929-944. doi:10.1016/j.cell.2014.06.049
3. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793):94-101. doi:10.1038/s41586-020-1943-3
4. Campbell PJ, Getz G, Korbel JO, et al. Pan-cancer analysis of whole genomes. *Nature*. 2020;578(7793):82-93. doi:10.1038/s41586-020-1969-6
5. Hoadley KA, Yau C, Hinoue T, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell*. 2018;173(2):291-304.e6. doi:10.1016/j.cell.2018.03.022
6. Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. *Nature*. 2013;502:333-339. doi:10.1038/nature12634
7. Watson IR, Takahashi K, Futreal PA, Chin L. Emerging patterns of somatic mutations in cancer. *Nat Rev Genet*. 2013;14(10):703-718. doi:10.1038/nrg3539
8. Hudson TJ, Anderson W, Aretz A, et al. International network of cancer genome projects. *Nature*. 2010;464(7291):993-998. doi:10.1038/nature08987
9. Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. *Science (80-)*. 2017;357(6348). doi:10.1126/science.aal2380
10. Saghafinia S, Mina M, Riggi N, Hanahan D, Ciriello G. Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors. *Cell Rep*. 2018;25(4):1066-1080.e8. doi:10.1016/j.celrep.2018.09.082
11. McGhee JD, Ginder GD. Specific DNA methylation sites in the vicinity of the chicken β-globin genes. *Nature*. 1979;208:419-420. doi:https://doi.org/10.1038/280419a0
12. Razin A, Howard C. DNA methylation and gene expression. *Microbiol Rev*. 1991;55(3):451-458. doi:https://doi.org/10.1128/mr.55.3.451-458.1991
13. Ball MP, Li JB, Gao Y, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol*. 2009;27(4):361-368. doi:10.1038/nbt.1533
14. Shen H, Laird PW. Interplay between the cancer genome and epigenome. *Cell*. 2013;153(1):38-55. doi:10.1016/j.cell.2013.03.008
15. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*. 1983;301:89-92. doi:https://doi.org/10.1038/301089a0
16. Baylin SB, Herman JG. DNA hypermethylation in tumorigenesis: Epigenetics joins genetics. *Trends Genet*. 2000;16(4):168-174. doi:10.1016/S0168-9525(99)01971-X
17. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet*. 2002;3:415-428. doi:10.1038/nrg816
18. Baylin SB, Esteller M, Rountree M, Bachman K, Schuebel K, Herman J. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum Mol Genet*. 2001;10(7):687-692. doi:10.1093/hmg/10.7.687
19. Esteller M. Epigenetic gene silencing in cancer: The DNA hypermethylome. *Hum Mol Genet*. 2007;16(R1):50-59. doi:10.1093/hmg/ddm018

20. Herman JG, Latif F, Weng Y, et al. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A*. 1994;91(21):9700-9704. doi:10.1073/pnas.91.21.9700
21. Dobrovic A, Simpfendorfer D. Methylation of the BRCA1 gene in sporadic breast cancer. *Cancer Res*. 1997;57(16):3347-3350.
22. Domann FE, Rice JC, Hendrix MJC, Futscher BW. Epigenetic silencing of maspin gene expression in human breast cancers. *Int J Cancer*. 2000;85(6):805-810. doi:10.1002/(SICI)1097-0215(20000315)85:6<805::AID-IJC12>3.0.CO;2-5
23. Hovestadt V, Jones DTW, Picelli S, et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature*. 2014;510:537-541. doi:10.1038/nature13268
24. Yang X, Gao L, Zhang S. Comparative pan-cancer DNA methylation analysis reveals cancer common and specific patterns. *Brief Bioinform*. 2017;18(5):761-773. doi:10.1093/bib/bbw063
25. Zhang B, Xing XY, Li J, et al. Comparative DNA methylome analysis of endometrial carcinoma reveals complex and distinct deregulation of cancer promoters and enhancers. *BMC Genomics*. 2014;15(868). doi:10.1186/1471-2164-15-868
26. Nagarajan RP, Zhang B, Bell RJA, et al. Recurrent epimutations activate gene body promoters in primary glioblastoma. *Genome Res*. 2014;24(5):761-774. doi:10.1101/gr.164707.113
27. Chong A, Teo JX, Ban KHK. Distinct epigenetic signatures elucidate enhancer-gene relationships that delineate CIMP and non-CIMP colorectal cancers. *Oncotarget*. 2016;7(19):28027-28039. doi:10.18632/oncotarget.8473
28. Akhtar-Zaidi B, Cowper-Sallari R, Corradin O, et al. Epigenomic enhancer profiling defines a signature of colon cancer. *Science (80-)*. 2012;336(6082):736-739. doi:10.1126/science.1217277
29. Kurdistani SK. Enhancer dysfunction: How the main regulators of gene expression contribute to cancer. *Genome Biol*. 2012;13(5):4-6. doi:10.1186/gb-2012-13-5-156
30. Heyn H, Vidal E, Ferreira HJ, et al. Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. *Genome Biol*. 2016;17(11). doi:10.1186/s13059-016-0879-2
31. Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol*. 2013;14(3). doi:10.1186/gb-2013-14-3-r21
32. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317-330. doi:10.1038/nature14248
33. Ohgaki H, Kleihues P. The definition of primary and secondary glioblastoma. *Clin Cancer Res*. 2013;19(4):764-772. doi:10.1158/1078-0432.CCR-12-3002
34. Ostrom QT, Gittleman H, Liao P, et al. CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014. *Neuro Oncol*. 2017;19(5):v1-v88. doi:10.1093/neuonc/nox158
35. American Cancer Society. Cancer Facts & Figures 2019. *Atlanta Am Cancer Soc*. 2019.
36. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7-34. doi:10.3322/caac.21551
37. Colombo N, Creutzberg C, Amant F, et al. ESMO-ESGO-ESTRO consensus conference on endometrial cancer: Diagnosis, treatment and follow-up. *Ann Oncol*. 2016;27(1):16-41. doi:10.1093/annonc/mdv484

38. The Cancer Genome Atlas Research Network, Levine DA, Getz G, et al. Integrated genomic characterization of endometrial carcinoma. *Nature*. 2013;497:67-73. doi:10.1038/nature12113
39. Amant F, Moerman P, Neven P, Timmerman D, Van Limbergen E, Vergote I. Endometrial cancer. *Lancet*. 2005;366(9484):491-505. doi:10.1016/S0140-6736(05)67063-8
40. Li D, Zhang B, Xing X, Wang T. Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. *Methods*. 2015;72(C):29-40. doi:10.1016/j.ymeth.2014.10.032
41. Zhang B, Zhou Y, Lin N, et al. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Res*. 2013;23(9):1522-1540. doi:10.1101/gr.156539.113
42. Andersson R, Gebhard C, Miguel-Escalada I, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014;507:455-461. doi:10.1038/nature12787
43. Lizio M, Harshbarger J, Shimoji H, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol*. 2015;16(22). doi:10.1186/s13059-014-0560-6
44. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser - A database of tissue-specific human enhancers. *Nucleic Acids Res*. 2007;35(SUPPL. 1):88-92. doi:10.1093/nar/gkl822
45. Khan A, Zhang X. DbSUPER: A database of Super-enhancers in mouse and human genome. *Nucleic Acids Res*. 2016;44(D1):D164-D171. doi:10.1093/nar/gkv1002
46. Ernst J, Kellis M. ChromHMM: Automating chromatin-state discovery and characterization. *Nat Methods*. 2012;9:215-216. doi:10.1038/nmeth.1906
47. Dunham I, Kundaje A, Aldred SF, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57-74. doi:10.1038/nature11247
48. Esteller M, Herman JG. Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. *J Pathol*. 2002;196:1-7. doi:10.1002/path.1024
49. Baylin SB. DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol*. 2005;2:S4-11. doi:10.1038/ncponc0354
50. Herman JG, Baylin SB. Gene Silencing in Cancer in Association with Promoter Hypermethylation. *N Engl J Med*. 2003;349(21):2042-2054. doi:10.1056/nejmra023075
51. Soengas MS, Alarcón RM, Yoshida H, et al. Apaf-1 and caspase-9 in p53-dependent apoptosis and tumor inhibition. *Science (80-)*. 1999;284(5411):156-159. doi:10.1126/science.284.5411.156
52. Soengas MS, Capodieci P, Polsky D, et al. Inactivation of the apoptosis effector Apaf-1 in malignant melanoma. *Nature*. 2001;409:207-211. doi:https://doi.org/10.1038/35051606
53. Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA methylation. *Nat Rev Genet*. 2016;17(9):551-565. doi:10.1038/nrg.2016.83
54. Schübeler D. Function and information content of DNA methylation. *Nature*. 2015;517:321-326. doi:10.1038/nature14192
55. Dutta S, Roy S, Polavaram NS, et al. Neuropilin-2 regulates endosome maturation and EGFR trafficking to support cancer cell pathobiology. *Cancer Res*. 2016;76(2):418-428. doi:10.1158/0008-5472.CAN-15-1488
56. Simonavicius N, Robertson D, Bax DA, Jones C, Huijbers IJ, Isacke CM. Endosialin (CD248) is a marker of tumor-associated pericytes in high-grade glioma. *Mod Pathol*. 2008;21(3):308-315. doi:10.1038/modpathol.3801006

57. Di Benedetto P, Liakouli V, Ruscitti P, et al. Blocking CD248 molecules in perivascular stromal cells of patients with systemic sclerosis strongly inhibits their differentiation toward myofibroblasts and proliferation: A new potential target for antifibrotic therapy. *Arthritis Res Ther*. 2018;20(223). doi:10.1186/s13075-018-1719-4
58. Heinz S, Benner C, Spann N, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell*. 2010;38(4):576-589. doi:10.1016/j.molcel.2010.05.004
59. Burns KH. Transposable elements in cancer. *Nat Rev Cancer*. 2017;17(7):415-424. doi:10.1038/nrc.2017.35
60. Jang HS, Shah NM, Du AY, et al. Transposable elements drive widespread expression of oncogenes in human cancers. *Nat Genet*. 2019;51(4):611-617. doi:10.1038/s41588-019-0373-3
61. Miao B, Fu S, Lyu C, Gontarz P, Wang T, Zhang B. Tissue-specific usage of transposable element-derived promoters in mouse development. *Genome Biol*. 2020;21(255). doi:10.1186/s13059-020-02164-3
62. Sandoval J, Heyn HA, Moran S, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics*. 2011;6(6):692-702. doi:10.4161/epi.6.6.16196
63. Bibikova M, Barnes B, Tsan C, et al. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011;98(4):288-295. doi:10.1016/j.ygeno.2011.07.007
64. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics*. 2016;8(3):389-399. doi:10.2217/epi.15.114
65. Sur I, Taipale J. The role of enhancers in cancer. *Nat Rev Cancer*. 2016;16(8):483-493. doi:10.1038/nrc.2016.62
66. Maunakea AK, Nagarajan RP, Bilenky M, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*. 2010;466(7303):253-257. doi:10.1038/nature09165
67. Roy N, Hebrok M. Regulation of Cellular Identity in Cancer. *Dev Cell*. 2015;35(6):674-684. doi:10.1016/j.devcel.2015.12.001
68. Xie M, Hong C, Zhang B, et al. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet*. 2013;45(7):836-841. doi:10.1038/ng.2649
69. Haeussler M, Zweig AS, Tyner C, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res*. 2019;47(D1):D853-D858. doi:10.1093/nar/gky1095
70. Zhao M, Sun J, Zhao Z. TSGene: A web resource for tumor suppressor genes. *Nucleic Acids Res*. 2013;41(D1):970-976. doi:10.1093/nar/gks937
71. Zhao M, Kim P, Mitra R, Zhao J, Zhao Z. TSGene 2.0: An updated literature-based knowledgebase for Tumor Suppressor Genes. *Nucleic Acids Res*. 2016;44(D1):D1023-D1031. doi:10.1093/nar/gkv1268
72. McLean CY, Bristor D, Hiller M, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010;28(5):495-501. doi:https://doi.org/10.1038/nbt.1630
73. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28(1):27-30. doi:10.1093/nar/28.1.27
74. Zhou X, Maricque B, Xie M, et al. The human epigenome browser at Washington

University. *Nat Methods*. 2011;8(12):989-990. doi:10.1038/nmeth.1772
75. Li D, Hsu S, Purushotham D, Sears RL, Wang T. WashU Epigenome Browser update 2019. *Nucleic Acids Res*. 2019;47(W1):W158-W165. doi:10.1093/nar/gkz348

Chapter 3: Multiomic comparison of paired primary and brain metastatic non-small cell lung cancer suggests tumor cell reprogramming toward a glial cell phenotype

The contents of this chapter have been adapted from the submitted manuscript:

> "Karlow, J.A.*, Devarakonda, S.*, Sankararaman, S., et al. Multiomic comparison of paired primary and brain metastatic non-small cell lung cancer suggests tumor cell reprogramming towards a glial cell phenotype." *These authors contributed equally to this work.

3.1 Abstract

Brain metastasis is a common cause of mortality in patients with non-small cell lung cancer (NSCLC); a better understanding of recurrent molecular events that promote this process could produce improved diagnostics and targeted therapeutics to mitigate tumor spread to the brain and improve overall patient survival. Patient-matched primary and brain metastatic tumors from NSCLC patients were evaluated by comprehensive exome (*n*=30), transcriptome (*n*=30), and DNA methylome (*n*=12) analyses. Metastatic brain tumors showed few recurrent or novel genomic alterations relative to their primary tumor counterparts. However, transcriptome analyses of brain metastases demonstrated specific and recurrent patterns of "oligodendrocyte-like" gene expression which could not be explained by stromal contamination. Comparative methylation analyses also demonstrated recurrent epigenetic changes in brain-specific active enhancers and brain super enhancers, correlating with expression changes of nearby genes

regulating glial progenitor cell differentiation. This analysis suggests that despite an apparent absence of 'brain metastasis' driver mutations, brain metastatic NSCLC tumor cell populations may assume a glial-like phenotype through epigenomic reprogramming, possibly creating a growth advantage by co-opting mechanisms for survival in their metastatic niche.

3.2 Introduction

Despite advances in systemic therapy, brain metastasis remains a significant cause of mortality in non-small cell lung cancer (NSCLC) patients. Nearly 26% of patients with lung cancer at diagnosis develop brain metastases[1]. Numerous studies have characterized the genomic landscape of primary NSCLC tumors, which is both complex and varies widely across patients[2–4]. The molecular evolution of NSCLC cells during their progression to metastatic disease and more specifically, to brain metastases, is equally varied and has only been evaluated in relatively few individual cases[5–9]. Although many canonical biological pathways and key genetic components associated with metastasis have been well defined[10], these advances have not yet led to robust clinical biomarkers for predicting or mitigating brain-specific (CNS) metastatic behavior of primary NSCLC.

Since NSCLC is typically characterized by a high mutation burden, genomic complexity, and substantial inter-patient heterogeneity, we hypothesized that direct, pairwise comparisons of the genomic, transcriptomic, and epigenomic profiles of primary NSCLC and metastatic brain tumors of individual patients would provide valuable insight into molecular events that promote metastasis and explain CNS-specific organotropism in lung cancer. To this end, matched normal lung, primary lung tumor, and surgically resected brain metastasis (metastasis) specimens from NSCLC patients were analyzed by whole-exome sequencing (WES, n=30) and DNA methylation analysis (MeDIP-seq and MRE-seq, n=12). Transcriptome analyses were performed on paired

primary and metastasis samples (n=30), and WES was performed on matched normal and metastasis samples from an additional 3 patients. Clinical and histological characteristics of patients are summarized in Table 3.1.

3.3 Results

Overall, primary-metastasis pairs demonstrated a significant difference in tumor mutation burden (TMB) (median TMB: 5.6 vs 6.2 mutations/MB, respectively, p=0.008728, paired Wilcox test). However, the fraction of genome altered (FGA) was not significantly different between primary-metastasis pairs (median FGA: 3.5% vs 4.8%, p=0.3931, paired Wilcox test, Figure 3.1A and Figure 3.1B, Table 3.1). Non-synonymous mutation signature patterns were analyzed for 30 primary-metastasis pairs using whole-exome data. Signature patterns in each sample were decomposed into single-base-substitution (SBS) COSMIC signatures[11]. Hierarchical clustering based on similarity of mutation profiles demonstrated 12 samples clustering together in primary-metastasis pairs (patient IDs: 6, 9, 11, 17, 18, 19, 24, 30, 32, 33, 34, 50), suggesting a common evolutionary mechanism driving primary and metastatic tumor growth (Figure 3.1C). Tobacco smoke-associated SBS signature, SBS4, was dominant in 61.6% (37/60) of samples, concordant with self-reported smoking status. Clock-like signatures, SBS1 and SBS5, were dominant in 33.33% (20/60) of samples. As expected, cosine-similarity metrics for other SBS signatures, including those attributed to common environmental mutagens[12], distinguished tumors from patients with no smoking history from those associated with mutagens present in tobacco smoke (Figure 3.2).

Because cancer subclones at metastatic and primary sites evolve independently, we hypothesized that comparing the genomic similarity between metastases and primary tumors would offer insight into the evolutionary timing of metastatic dissemination. The ratio of somatic

mutations shared between each primary-metastasis pair to the total number of mutations observed in the pair (Jaccard score) varied widely (0.03-0.90, median: 0.34), suggesting that the timing of metastatic dissemination in our patient cohort was highly variable (Figure 3.1A and Figure 3.1B, Table 3.1). Although these analyses were limited by the small sample size, Jaccard scores also did not vary significantly between known synchronous (n=5) and metachronous (n=9) cases (p=0.2398, Wilcoxon test). In fact, we observed a high degree of Jaccard similarity (0.73 and 0.59) in two primary-metastasis pairs that surprisingly demonstrated different histology at primary and metastatic sites (adenocarcinoma metastasis with squamous cell carcinoma primary (patient ID: 46) and large cell neuroendocrine metastasis with adenocarcinoma primary (patient ID: 16)) (Table 3.1).

Both primary tumors and their corresponding brain metastases demonstrated common mutations, deletions, and amplifications in canonical "driver genes". In primary tumors, *TP53* and *STK11* were the only significantly recurrent mutated genes while mutations in *TP53*, *STK11*, *POLDIP2*, *KRAS*, *KEAP1*, and *DCDC1* were significantly recurrent in metastatic lesions (q<0.05) (Appendix B.1). Reasoning that mutations promoting a brain-specific metastatic phenotype may be clonally enriched in brain metastases relative to primary tumors, we also leveraged our pairwise comparison to identify genes that demonstrated recurrent "metastasis-enriched variants" (MEVs) - variants that were undetectable or at very low (<10%) variant allele frequency (VAF) in primary tumors with coverage of greater than 30x while also present at greater than 30% VAF in the corresponding brain metastasis. The average VAF of MEVs was 2.4% in primary tumors and 46.7% in corresponding metastases. Histological review and molecular assessment (see below) were utilized to discount the possibility that VAF differences were due only to differences in tumor cell purity. We observed recurrent MEVs in 75 genes,

including *TP53* (*n*=4), *LRP1B* (*n*=4), *CDKN2A* (*n*=3), *STK11* (*n*=2), and *KEAP1* (*n*=2) (Figure 3.3A, Figure 3.3B and Figure 3.4). *TP53* alterations were found more frequently in metastatic lesions (*n*=15) compared to their primary tumor progenitors (*n*=10), suggesting that *TP53* could be an important determinant of progression to CNS metastasis (Figure 3.3A and Figure 3.3B). In addition, three patients (IDs: 17, 19, 20) also exhibited MEVs in *TET1*, a DNA demethylase enzyme, previously implicated in contributing to metastatic dissemination[13]. Although few singular genes demonstrated recurrent MEVs, biological pathway analyses demonstrated significantly recurrent MEVs present in genes implicated in processes of both focal adhesion and extracellular matrix receptor interactions (Appendix B.2).

A comparative analysis of copy number alterations (CNAs) demonstrated metastasis-specific recurrent losses in chromosomal regions containing tumor suppressor genes such as *PTEN*, *TP53*, and *GSTT1* (Figure 3.3C). Recurrent metastasis-specific losses at the gene level were examined by determining genes within recurrent chromosomal losses that also displayed a loss of expression in metastasis combined with genes containing MEVs. Leveraging both deletion and mutation information identified several genes with a metastasis-specific losses, such as the tumor suppressor *SAMD9L* and the immune-related gene *KIR3DL1* (Figure 3.3D). Recurrent metastasis-enriched gains were also observed in regions present on the long arms of several chromosomes (Figure 3.3E), including chromosome 3q26.33 which harbors *SOX2*, a gene frequently overexpressed in brain metastatic lesions (see below) and involved in CNS glial cell differentiation[14]. In addition to *SOX2*, genes with increased expression in metastases that fell within recurrently amplified chromosomal regions included those with various developmental and neuronal functions, such as *ZAR1*, *MARCH6*, and *ECE2*.

In contrast to the intra-patient similarity but inter-patient heterogeneity between genomic features of primary and brain metastatic tumors, we observed several consistent and programmatic differences in gene expression across patients (Figure 3.5A, Appendix B.3). Globally, brain metastases were relatively devoid of gene transcripts associated with immune and inflammatory responses (q=1.61E-10 and q=2.23E-10, respectively, BH-corrected) and demonstrated enriched expression associated with central nervous system development (q=0.01, BH-corrected) (Figure 3.5A and Figure 3.5B, Appendix B.3). At the level of individual primary-metastasis pairs, single sample gene set enrichment analysis (ssGSEA), utilizing previously published expression signatures (Appendix B.4), demonstrated significant differences in canonical patterns of tissue-specific gene expression (Figure 3.5C).[15] While ssGSEA enrichment scores for lung-specific transcripts were not significantly different between paired primary and brain metastasis samples (Figure 3.6, Appendix B.3), metastatic specimens demonstrated significant enrichment scores for genes expressed specifically in the brain (p=0.0007) (Figure 3.6, Appendix B.3). This finding could not be attributed simply to brain stromal contamination in metastatic specimens, since a similar comparison between genome-based estimates of tumor purity for each metastatic specimen with brain-specific ssGSEA enrichment scores demonstrated no correlation (p=0.44, linear regression) (Figure 3.5D). Moreover, in eight patients, we were able to identify incidental tumor-specific variants in genes that comprised the 'glial gene signature' (e.g. *APP*, *MBP*, *TAGLN3*, *PRRT2*, *SCG2*) and these same variants were detected in expressed RNA sequence reads from the corresponding brain-metastatic lesion, further suggesting that the expression signature emanated from tumor cells rather than contaminating brain parenchyma. Because of the relative *SOX2* amplification seen in 15.8% of brain metastases (with both exome-seq and RNA-seq data), the known role of *SOX2* in oligodendrocyte

differentiation and tissue plasticity[14], and the enrichment of oligodendrocyte-related gene expression seen in these same samples, we evaluated the enrichment scores for a *SOX2*-driven expression program and also found it to be significantly enriched in metastatic lesions as compared to matched primary tumors (p<0.0001) (Figure 3.5C). Since many patients did not demonstrate obvious *SOX2* gene amplification in the metastatic samples despite showing an upregulation of ssGSEA scores for *SOX2*-driven gene expression, we hypothesized that *SOX2* expression programs could be up-regulated by epigenetic activation of *SOX2*.

In fact, genomic analyses revealed increased VAFs in several classes of epigenetic regulator proteins (Figure 3.7), including 2 predicted deleterious variants within the SAM1 domain of *PHC3*. Therefore, we performed comparative DNA methylation analyses on a subset of 12 patients to determine if epigenomic changes may be contributing to the deregulation of *SOX2* and other neuronal expression programs. Differentially methylated regions (DMRs) were identified between primary tumor and metastasis in each patient, but like genomic alterations, demonstrated considerable inter-patient variability (Figure 3.8A). However, metastasis samples demonstrated a significant increase of methylation within DNA methylation valleys (DMVs)- i.e. genomic regions devoid of methylation in non-malignant lung tissue (Figure 3.8B). These DMVs encompassed 735 unique autosomal genes, many related to CNS development, suggesting that methylation changes within these regions may contribute to transcriptional changes in developmental programs. Recurrently hypermethylated sequence motifs within these DMVs also corresponded to binding sites for several transcription factors (TFs) related to development (Appendix B.5), and several of these TFs, such as those belonging to the KLF and HOX families, are known to preferentially bind methylated CpG recognition sites.[16] Consistent with this observation, there was a general increase in expression of DMV resident genes proximal to these

recurrently hypermethylated regulatory regions (Figure 3.8C and Figure 3.8D). Recurrent metastasis enriched hypermethylation in several individual and biologically relevant genes outside of these DMVs was also observed. For example, *TBX4*, a gene previously implicated in cancer[17], demonstrated a recurrent gain of promoter methylation in 6 of 12 patient metastases analyzed, accompanied by a significant loss of expression in metastasis relative to primary tumors (Figure 3.8C, Figure 3.8D).

Since it is well known that enhancers often regulate cell type-specific transcriptional programs, we also evaluated whether methylation changes within enhancers contribute to metastasis-specific changes in transcriptional programs. Although there was no overall difference in enhancer methylation between primary and metastatic tumors (Figure 3.8B), brain-specific active enhancers defined by the Epigenome Roadmap (3.4 Methods) revealed a significant decrease in average methylation in brain metastases (Figure 3.8B). Enhancers with the most dramatic loss of methylation in metastases were proximal to brain-related genes that exhibited concordant expression increases, such as *MBP* (Figure 3.8E). Similarly, while there was no overall difference in 'super enhancer' methylation between primary and metastatic tumors (Figure 3.8B), specific patterns of increased methylation in super enhancers associated with immune-related genes, and decreased methylation in those associated with CNS genes (Figure 3.8G), consistent with transcriptome analyses, was observed.

Metastatic dissemination is a complex, multi-step process that involves tumor intrinsic and microenvironment changes at the primary and metastatic sites. We sought to identify recurrent genomic alterations that were "drivers" of the brain metastatic phenotype in patients with NSCLC, but found that, except in rare instances, primary and brain metastatic NSCLC were both surprisingly similar for a given patient and diverse across patients, regardless of clinical

presentation of disease. As previously noted, an enrichment of mutations in *KRAS*[7], *TP53*[7], *CDKN2A*[5,6], and *KEAP1*[7] was observed in metastatic tumors relative to their primary tumor progenitors. *KEAP1*, a commonly altered tumor suppressor, is known to suppress migration and invasion in lung cancer cell lines through *NRF2* signaling[18]. Nonetheless, an otherwise high degree of genomic similarity between primary and metastatic lesions, a common finding in many solid tumors, may limit the diagnostic utility of genomic sequencing alone for predicting brain metastatic progression.

The multi-omic analyses in this study suggest that epigenomic-based tumor cell reprogramming to co-opt mechanisms of growth and survival in the CNS may be a more important determinant of the brain metastatic phenotype in NSCLC. We observed both relative amplification and a corresponding enrichment in *SOX2* gene expression signatures in metastatic tumors. *SOX2* is a known oncogene in lung cancer but also plays an important role in CNS differentiation. Given that NSCLC brain metastases also demonstrated a significant enrichment for neuronal expression signatures, independent of any contribution from stromal contamination (Figure 3.5D), it is possible that *SOX2* facilitates CNS colonization by promoting plasticity and de-differentiation along a neuronal lineage in lung cancer cells. Similar findings have been reported in several mouse xenograft studies[19,20]. The observed gain in methylation over brain-specific active enhancers and corresponding changes in expression of proximal CNS-related genes suggests that primary tumors may adopt a phenotype that promotes survival in the brain microenvironment. Therapies targeting unique vulnerabilities conferred by such epigenetic reprogramming and diagnostics to identify these harbingers in primary tumors would likely mitigate metastatic disease and improve survival in patients with NSCLC.

3.4 Methods

3.4.1 Patients and specimens, and sample preparation

Patient-matched non-malignant lung tissue, primary NSCLC, and brain metastasis specimen trios together with available, de-identified clinical data were obtained from existing diagnostic formalin-fixed, paraffin embedded (FFPE) pathology tissue blocks under the corresponding institutional IRB approvals (HRPO # 201301001). Each specimen was sectioned, stained with hemotoxylin and eosin, and reviewed by a pathologist to ensure that it was representative of the pathology diagnosis of record (non-malignant / primary tumor / CNS metastasis). Serial sections from each block were then cut either to slides or tissue curls for immediate nucleic acid isolation. Isolated nucleic acid was quantified by Qubit fluorometry and RNA was qualitatively assessed using an Agilent bioanalyzer. After case selection, tissue review, and DNA isolation, a total of 45 cases were available with DNA from representative non-malignant, primary tumor, and metastasis lesions for analysis.

3.4.2 Exome sequencing, alignment, and variant calling

Library preparations for whole exome sequencing (WES) from genomic DBA was performed using the Agilent Exome SureSelect V5 kit. The resulting library fragments were sequenced from each end on an Illumina HiSeq-2500 instrument to an average depth of 30X. Sequence reads (FASTQ files) were mapped to the human reference genome (GRCh37/hg19) and BAM files were generated using Novoalign tool v3.02.06 (http://www.novocraft.com) followed by marking and removing PCR duplicates by PICARD tools v1.67 (http://broadinstitute.github.io/picard/). Genome Analysis Toolkit (GATK) v3.3.0 was used for local realignment and base quality score recalibration. Local realignment around indels was performed by *RealignerTargetCreator* and *IndelRealigner* tools of GATK followed by Base Quality Score Recalibration by *BaseRecalibrator*. Finally, *PrintReads* tool was used to generate analysis-ready BAM files.

3.4.3 Mutation calling and significantly mutated genes (SMGs)

Mutect v1.14[21] was used to detect somatic point mutations. The input for Mutect was BAM files generated by GATK and paired sample mode was used to call somatic SNVs and small indels by Mutect. GATK IndelGenotyperV2 (https://software.broadinstitute.org/cancer/cga/indelocator) was used to call somatic indels with default parameters. The variants were annotated and Mutation Annotation Format (MAF) files were generated using Oncotator[22]. Significantly mutated genes (SMGs) were identified using MutSigCV[23] with a q-value threshold of 0.05.

3.4.4 Copy number analysis

The R package, cn.mops[24] was used to detect copy number variations from the whole exome sequencing data. Paired analysis was run for each patient with BAM files from non-malignant, primary tumor, and malignant samples. The *referencecn.mops* function was specifically used to detect copy number variations (CNVs). A segment mean < 0.5 was considered a loss and a segment mean >0.6 was considered a gain. An in-house script was used to find genes residing in the segments identified to be amplified/deleted by cn.mops.

Furthermore, we correlated the findings of cn.mops with transcriptomics data. We identified genes in amplified segments with high expression (segment mean neutral in primary tumor, segment mean > 0.6 in metastatic sample, and TPM fold change > 2 in metastasis compared to primary tumor sample) in RNA-seq data. Similarly, genes in segments identified as loss by

cn.mops were correlated with low expression values in RNA-seq data (segment mean neutral in primary tumor, segment mean < -0.5 in metastatic sample, and TPM fold change < 0.5 in metastasis compared to primary tumor sample).

3.4.5 Metastasis Enriched Variant (MEV) analysis

For each patient, unique somatic variants (SNVs and small Indels) were extracted, and their variant allele frequency (VAF) was determined for primary tumor and brain metastasis samples. Allelic counts and depth were calculated using samtools mpileup[25] and an in-house script was used to parse the pileup file format. Metastasis Enriched Variants" (MEVs) were chosen based on the following criteria - VAF in brain metastasis sample >30, VAF in paired primary tumor sample < 10 and depth in primary tumor >30.

3.4.6 Purity and ploidy estimates

R package Sequenza[26] was used to estimate purity and ploidy of the primary tumor and brain metastasis samples. The normal and tumor BAM files for a patient were processed together by *sequenza-utils* to generate a *seqz* file which served as the input for the tool. Purity and ploidy calculations were performed using default parameters by tool *sequenza.fit*. Out of all the solutions put forward by the model we selected the purity-ploidy pair with the highest posterior probability.

3.4.7 Tumor Mutation Burden (TMB), Fraction of Genome Altered (FGA), and Jaccard score estimates

Tumor mutation burden (TMB) was calculated as the number of non-synonymous somatic mutations (SNVs and small indels) per megabase in coding regions (mut/Mb). The total genomic region captured by Agilent Exome SureSelect V5 is 50mb. TMB was thus calculated as the total number of nonsynonymous somatic mutations in a sample / 50. Fraction of genome altered (FGA) was calculated as the total genomic region altered in a sample as detected by cn.mops per base of the human genome. Total genomic region altered for a sample was measured by summing up the base pairs of all segments identified as copy number alterations by cn.mops.

Jaccard scores for primary-metastasis sample pairs were calculated by Bedtools' jaccard tool (https://bedtools.readthedocs.io/en/latest). Jaccard scores ranged from 0.0 to 1.0, where 0.0 represents no mutational overlap between primary and metastasis samples and 1.0 represents complete overlap.

3.4.8 Mutation signature analysis

The R package, SigProfilerMatrixGeneratorR, was used to calculate mutation matrices associated with single-base-substitution signatures (SBS), double-base signatures (DBS) and short indels detected in whole exome sequencing data (https://github.com/AlexandrovLab/SigProfilerMatrixGenerator). The SBS matrix calculated in 96 trinucleotide contexts was used in the Python package SigProfilerExtractor (https://github.com/AlexandrovLab/SigProfilerExtractor) for a sample-wise decomposition into predominant COSMIC signatures (version 3.1)[11]. Custom code was written in R to calculate the cosine-similarity of the sample SBS signatures and mutagen-associated SBS signatures. The mutagen SBS signatures were obtained from data reported in[12]. The z-score of mutations associated with each COSMIC signature and the cosine-similarity values were hierarchically clustered and plotted using the R function pheatmap.

3.4.9 RNA-sequencing analysis

Total RNA from 30 pairs of primary and metastasis FFPE tissue blocks was isolated using Qiagen RNseasy protocol with manufacturer's recommended modification for FFPE tissue[27] by

the Siteman Cancer Center Tissue Procurement Core. RNA was quantified by nanodrop spectrophotometer and quality assessed using Agilent TapeStation. RNA (2ug) was depleted of rRNA fragments using RiboZero reagent and subject to library construction and sequencing by the Washington University Genome Technology Access Center (GTAC). Sequencing files were reviewed using FastQC and adapter sequences and low quality 3' ends were removed, satisfying a *q*-value threshold of 20. Reads were then aligned to hg19 rRNA sequences from NCBI using Bowtie2[28], and reads mapping to rRNA were discarded. Remaining reads were then mapped using STAR[29] to the hg19 genome with the NCBI RefSeq GTF file downloaded from the UCSC Genome Browser[30]. featureCounts[31] was used to assign mapped reads to genes using the same GTF file described above and a minimum mapping quality score of at least 20. Transcripts per million reads (TPM) values were calculated for each gene for each sample individually.

3.4.10 Gene ontology analysis of significantly differentially expressed genes

Gene counts obtained using featureCounts[31] (described above) were input to DESeq2[32] and genes significantly differentially expressed (adjusted *p*-value < 0.05) between the primary and metastasis samples were identified. 401 genes were determined to have a significant increase in expression in the metastasis samples compared to primaries (Appendix B.3). The list of 401 genes was input to DAVID[33,34], where 398 gene names were recognized (genes unrecognized: *LINC01806*, *DCST1-AS1*, *ZNF8-ERVK3-1*). GOTERM_BP_DIRECT results were then filtered to include only terms with a Benjamini Hochberg corrected value of < 0.05. 534 genes were determined to have a significant decrease in expression in the metastasis samples compared to the primaries (Appendix B.3). The list of 534 genes was input to DAVID, and 533 were recognized (gene unrecognized: *LINC01934*). GOTERM_BP_DIRECT results were then filtered to include only terms with a Benjamini Hochberg corrected value of < 0.05.

3.4.11 Single-sample gene set enrichment analysis (ssGSEA)

Gene lists were obtained from the following sources: Brain (GTEx; top 100 expressed genes from each of the 12 brain tissues available were obtained, duplicates removed, and any overlapping the top 100 expressed lung genes were removed; *n*=169), Lung (GTEx; top 100 expressed genes from lung were obtained, and any overlapping the top 100 expressed brain genes from any of 12 available brain tissues were removed; *n*=48), Allograft rejection (GSEA, gene set "M5950": "HALLMARK_ALLOGRAFT_REJECTION")[35], Oligodendrocyte (top human expressing genes with rank < 50 were obtained and filtered for unique genes per cell type; *n*=18)[36], Astrocyte (top human expressing genes with rank < 50 were obtained and filtered for unique genes per cell type)[36], and SOX2[37]. Single-sample Gene Set Enrichment Analysis was then run on each gene set, using RNA-seq data from paired primary and metastasis tumors (*n*=30 pairs) as input[15].

3.4.12 DNA methylation profiling

DNA samples from non-malignant lung tissue, primary lung tumor, and brain metastasis from 12 patients were obtained (Table 3.1). For each patient sample trio, global methylation profiling was performed using Methylation Dependent Immunoprecipitation followed by sequencing (MeDIP-seq) and Methylation-sensitive Restriction Enzyme digestion followed by sequencing (MRE-seq) as previously described[38,39]. All samples passed sequencing depth threshold requirements (Table 3.1). Single CpG-resolution methylation values for all MeDIP-seq + MRE-seq samples were estimated using methylCRF (where -c was either set to 0 or 3 depending on sequencing)[40]. Differentially methylated regions (DMRs) across primary-metastasis pairs were calculated using

MnM with default parameters and a stringency threshold of q=0.00001, as previously described[41], using MeDIP-seq and MRE-seq data from each sample as input and excluding blacklisted CpG regions. Data were visualized on the Washington University Epigenome Browser[42,43].

3.4.13 DNA methylation characterization

Promoters, CpG islands, exons, introns, intergenic regions, and repeats were defined using the NCBI RefSeq annotations for hg19 downloaded from the UCSC genome browser[30]. Promoter regions were defined as +1.5 kb upstream the transcription start site (TSS) to -.5 kb downstream the TSS. Partially methylated domains (PMDs) were calculated based on a previously described method[44]. In short, the genome was divided into 10 kb bins, and the average methylation per 10 kb bin was calculated for each of the 12 normal lung samples, omitting blacklisted CpGs (ENCODE accession: wgEncodeEH001432). 10 kb bins with an average methylation less than 70% in all normal samples and containing at least 10 CpGs were called PMDs. Adjacent PMDs were merged. DNA methylation valleys (DMVs) were calculated as previously described[45]. In short, the genome was divided into 1 kb bins and the average methylation per 1 kb bin was calculated for each of the 12 normal lung samples. A sliding window was then applied to calculate the average methylation across 5 kb windows, incrementing 1 kb at a time. Average methylation across 5 consecutive 1 kb bins was calculated, and regions with less than or equal to 15% methylation were classified as DMVs. Consecutive DMVs were then merged. Human super enhancers for all available tissues and cell types (n=99) were downloaded directly from the dbSUPER database[46]. To determine the locations of brain-specific active enhancers, all active enhancers (defined previously according to ROADMAP; chromatin states "9_EnhA1" and "10_EnhA2" in 18-state model predictions[47]) were merged, and any active enhancers defined in any other cell type were removed. Paired t-tests were performed to determine if there were significant changes in methylation over any genomic category between normal-primary, primary-metastasis, and normal-primary. Resulting p-values were then corrected using BH.

3.4.14 Identification of genes with recurrent differentially methylated regulatory regions

To determine which genes contained differentially methylated regulatory elements across patients, we first identified DMRs associated with promoters and enhancers. Promoters were defined based on the NCBI RefSeq annotations for hg19 downloaded from the UCSC genome browser, spanning +1.5 kb upstream the TSS to -.5 kb downstream the TSS[30]. As long as any portion of the DMR overlapped with the promoter, we classified the promoter as containing a DMR. Note that multiple TSSs could potentially overlap a single DMR, and each would be counted. Regions classified as active enhancers (chromatin states "9_EnhA1" and "10_EnhA2" in 18-state model predictions) in at least a single ROADMAP cell type[47] not overlapping promoters were identified, intersected with DMRs, and assigned to genes based on the closest TSS.

3.4.15 Change in expression of genes with recurrent differentially methylated regulatory regions and those assigned to brain-specific active enhancers

RNA-seq data was processed as described above for all samples. Beginning with the gene counts matrix output from featureCounts[31], differential expression between primary and metastasis samples was calculated using DESeq2 v1.26.0[32] for the 12 samples where methylation analysis

was also done (Table 3.1). Log2 fold changes in expression were used directly from the DESeq2 output and *p*-values were adjusted for gene sets using the BH correction method.

3.4.16 Identification of enriched transcription factor binding motifs (TFBMs) within recurrent DMV hyperDMRs

Autosomal hyperDMRs from primary to metastasis shared in at least 2 patients that overlapped DMVs was used as input to Analysis of Motif Enrichment (MEME Suite 5.3.3)[32] for the identification of enriched TFBMs. The motif database searched was HOCOMOCO Human (v11, full), and default parameters were used. All significantly enriched TFBMs are reported, as well as their predicted methylation binding status, as reported in[16] (Appendix B.5).

3.4.17 Assigning brain-specific active enhancers to genes

The distance from each brain-specific active enhancer to each TSS was calculated. The top 3 genes with the shortest distance from one of their TSSs to the enhancer were considered for each enhancer.

3.5 Figures

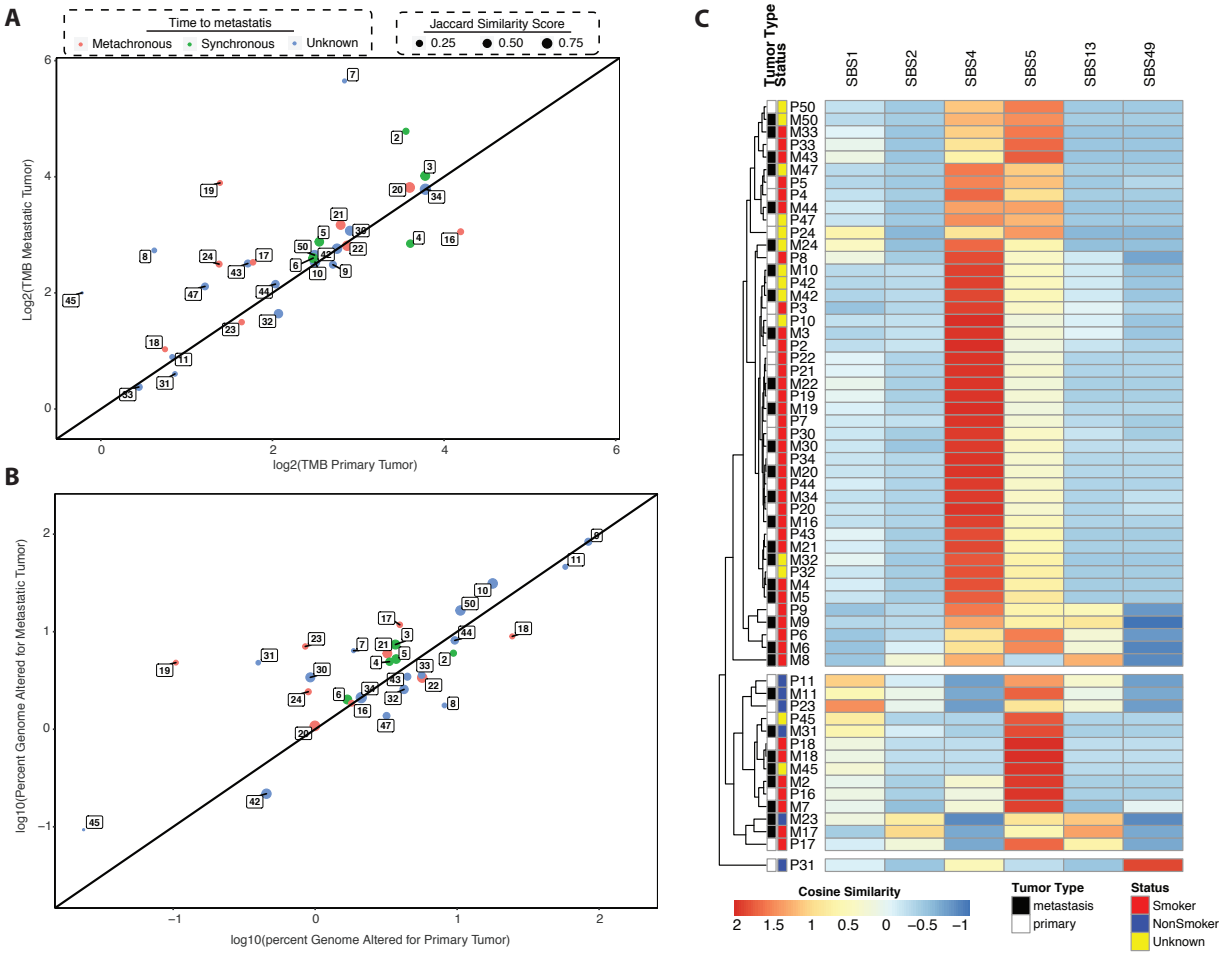


Figure 3.1. Comparative genomic landscape between primary and brain metastatic NSCLC.

(A) Comparative tumor mutation burden (TMB), and (B) comparative fraction of genome altered (FGA) between primary tumor and metastatic pairs. Patients are numbered (see Table 3.1) and color-coded by metastatic presentation while the size of each point is proportional to the Jacaard similarity score between patient pairs. (C). Profile of six major single-base-substitution (SBS) signatures detected between primary (P) and metastasis (M) pairs for each patient, with SBS cosine similarity for each signature depicted in color scale. Hierarchical clustering of individual samples based on SBS signature profile is shown and annotated with smoking history.

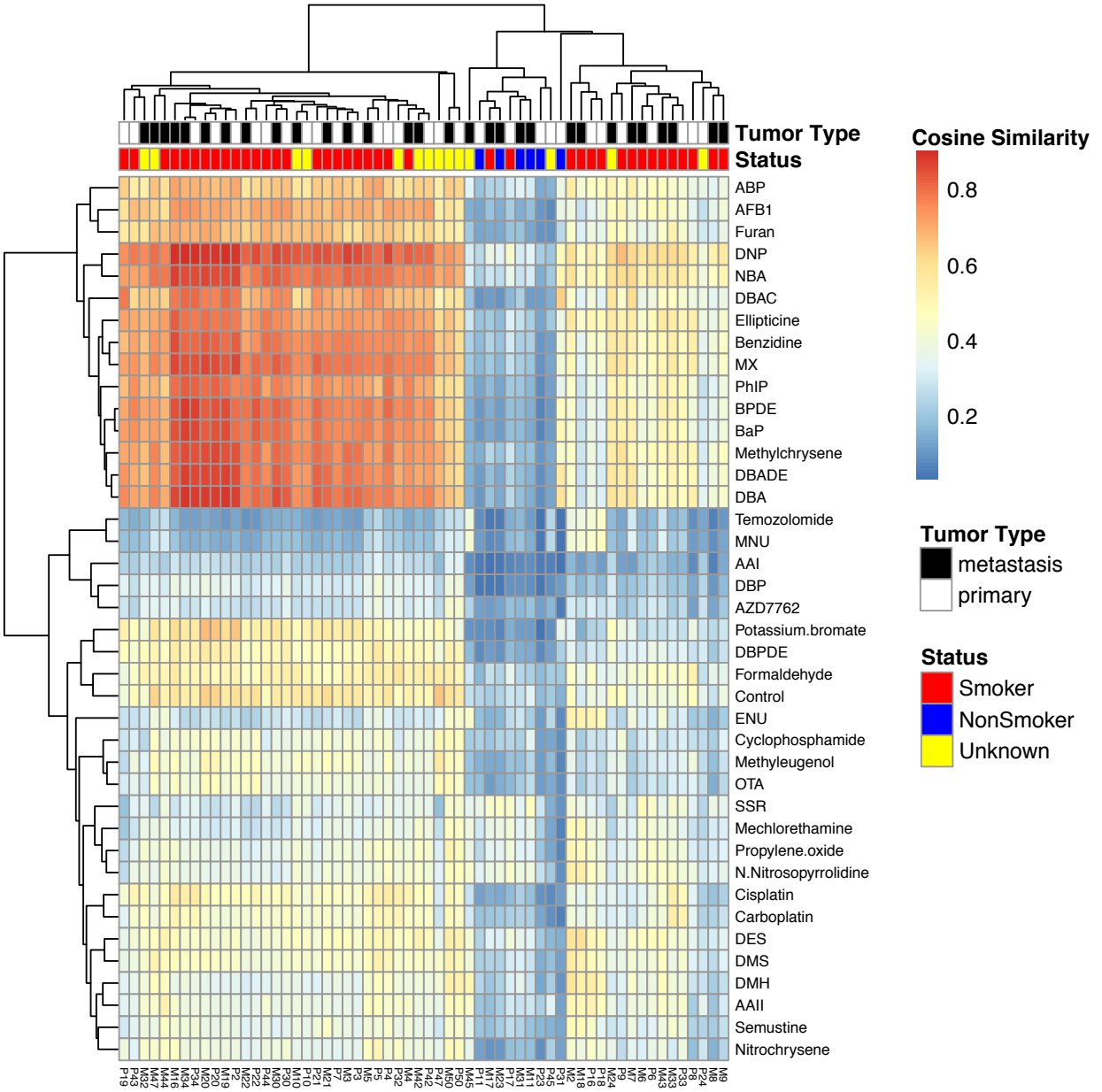


Figure 3.2. Cosine-similarity values between samples and mutagen SBS signatures.
Single-Base-Substitution (SBS) signatures associated with different canonical environmental mutagens, expressed in color-scale of cosine-similarity values, for all primary tumor (P) and corresponding brain metastasis (M) samples. Patient numbers are indicated. All samples are represented as a hierarchical cluster based on similarity of their overall SBS signature, and known smoking history for each patient is also shown.

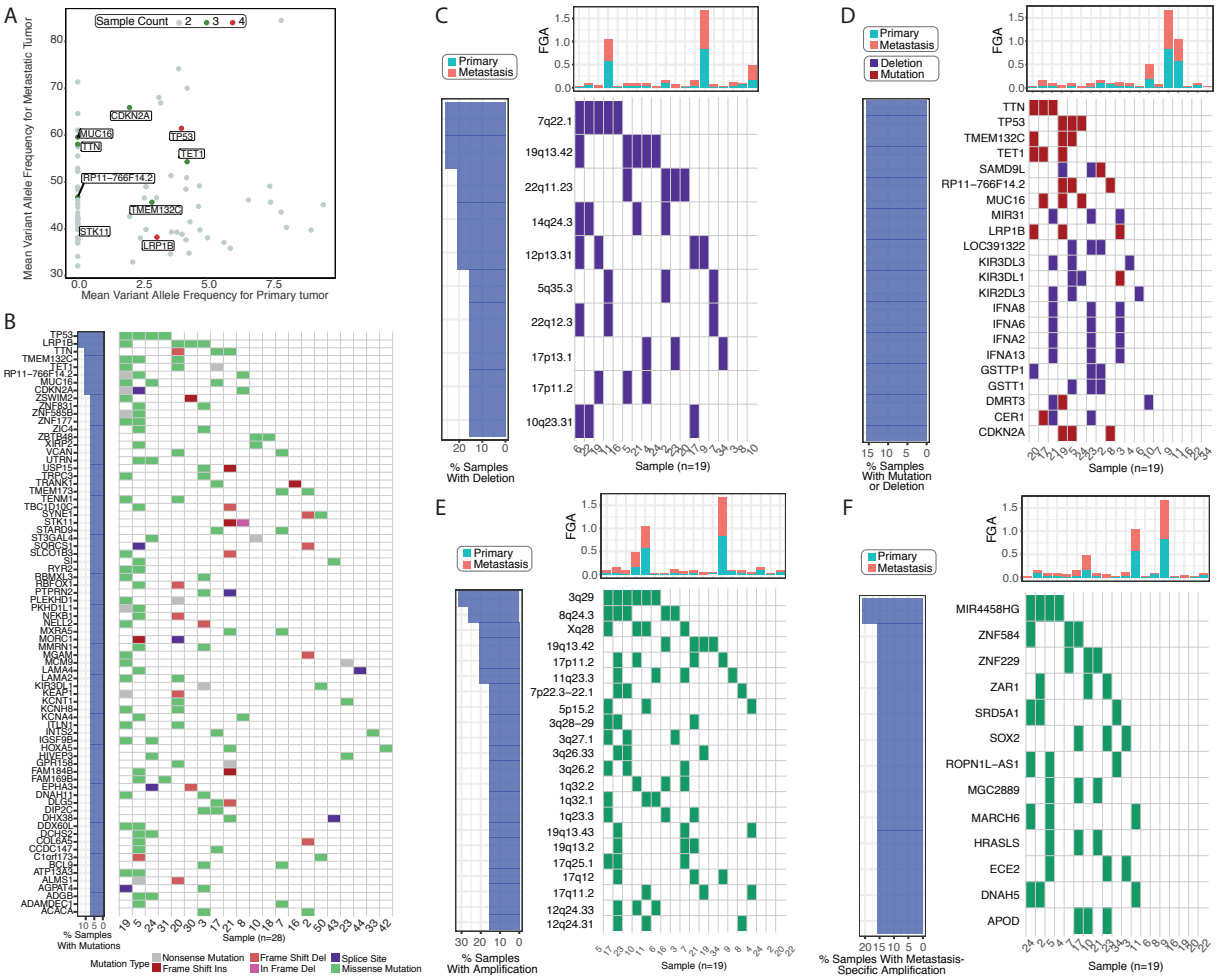


Figure 3.3. Metastasis enriched variants (MEVs) and relative copy number changes in primary vs. metastatic tumors.

(A) Dot plot comparing recurrent MEVs in primary tumor (x-axis) and paired CNS metastasis (y-axis). Genes with variants present in 2, 3, or 4 patients are color-coded. (B) Matrix plot of genes with recurrent MEVs. (C-D) Recurrent metastasis-specific losses at the chromosomal region level (C) and individual gene level (D), where somatic mutations are also denoted in (D). (E-F) Recurrent metastasis-specific amplifications at the chromosomal region level (E) and individual gene level (F). In panels A-F, the overall fraction of genome alterations (FGA) for each primary-metastasis pair is also shown for reference.

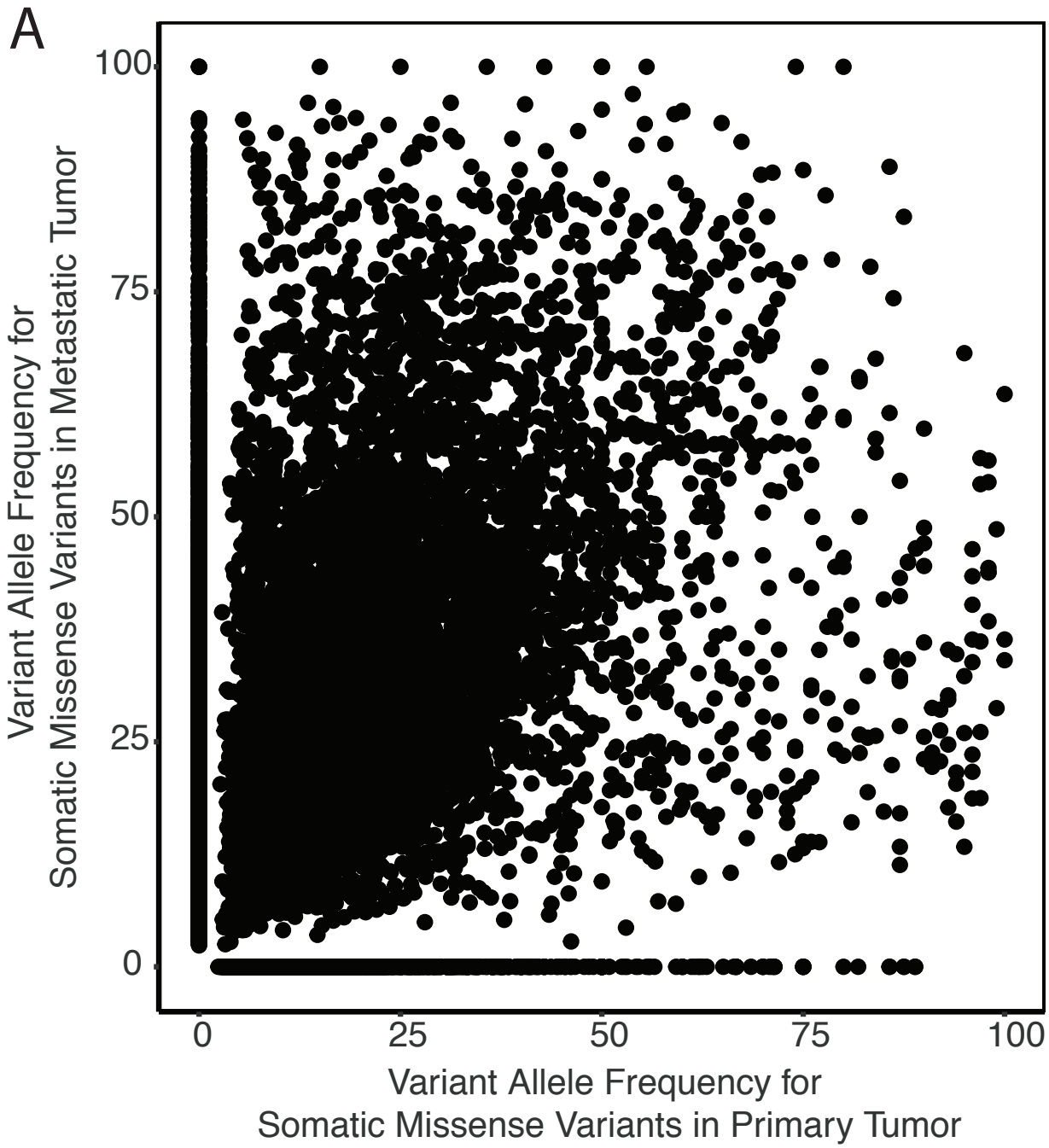


Figure 3.4. Comparison of all variant allele frequencies in primary vs metastatic samples. Dot plot comparing variant allele frequencies (VAFs) of all somatic variants identified in primary tumor (x-axis) and paired CNS metastasis (y-axis).

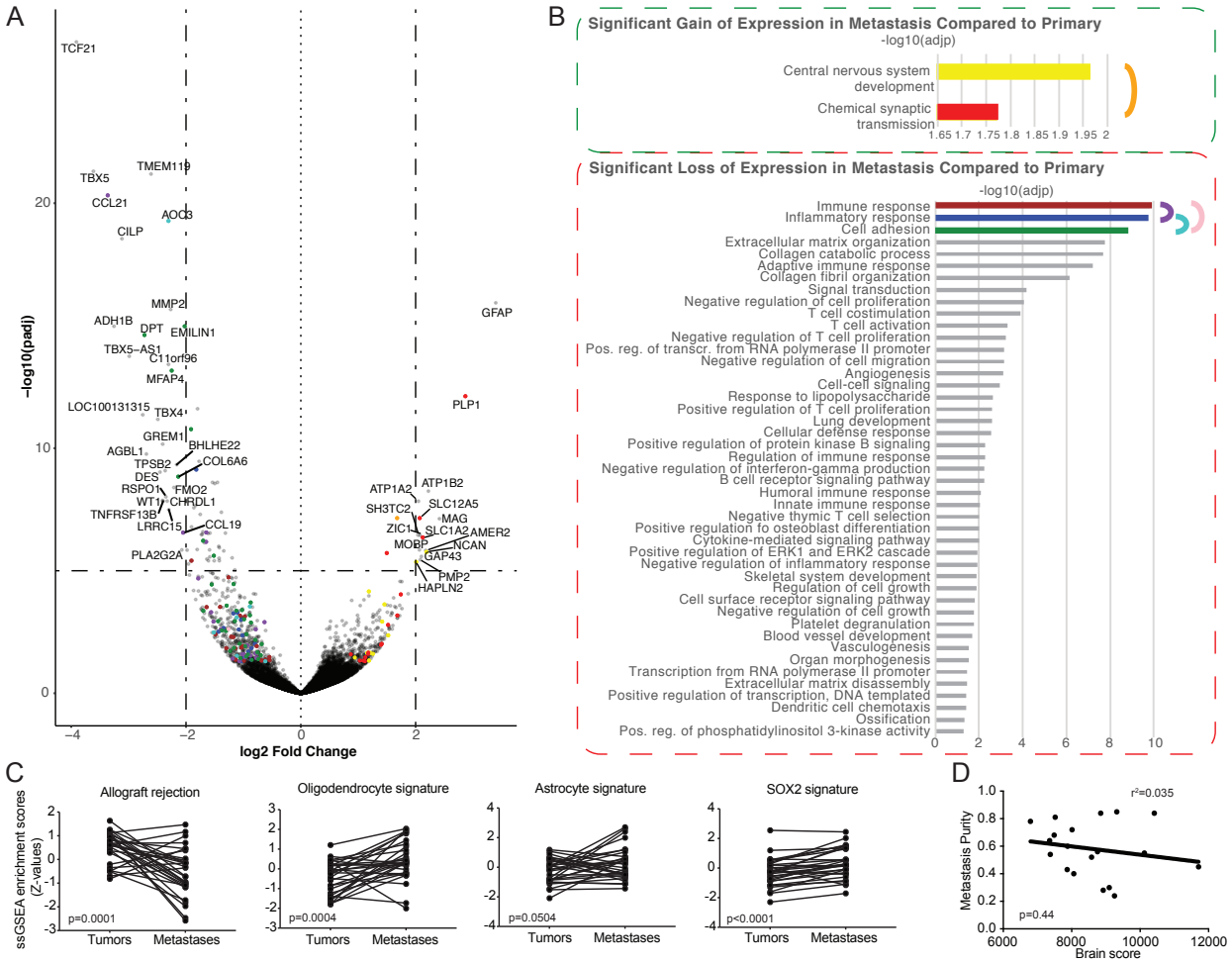
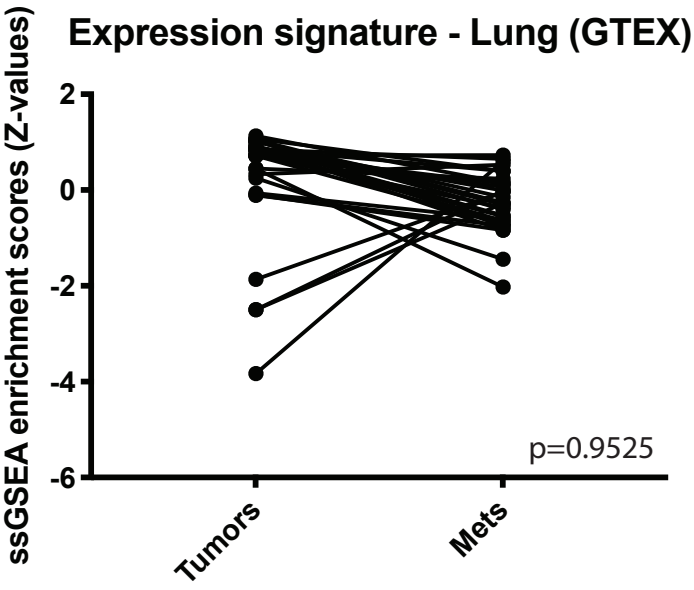


Figure 3.5. "Glial-like" differential gene expression in primary vs metastatic NSCLC.
(A) Volcano plot demonstrating group-wise differences in gene expression between primary tumors and brain metastases. Genes belonging to gene ontology groups in (B) are denoted by color. (B) Gene ontology analysis demonstrating statistical enrichment of genes associated with neuronal / glial phenotypes in metastatic lesions and corresponding down-regulation of immune-related pathways. (C) Single-sample gene set enrichment analysis (ssGSEA) of primary-metastasis sample pairs from individual patients (n=30) showing recurrent gene expression enrichment of oligodendrocyte, astrocyte, and SOX2-regulated gene sets in metastatic tumors as compared to their primary counterparts. (D) Scatter plot demonstrating lack of correlation of overall enrichment of brain gene expression scores (GTEx, see 3.4 Methods) to tumor cellularity in the brain metastatic lesions.

A



B

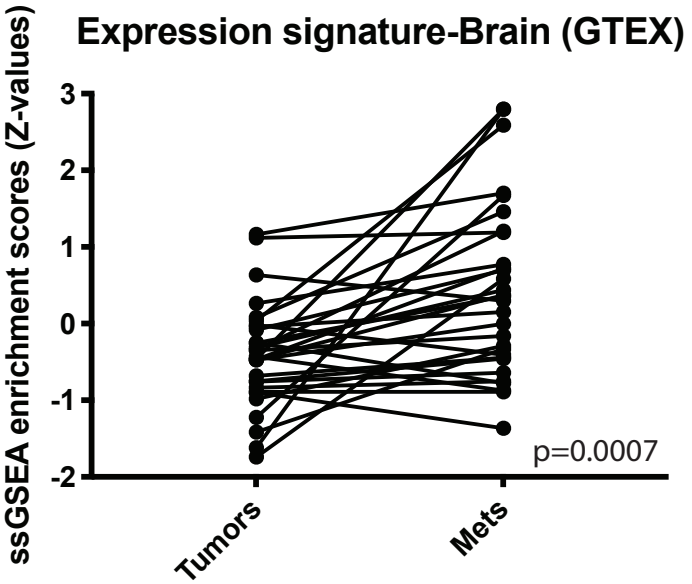


Figure 3.6. singe-sample gene set enrichment analysis comparisons between paired primary and metastasis samples.

(A) Comparison of single-sample gene set enrichment analysis (ssGSEA) z-scores of lung expression in primary tumors compared to matched metastases (paired t-test). (B) Comparison of single-sample gene set enrichment analysis (ssGSEA) z-scores of brain expression in primary tumors compared to matched metastases (paired t-test).

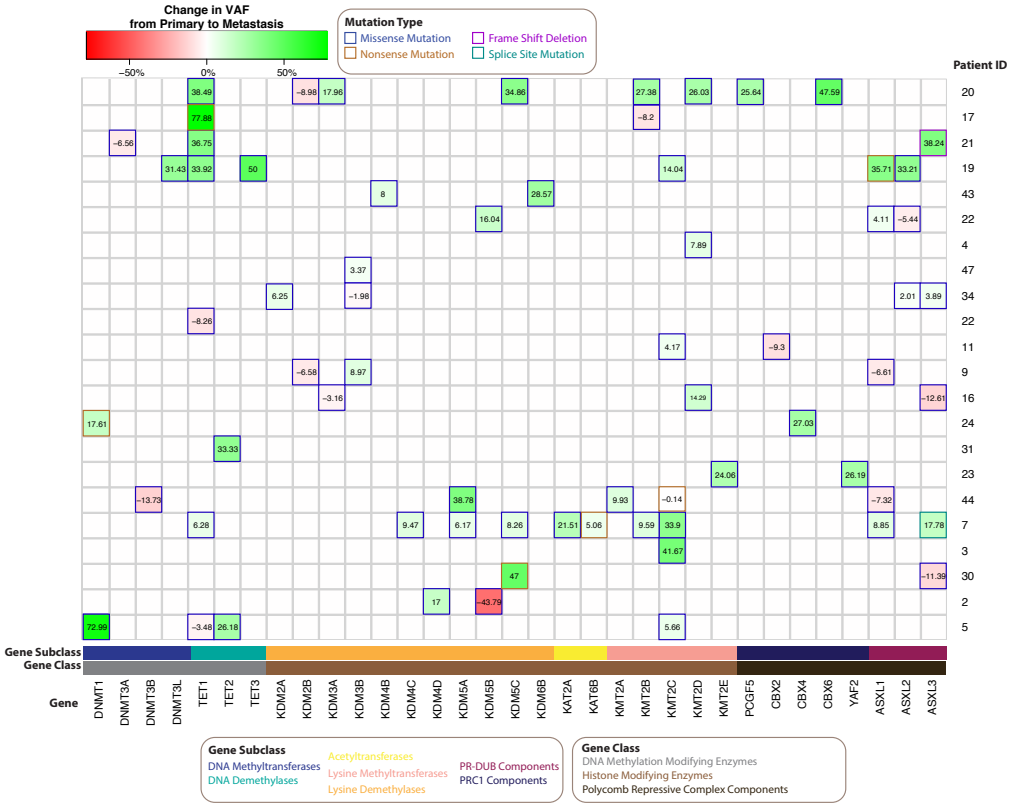


Figure 3.7. Change in variant allele frequencies (VAFs) of epigenetic gene mutations. Changes in VAFs found in canonical epigenomic regulatory protein genes. Max change in VAF is shown if there were multiple mutations within a gene. Only predicted deleterious mutaitons with a primary depth of at least 30x are displayed.

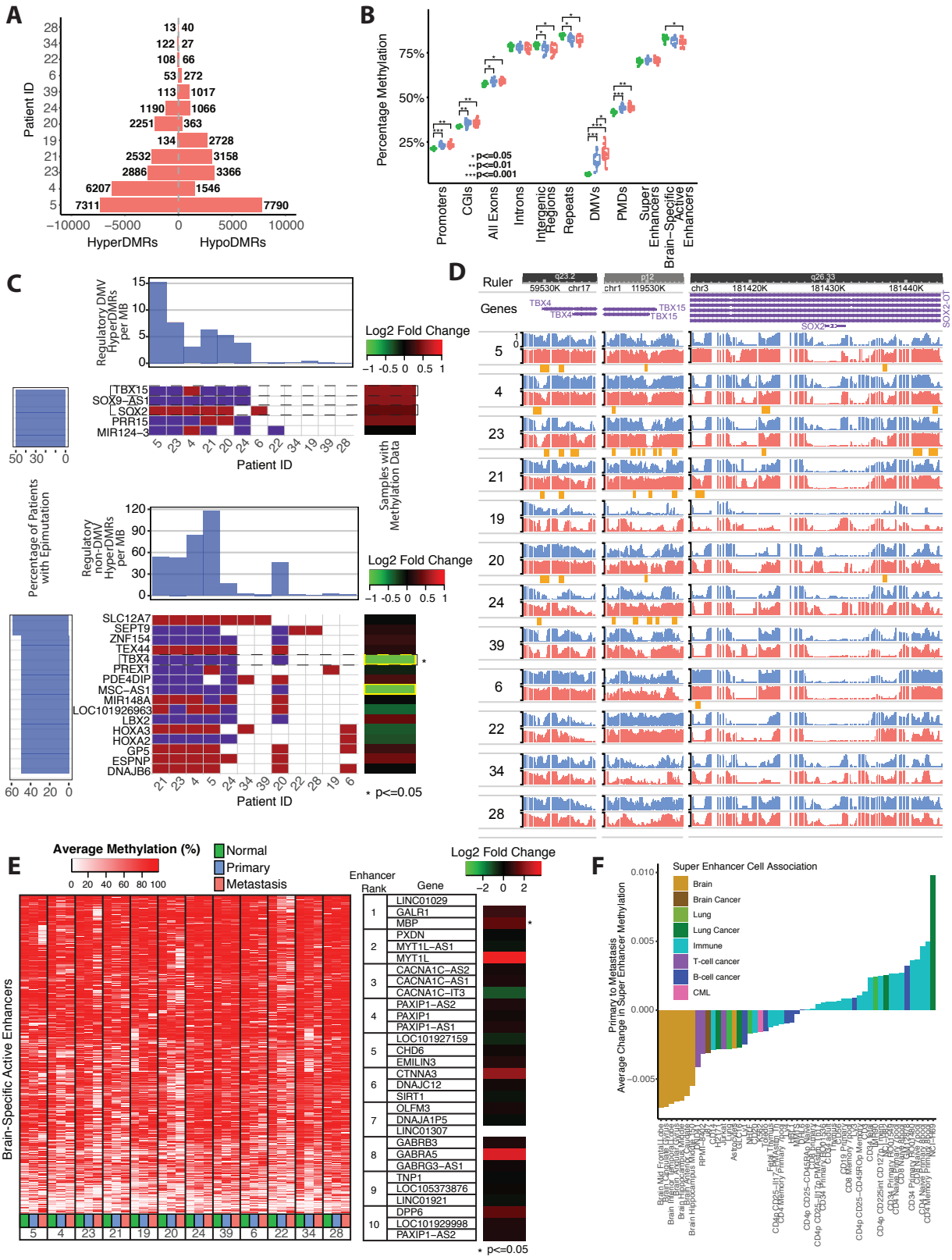

A
Patient ID
HyperDMRs
HypoDMRs
B
Percentage Methylation
Promoters
CGIs
All Exons
Introns
Intergenic Regions
Repeats
DMVs
PMDs
Super Enhancers
Brain−Specific Active Enhancers
C
Regulatory DMV HyperDMRs per MB
Log2 Fold Change
Samples with Methylation Data
Percentage of Patients with Epimutation
Regulatory non-DMV HyperDMRs per MB
Patient ID
p<=0.05
D
Ruler
Genes
E
Average Methylation (%)
Normal
Primary
Metastasis
Brain-Specific Active Enhancers
Enhancer Rank
Gene
F
Super Enhancer Cell Association
Brain
Brain Cancer
Lung
Lung Cancer
Immune
T-cell cancer
B-cell cancer
CML
Primary to Metastasis Average Change in Super Enhancer Methylation

Figure 3.8. Regulatory methylation changes from primary to metastasis reveal increased brain-specific enhancer activation and loss of active immune-related enhancers, accompanying correlated expression changes.
(A) Number of differentially methylated regions (DMRs) identified between the primary and metastasis sample for each of 12 patients. (B) Average methylation over genomic categories, separated by disease status (*t*-tests, BH corrected, only significant comparisons are highlighted). (C) Genes with the most recurrent regulatory methylation changes and corresponding expression changes. Top: genes within DNA methylation valleys (DMVs); bottom: genes outside DMVs. Values reaching saturation (log2 fold change of < -1) are highlighted in yellow. (D) Browser view showing the level of methylation (0-1) across regulatory regions for *TBX4*, *TBX15*, and *SOX2*, for primary (blue) and metastasis (red) samples. Brown rectangles beneath each primary-metastasis pair indicate hyperDMRs called in that patient. (E) Left: average methylation across brain-specific active enhancers in normal, primary, and metastasis samples. Right: 3 closest genes to brain-specific active enhancers ranked according to greatest loss of methylation (primary to metastasis) and corresponding expression changes. Expression change *p*-values output by DESeq2 were BH corrected. (F) Average change in super-enhancer methylation (primary to metastasis), grouped by cell type association.

3.6 Tables

Table 3.1 Summary of sample clinical and experimental values

Master Patient ID	Met type	Gender	Race	Smoking History	Histology	Included in Exome-seq analysis (YES/NO)	Mutational signatures data (YES/NO)	Tumor Max Cellularity (Sequenza)	Tumor Max Ploidy (Sequenza)	Metastasis Max Cellularity (Sequenza)	Metastasis Max Ploidy (Sequenza)	TMB (Met)	TMB (Tumor)	FGA (Met)	FGA (Tumor)	Jaccard	Included in RNA-seq analysis (YES/NO)	Included in methylation analysis (YES/NO)	Number of Non-Redundant, Uniquely Mapped MeDIP Fragments (Normal)	Number of MRE filtered reads (Normal)	Number of CpG sites sampled (MRE) (Normal)	Number of Non-Redundant, Uniquely Mapped MeDIP Fragments (Primary)	Number of MRE filtered reads (Primary)	Number of CpG sites sampled (MRE) (Primary)	Number of Non-Redundant, Uniquely Mapped MeDIP Fragments (Metastasis)	Number of MRE filtered reads (Metastasis)	Number of CpG sites sampled (MRE) (Metastasis)
1	ND	FEMALE	WHITE	SMOKER	ADENOCARCINOMA	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
2	Syn	FEMALE	AFRICAN AMERICAN	SMOKER	ADENOCARCINOMA	YES	YES	0.18	5.8	0.56	2.8	27.54	11.76	0.060241	0.094205	0.240174	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
3	Syn	MALE	WHITE	SMOKER	ADENOCARCINOMA, POORLY DIFFERENTIATED	YES	YES	0.27	3.4	0.68	3.4	16.18	13.74	0.073851	0.036754	0.645973	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
4	Syn	MALE	WHITE	SMOKER	ADENOCARCINOMA	YES	YES	0.52	3.3	0.52	3.3	7.2	12.18	0.049276	0.033421	0.405224	YES	YES	73416623	48367431	1594200	45759761	27440261	1931307	38601951	31439552	1735960
5	Syn (LCNEC brain)	FEMALE	WHITE	SMOKER	ADENOCARCINOMA	YES	YES	0.16	2.8	0.85	3.6	7.36	5.84	0.052445	0.037222	0.585938	YES	YES	36628791	37860547	1533749	36874825	26780372	1620655	39390251	33124201	1484732
6	Syn	FEMALE	WHITE	SMOKER	SQUAMOUS CELL CARCINOMA	YES	YES	0.36	3.4	0.54	3.8	6.06	5.56	0.020177	0.016849	0.629776	YES	YES	47955480	34098626	1340320	43943922	27308844	1479275	40680804	27606556	1436488
7	ND	FEMALE	WHITE	SMOKER	NON-SMALL CELL CARCINOMA	YES	YES	0.47	3.3	0.64	3.6	50.22	7.16	0.063871	0.018635	0.07838	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
8	ND	FEMALE	WHITE	SMOKER	ADENOCARCINOMA	YES	YES	0.15	3.9	0.3	3.6	6.64	1.54	0.01754	0.081453	0.120205	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
9	ND	MALE	WHITE	SMOKER	ADENOCARCINOMA, POORLY DIFFERENTIATED	YES	YES	0.99	2.3	0.84	2.4	5.6	6.52	0.836698	0.839681	0.343215	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
10	ND	MALE	WHITE	SMOKER	ADENOCARCINOMA	YES	YES	0.61	3.5	0.84	3.4	5.78	5.64	0.312332	0.178039	0.75387	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
11	ND	FEMALE	WHITE	NONSMOKER	ADENOCARCINOMA	YES	YES	0.24	3.3	0.24	6.8	1.86	1.78	0.462246	0.577811	0.134426	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
13	ND	MALE	WHITE	SMOKER	ADENOCARCINOMA	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
14	ND	FEMALE	WHITE	SMOKER	ADENOCARCINOMA	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
15	ND	FEMALE	AFRICAN AMERICAN	SMOKER	ADENOCARCINOMA	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
16	Meta	MALE	WHITE	SMOKER	MIXED ADENO/NEURO	YES	YES	0.38	4.1	0.55	3.7	8.3	18.3	0.018469	0.018057	0.21519	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
17	Meta	FEMALE	ASIAN	SMOKER	ADENOCARCINOMA	YES	YES	0.48	2	0.81	3.7	5.76	3.42	0.117777	0.039471	0.178492	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
18	Meta	MALE	ND	SMOKER	SQUAMOUS CELL CARCINOMA	YES	YES	0.39	3.3	0.59	3.5	2.04	1.68	0.089557	0.244522	0.137019	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
19	Meta	MALE	WHITE	SMOKER	ADENOCARCINOMA	YES	YES	0.11	2	0.45	4.8	14.86	2.62	0.048191	0.001045	0.128778	YES	YES	41160590	55189685	1905327	45966175	24589542	1512191	27680417	27467013	1376658
20	Meta	FEMALE	AFRICAN AMERICAN	SMOKER	ADENOCARCINOMA	YES	YES	0.28	2.6	0.72	2.3	14.08	12.12	0.010823	0.009957	0.775339	YES	YES	28508944	25769771	1242424	32292234	13236869	1435900	30494631	23454005	1394917
21	Meta	FEMALE	WHITE	SMOKER	ADENOCARCINOMA, POORLY DIFFERENTIATED	YES	YES	0.15	5.4	0.6	3.7	9	6.94	0.059845	0.032256	0.621094	YES	YES	42036159	22152065	1270311	31945145	20037171	1450734	49004874	21784884	1535680
22	Meta	FEMALE	WHITE	SMOKER	ADENOCARCINOMA	YES	YES	0.34	5.3	0.28	4.9	7.02	7.32	0.033811	0.056735	0.730676	YES	YES	42272773	26487553	1461826	39707432	36042971	2092491	34312156	23980378	1900609
23	Meta	FEMALE	ASIAN	NONSMOKER	ADENOCARCINOMA	YES	YES	0.23	3.1	0.78	4.1	2.82	3.12	0.070593	0.008519	0.167241	YES	YES	47756141	32837832	1523850	41808806	11423436	1151597	33637329	25264348	1506970
24	Meta	FEMALE	WHITE	ND	CARCINOMA, POORLY DIFFERENTIATED	YES	YES	0.43	2.1	0.4	3.6	5.64	2.6	0.024224	0.008948	0.223859	YES	YES	48839534	23495002	1504045	41818421	14640997	1420906	40003428	27713864	1292284
25	ND	MALE	WHITE	SMOKER	ADENOCARCINOMA	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
26	ND	FEMALE	WHITE	SMOKER	ADENOCARCINOMA	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
27	Meta	MALE	WHITE	SMOKER	ADENOCARCINOMA	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
28	ND	MALE	ND	ND	ADENOCARCINOMA	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	YES	YES	39964554	54108608	1399785	36783513	19291825	1130174	44585459	43244090	1517505
29	ND	MALE	ND	ND	ADENOCARCINOMA	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
30	ND	MALE	WHITE	SMOKER	ADENOCARCINOMA	YES	YES	0.2	3.4	0.41	3.3	8.4	7.48	0.034044	0.009245	0.654995	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
31	ND	FEMALE	WHITE	NONSMOKER	ACINAR CELL CARCINOMA	YES	YES	0.26	2.9	0.75	2.8	1.52	1.82	0.0481	0.003973	0.111111	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
32	ND	MALE	WHITE	ND	ADENOCARCINOMA	YES	YES	0.49	4.6	0.28	3.8	3.12	4.2	0.025624	0.042291	0.544872	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
33	ND	MALE	WHITE	SMOKER	ADENOCARCINOMA	YES	YES	0.17	6.3	0.17	6.3	1.3	1.36	0.035996	0.05629	0.303318	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
34	ND	FEMALE	WHITE	SMOKER	ADENOCARCINOMA	YES	YES	0.36	2.4	0.43	2.4	13.78	13.76	0.021088	0.021119	0.902271	YES	YES	45655695	20508236	1789400	44620008	20462172	2470336	36534456	19348471	2314652
39	ND	FEMALE	ND	ND	ADENOCARCINOMA	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	YES	YES	43487063	40791564	1475551	39159205	21208134	1442202	45807803	30846187	1463536
40	ND	FEMALE	ND	ND	CARCINOMA, POORLY DIFFERENTIATED	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	YES	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
41	ND	FEMALE	WHITE	ND	ADENOCARCINOMA	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
42	ND	MALE	AFRICAN AMERICAN	ND	ADENOCARCINOMA	YES	YES	0.29	3.3	0.3	3.5	6.78	6.74	0.002182	0.004541	0.724	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
43	ND	FEMALE	ND	SMOKER	ADENOCARCINOMA	YES	YES	0.42	2	0.78	1.9	1.68	3.28	0.034581	0.04463	0.34413	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
44	ND	FEMALE	WHITE	SMOKER	ADENOCARCINOMA	YES	YES	0.24	2.7	0.33	3	4.44	4.1	0.081435	0.096285	0.400332	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
45	ND	MALE	AFRICAN AMERICAN	ND	ADENOCARCINOMA	YES	YES	0.24	2.1	0.27	2.4	4	0.86	0.000933	0.000234	0.031782	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
47	ND	MALE	WHITE	ND	ADENOCARCINOMA	YES	YES	0.62	2.1	0.19	2.7	4.32	2.32	0.013723	0.031854	0.319101	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
48	ND	MALE	WHITE	ND	NON-SMALL CELL CARCINOMA	YES (duo)	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
50	ND	FEMALE	OTHER	ND	SQUAMOUS CELL CARCINOMA	YES	YES	0.36	5.1	0.74	3.8	6.28	5.62	0.165172	0.105517	0.725881	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
51	ND	MALE	HISPANIC	ND	ADENOCARCINOMA	YES (duo)	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND
52	ND	MALE	ND	ND	ADENOCARCINOMA	YES (duo)	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND	NO	NO	ND	ND	ND	ND	ND	ND	ND	ND	ND

- ND = No Data
- duo = 2 samples from an individual patient; included in VAF analyses

3.7 References

1. Waqar SN, Samson PP, Robinson CG, et al. Non–small-cell Lung Cancer With Brain Metastasis at Presentation. *Clin Lung Cancer*. 2018;19(4):e373-e379. doi:10.1016/j.cllc.2018.01.007
2. Collisson EA, Campbell JD, Brooks AN, et al. Comprehensive molecular profiling of lung adenocarcinoma: The cancer genome atlas research network. *Nature*. 2014;511(7511):543-550. doi:10.1038/nature13385
3. Riess JW, Gandara DR, Frampton GM, et al. Diverse EGFR Exon 20 Insertions and Co-Occurring Molecular Alterations Identified by Comprehensive Genomic Profiling of NSCLC. *J Thorac Oncol*. 2018;13(10):1560-1568. doi:10.1016/j.jtho.2018.06.019
4. Skoulidis F, Heymach J V. Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nat Rev Cancer*. 2019;19(9):495-509. doi:10.1038/s41568-019-0179-8
5. Brastianos PK, Carter SL, Santagata S, et al. Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov*. 2015;5(11):1164-1177. doi:10.1158/2159-8290.CD-15-0369
6. Shih DJH, Nayyar N, Bihun I, et al. Genomic characterization of human brain metastases identifies drivers of metastatic lung adenocarcinoma. *Nat Genet*. 2020;52(4):371-377. doi:10.1038/s41588-020-0592-7
7. Aljohani HM, Aittaleb M, Furgason JM, et al. Genetic mutations associated with lung cancer metastasis to the brain. *Mutagenesis*. 2018;33(2):137-145. doi:10.1093/mutage/gey003
8. Song Z, Yang L, Zhou Z, et al. Genomic profiles and tumor immune microenvironment of primary lung carcinoma and brain oligo-metastasis. *Cell Death Dis*. 2021;12(1). doi:10.1038/s41419-021-03410-7
9. Laughney AM, Hu J, Campbell NR, et al. *Regenerative Lineages and Immune-Mediated Pruning in Lung Cancer Metastasis*. Vol 26. Springer US; 2020. doi:10.1038/s41591-019-0750-6
10. Ganesh K, Massagué J. Targeting metastatic cancer. *Nat Med*. 2021;27(1):34-44. doi:10.1038/s41591-020-01195-4
11. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793):94-101. doi:10.1038/s41586-020-1943-3
12. Kucab JE, Zou X, Morganella S, et al. A Compendium of Mutational Signatures of Environmental Agents. *Cell*. 2019;177(4):821-836.e16. doi:10.1016/j.cell.2019.03.001
13. Hsu CH, Peng KL, Kang ML, et al. TET1 Suppresses Cancer Invasion by Activating the Tissue Inhibitors of Metalloproteinases. *Cell Rep*. 2012;2(3):568-579. doi:10.1016/j.celrep.2012.08.030
14. Graham V, Khudyakov J, Ellis P, Pevny L. SOX2 functions to maintain neural progenitor identity. *Neuron*. 2003;39(5):749-765. doi:10.1016/S0896-6273(03)00497-5
15. Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*. 2009;462(7269):108-112. doi:10.1038/nature08460
16. Yin Y, Morgunova E, Jolma A, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science (80-)*. 2017;356(6337).

doi:10.1126/science.aaj2239
17. Zong M, Meng M, Li L. Low expression of TBX4 predicts poor prognosis in patients with stage II pancreatic ductal adenocarcinoma. *Int J Mol Sci*. 2011;12(8):4953-4963. doi:10.3390/ijms12084953
18. Ko E, Kim D, Min DW, Kwon SH, Lee JY. Nrf2 regulates cell motility through RhoA–ROCK1 signalling in non-small-cell lung cancer cells. *Sci Rep*. 2021;11(1):1-9. doi:10.1038/s41598-021-81021-0
19. Berezovsky AD, Poisson LM, Cherba D, et al. Sox2 promotes malignancy in glioblastoma by regulating plasticity and astrocytic differentiation. *Neoplasia*. 2014;16(3):193-206.e25. doi:10.1016/j.neo.2014.03.006
20. Wang J, Xu SL, Duan JJ, et al. Invasion of white matter tracts by glioma stem cells is regulated by a NOTCH1–SOX2 positive-feedback loop. *Nat Neurosci*. 2019;22(1):91-105. doi:10.1038/s41593-018-0285-z
21. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213-219. doi:10.1038/nbt.2514
22. Ramos AH, Lichtenstein L, Gupta M, et al. Oncotator: Cancer variant annotation tool. *Hum Mutat*. 2015;36(4):E2423-E2429. doi:10.1002/humu.22771
23. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499(7457):214-218. doi:10.1038/nature12213
24. Klambauer G, Schwarzbauer K, Mayr A, et al. Cn.MOPS: Mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res*. 2012;40(9):1-14. doi:10.1093/nar/gks003
25. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-2993. doi:10.1093/bioinformatics/btr509
26. Favero F, Joshi T, Marquard AM, et al. Sequenza: Allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol*. 2015;26(1):64-70. doi:10.1093/annonc/mdu479
27. von Ahlfen S, Missel A, Bendrat K, Schlumpberger M. Determinants of RNA quality from FFPE samples. *PLoS One*. 2007;2(12):1-7. doi:10.1371/journal.pone.0001261
28. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359. doi:10.1038/nmeth.1923
29. Dobin A, Davis CA, Schlesinger F, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635
30. Karolchik D, Hinricks AS, Furey TS, et al. The UCSC table browser data retrieval tool. *Nucleic Acids Res*. 2004;32(DATABASE ISS.):493-496. doi:10.1093/nar/gkh103
31. Liao Y, Smyth GK, Shi W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930. doi:10.1093/bioinformatics/btt656
32. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):1-21. doi:10.1186/s13059-014-0550-8
33. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57.

doi:10.1038/nprot.2008.211
34. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37(1):1-13. doi:10.1093/nar/gkn923
35. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550. doi:10.1073/pnas.0506580102
36. McKenzie AT, Wang M, Hauberg ME, et al. Brain Cell Type Specific Gene Expression and Co-expression Network Architectures. *Sci Rep*. 2018;8(1):1-19. doi:10.1038/s41598-018-27293-5
37. Chung JH, Jung HR, Jung AR, et al. SOX2 activation predicts prognosis in patients with head and neck squamous cell carcinoma. *Sci Rep*. 2018;8(1):1-11. doi:10.1038/s41598-018-20086-w
38. Li D, Zhang B, Xing X, Wang T. Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. *Methods*. 2015;72(C):29-40. doi:10.1016/j.ymeth.2014.10.032
39. Xing X, Zhang B, Li D, Wang T. Comprehensive Whole DNA Methylome Analysis by Integrating MeDIP-seq and MRE-seq. In: Tost J, ed. *DNA MEthylation Protocols*. New York, NY: Humana Press, New York, NY; 2018:209-246. doi:https://doi.org/10.1007/978-1-4939-7481-8_12
40. Stevens M, Cheng JB, Li D, et al. Estimating absolute methylation levels at single-CpG resolution from methylation enrichment and restriction enzyme sequencing methods. *Genome Res*. 2013;23(9):1541-1553. doi:10.1101/gr.152231.112
41. Zhang B, Zhou Y, Lin N, et al. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Res*. 2013;23(9):1522-1540. doi:10.1101/gr.156539.113
42. Zhou X, Maricque B, Xie M, et al. The human epigenome browser at Washington University. *Nat Methods*. 2011;8(12):989-990. doi:10.1038/nmeth.1772
43. Li D, Hsu S, Purushotham D, Sears RL, Wang T. WashU Epigenome Browser update 2019. *Nucleic Acids Res*. 2019;47(W1):W158-W165. doi:10.1093/nar/gkz348
44. Lister R, Pelizzola M, Dowen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009;462(7271):315-322. doi:10.1038/nature08514
45. Xie W, Schultz MD, Lister R, et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell*. 2013;153(5):1134-1148. doi:10.1016/j.cell.2013.04.022
46. Khan A, Zhang X. dbSUPER: A database of Super-enhancers in mouse and human genome. *Nucleic Acids Res*. 2016;44(D1):D164-D171. doi:10.1093/nar/gkv1002
47. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317-330. doi:10.1038/nature14248

Chapter 4: Epigenetic reprogramming of brain development pathways during non-small cell lung cancer metastasis to brain

The contents of this chapter have been adapted from the manuscript under consideration:

> "Karlow, J.A., Devarakonda, S., Xing, X., et. al. Epigenetic reprogramming of brain development pathways during non-small cell lung cancer metastasis to brain."

4.1 Abstract

Non-small cell lung cancer (NSCLC) is one of the most commonly diagnosed and deadliest cancer worldwide, owing in large part to a high rate of metastatic disease. While the major events in the metastatic cascade have been identified, a mechanistic understanding of how NSCLC routinely, successfully colonizes the brain is largely unknown. Recent studies have begun demonstrating the role of epigenetic mis-regulation within tumorigenesis as well as in the metastatic process, including widespread changes in DNA methylation and histone modifications. To better understand the role of DNA methylation alterations in NSCLC metastasis to brain, we measured DNA methylation during disease progression for 12 patients, globally profiling the methylation of their normal lung, primary lung tumor, and brain metastasis. We found that the variation in methylation is just a great; albeit less coordinated across genomic features, during metastatic spread as compared to primary tumorigenesis. The greatest recurrent methylation changes during metastatic progression occurred over DNA methylation valleys

(DMVs) harboring H3K9me3 as well as bivalent marks H3K27me3 and H3K4me1 enrichment in normal lung. Mapping EZH2, the catalytic subunit of polycomb repressive complex 2 (PCR2), binding locations in H1299, a lymph node-derived lung cancer cell line, revealed a loss of EZH2 binding within DMVs accompanied by an increase in DNA methylation, exemplifying epigenetic switching. The vast majority of these DMR-associated DMVs harbor developmental genes, suggesting that alteration in epigenetic regulation of developmentally important genes may confer a selective advantage during metastatic progression.

4.2 Introduction

Lung cancer is frequently one of the most highly diagnosed and deadliest cancers globally each year. In 2020, more than 2 million new lung cancer cases were diagnosed and more than 1.7 million deaths attributable, making lung cancer the most deadly cancer and second most newly diagnosed cancer of 2020[1]. Collectively, lung cancer is comprised of two main categories: small cell lung cancer and non-small cell lung cancer (NSCLC), which account for roughly 15% and 85% of all lung cancer cases, respectively[2]. With a 5-year survival rate ranging from 41% - 0% for stage III and IV tumors, advancements in the detection and treatment of lung cancer are eminent[3].

One major contributor to the poor prognosis of the disease is that at time of diagnosis, roughly 46.8% - 61.2% of all NSCLC patients present with both primary and metastatic lesions[4]. Although most cancer-related deaths can be attributed to metastatic spread, the molecular and genetic basis for this process is still largely elusive, limiting treatment options for widespread disease. Fortunately, the largescale steps involved in metastatic dissemination and colonization have been identified[5–7], and major efforts are in place to understand these processes at finer scales. Broadly, primary tumor cells must undergo several changes prior to seeding a secondary

lesion. First, cells must acquire a phenotype capable of detaching from the primary mass, traverse the extracellular matrix, and enter into the bloodstream or lymphatic system. While in the circulatory system, the cells must avoid immune surveillance and the shearing forces of blood flow. The cells must then extravasate the bloodstream, survive in a foreign microenvironment, and possess stem-like properties to allow for the seeing of a secondary lesion. Although many tumor cells may disseminate into the bloodstream, only a small fraction are able to survive the harsh conditions required to form secondary tumors[8–12], suggesting a selective bottleneck where cells successful in seeding the metastasis should be enriched for functional properties that allowed them to survive this journey as compared to their primary relatives.

Metastatic dissemination from primary NSCLC is non-random, most frequently migrating to the brain, but also commonly to the bones, liver, and adrenal glands[13]. Aside from common metastatic characteristics, the specific changes required of NSCLC cells to form brain metastases are largely unknown. Recent studies have identified heightened activation of the PI3K pathway within brain metastases compared to primary tumors[14,15], suggesting possible avenues for tumor colonization of brain. Still, to date, few recurrent genetic mutations have been identified in metastasis[16], and even fewer have been identified that contribute to metastasis in a cell type-specific fashion, and in particular, NSCLC metastasis to brain. On the contrary, several studies have demonstrated metastasis-specific expression signatures, arguing that epigenetic regulation of gene expression is likely involved in the metastatic process. In agreement with this, several classes of metastasis genes exhibit altered DNA methylation and show a corresponding change in expression within metastatic samples, the most classic example being hypermethylation of the *CDH1* promoter and subsequent reduction of expression[16]. In a cohort of 30 patients with matched primary NSCLC tumors and brain metastases, we identified a strong

increase in variant allele frequencies (VAFs) among several classes of epigenetic regulators[17], most notably including *TET1*, suggesting that mis-regulation of DNA methylation may provide a selective advantage in the metastatic colonization of NSCLC in brain.

To better understand how DNA methylation alterations may contribute to NSCLC metastatic colonization of brain, we traced DNA methylation changes during disease progression for 12 patients, globally assessing methylation profiles for their normal lung, primary lung tumor, and brain metastasis. Interestingly, we found that although methylation alterations between normal samples and their matched primary tumors were more coordinated across various genomic features, changes between primary tumors and their matched brain metastases were of a similar scale but exhibited broader variation across individuals. Identification of recurrent alterations during metastatic progression revealed a progressive gain in methylation over DNA methylation valleys (DMVs)[17], particularly those harboring bivalent enhancer marks H3K27me3 and H3K4me1, as well as the constitutive heterochromatin mark H3K9me3. The facultative repressive histone modification H3K27me3 is known to be deposited by EZH2, a member of the polycomb repressive complex 2 (PRC2) protein complex, and its occupancy in CpG islands is mutually exclusive with that of DNA methylation[18,19]. Increased methylation over genomic regions know to contain high levels of PRC2-deposited H3K27me3 suggests then a loss of PRC2 occupancy within DMVs during metastatic progression. To further explore this connection, we globally profiled the binding of EZH2 in a lymph node-derived metastatic lung cancer cell line, H1299. We found that EZH2 binding within DMVs was significantly reduced in the metastatic surrogate compared to normal lung fibroblasts, whereas DNA methylation was increased, exemplifying epigenetic switching within DMVs. Since the vast majority of these differentially methylated region (DMR)-associated DMVs harbor developmental genes, these

data suggest that alterations in the epigenetic regulation of developmentally important genes may confer a selective advantage during metastatic progression.

4.3 Results

4.3.1 DNA methylation changes are more coordinated from normal to primary than from primary to metastasis

Recent comparisons of average methylation across normal lung, primary non-small cell lung cancer (NSCLC) tumors, and matched brain metastases, revealed several genomic features with significant methylation changes during primary tumorigenesis, including promoters, CpG islands, exons, intergenic regions, repeats, DNA methylation valleys (DMVs), and partially methylated domains (PMDs)[17]. In contrast, the only genomic feature to undergo significant methylation changes from primary NSCLC to brain metastases across paired patient samples was DMVs[17]. To further illuminate the large-scale methylation differences between primary NSCLC tumorigenesis and metastatic dissemination to the brain, we characterized methylation changes for the same 12 patient "trios" (sets of matched normal lung, primary lung tumor, and brain metastasis) (Table 4.1, Table 4.2) over normal human lung fibroblast (NHLF) epigenetic states[20]. While significant changes in methylation were observed over many regions baring marks of transcriptional activation, enhancers, repressive polycomb, and quiescent regions during primary tumorigenesis, a significant loss of methylation from primary to metastasis samples was found within regions classified as zinc fingers and repeats, accompanied by a significant progressive loss of methylation over heterochromatic regions (Figure 4.1a). Additionally, regions annotated as bivalent enhancers in NHLF exhibited a significant gain of methylation from normals to primaries and a further gain from primaries to metastases (Figure 4.1a), independent of total methylation change from primary to metastasis (r^2=0.0130, p=0.7226, Linear Regression, Figure 4.2a), and changes in tumor purity (r^2=0.237, p=0.1534, Linear Regression, Figure 4.2b).

Expanding the comparison of methylation values to include all autosomal CpGs revealed a wide degree of variation in methylation shifts throughout disease progression among individual patients (Figure 4.2c, Figure 4.2d). However, significant systematic changes emerged when comparing the density distributions of average autosomal methylation values across individual CpGs through disease states. Specifically, we observed a significant global loss of methylation and significant promoter gain of methylation from normal lung to primary lung tumor, as has been widely reported[21–23], and a subtler significant global loss and slight promoter gain of methylation from primary to metastasis (Figure 4.1b, Figure 4.2d, Figure 4.2e), suggesting that methylation alterations from primary to metastasis may follow similar large-scale trends as primary tumorigenesis, albeit not as pronounced.

To compare the overall similarity of methylation profiles across patients and disease states, we calculated a Pearson correlation score for each pairwise combination of samples, using all autosomal CpG methylation values. All normal samples showed a high degree of correlation, as expected, ranging from 0.84 to 0.90 (Figure 4.1c, Figure 4.2f). Clustering correlation values revealed that two primary samples, P1 and P9, shared a high degree of similarity with normal samples, suggesting that perhaps these tumors retained a high degree of normal epigenetic patterning, consistent with also displaying a low fraction of genome altered in the primary tumor (Figure 4.2G)[17], and independent of tumor purity estimates (Figure 4.2H). Additionally, we found that the primary and metastatic sample from the same individual often clustered together and displayed a higher degree of similarity (median Pearson correlation score of 0.81) than did primary and metastasis samples from different patients (median Pearson correlation score of 0.74) (Figure 4.1C, Figure 4.2F). These observations were also seen when calculating correlation

scores over promoter regions only (Figure 4.2I), suggesting that metastasis samples remain more similar to their primary tumors than tumors from other patients.

To further interrogate the spatial degree of relatedness among all samples, we performed principle component analysis (PCA) using global autosomal methylation values. Considering the first two principle components (PCs), we confirmed that all normal samples grouped together, primary samples dispersed outward, and metastasis samples further dispersed outward (Figure 4.1D). To get a better idea of how "close" different samples' methylation profiles were to each other in high dimensional space, we calculated the Euclidean distance between each pair of points in PCA space, scaling the distance in each dimension (PC) by the percent variance explained in that PC (4.5 Methods). These results indicated that the distances between each primary and metastasis pair were just as great as the distances between each normal and primary pair (Figure 4.1E, no difference between distance distributions, KS-test). These results were also seen when considering only autosomal promoter CpG values (Figure 4.2J).

Taken together, these data suggest that methylation changes between primary NSCLC samples and their matched brain metastases are just as great as between normal lung samples and their NSCLC primaries, they are just less coordinated across genomically- and epigenomically-defined features. However, changes between primary and metastasis do reflect systematic methylation increases, primarily over DMVs.

4.3.2 Regions with recurrent gain of DNA methylation during metastatic spread enrich within DNA methylation valleys and display bivalent polycomb signatures across tissues

Similar to global methylation alterations, the recent identification of differentially methylated regions (DMRs) within patients between primary NSCLC tumor samples and brain metastases revealed considerable inter-patient variability[17] (Figure 4.3A, Figure 4.3B). However, as expected, the number of DMRs between normal samples and their matched metastases was

significantly greater than those between normal samples and their matched primaries or between primaries and metastases of the same patient (Figure 4.3C, Wilcox test). The fraction of DMRs representing regions of methylation gain (hyperDMRs) was significantly lower when comparing normal samples and their matched metastases than when comparing normals and matched primaries (Figure 4.3C, Wilcox tests), indicating that a greater fraction of methylation changes accrued gradually during metastatic spread, as opposed to during primary tumorigenesis, are regions of methylation loss.

Characterization of hyperDMRs revealed similar trends as seen when comparing global methylation profiles over features, most notably including an enrichment within DMVs across disease state comparisons (Figure 4.4A). While not captured when comparing global methylation levels over features[17], primary to metastasis hyperDMRs were also enriched within promoters, CpG islands, exons, and partially methylated domains (PMDs) (Figure 4.4A). We also observe at this resolution that regions flanking transcription start sites, bivalent transcription start sites, bivalent enhancers, and polycomb repressed regions are sites of hyperDMR enrichment throughout disease (Figure 4.4B), suggesting that specific regions flanking transcription, bivalent enhancers, bivalent transcription start sites, and polycomb-repressed regions have large gains in methylation, as opposed to a more globally diffuse increase in methylation over these features, since they were not detected as showing significant methylation gains globally (Figure 4.1A).

While most DMRs were common to only one or two trios, several were observed across many. In each comparison of disease states, a greater number of shared hyperDMRs than shared hypoDMRs was observed (Figure 4.4C). The most recurrent primary to metastasis hyperDMR, present in 6/12 patients, was found within the promoter of *ZNF154*, a known metastasis suppressor gene[24] (Figure 4.4C). To determine if this region could serve as a useful biomarker of

NSCLC disease progression, we examined the methylation status of the region in TCGA LUAD patients. A total of 579 patients were first separated into 4 quantiles based on average methylation over the region (Figure 4.4D). A Kaplan-Meyer curve was then generated, comparing the first and fourth quantiles, revealing a significant difference in survivability between the two groups, where patients with a higher degree of methylation (quantile 4) demonstrated lower survivability (Figure 4.4E). This suggests that methylation over the *ZNF154* promoter may serve as a useful biomarker of survivability, possibly through contributing to metastatic progression.

To optimize locating potential driver DNA methylation changes[16], we focused our analysis on recurrent DMRs across patients. To better understand possible DMR regulatory functions, we calculated gene ontology enrichment of hyperDMRs and hypoDMRs from primary to matched metastasis samples shared by at least two patients (n=3037 and n=1869, respectively)[25]. Recurrent hyperDMRs enriched most notably for patterning processes and embryonic developmental processes, while recurrent hypoDMRs enriched for kidney development terms (Figure 4.3D). Similar processes were also observed when calculating enrichment for each patient's DMRs individually (Appendix C.1). To gain a more comprehensive picture of the key genes that may be influenced by these shared DMRs, we identified enriched transcription factor binding motifs (TFBMs). Specifically, we identified 49 TFBMs enriched within primary to metastasis hyperDMRs shared by at least 2 patients, of which the strongest enrichment was for ZBTB14, a positive regulator of neuronal development[26] (Appendix C.2). Of the 49 genes with enriched motifs, 3 had significant expression changes from primary to metastasis: *WT1* and *EGR2* (decrease) and *ZIC1*, regulator of neurogenesis (increase). Analysis of the 49 genes with significantly enriched motifs revealed most notably enriched

processes in transcription, as well as cancer-related pathways and the HTLV-1 infection pathway (Appendix C.2). When considering primary to metastasis hypoDMRs shared by at least 2 patients, we found 215 enriched motifs, of which the strongest enrichment was for ZNF770. Of these 215 TFs with enriched motifs, 3 had significant expression changes from primary to metastasis, all of which showed decreased expression: *MAF*, *NFATC1*, and *ATF3*. Analysis of the 215 genes with significantly enriched motifs revealed most notably enriched processes in transcription, development, and differentiation (Appendix C.2).

When considering primary to metastasis hyperDMRs that were shared by at least 3 patients (n=737), we found that many regions were associated with bivalent transcriptional start sites, bivalent enhancers, and repressed polycomb regions across other tissue types, particularly in developmental, epithelial, and brain tissues (Figure 4.3E, Figure 4.4F, 4.5 Methods); whereas hypoDMRs shared by at least 3 patients (n=251) often shared similar annotations across different samples as those in normal lung (Figure 4.3F, Figure 4.4G).

Taken together, these results indicate that regions with gain of methylation from primary to metastasis are more often common across multiple patients, whereas loss of methylation events are typically private to the individual or shared by only a few. Furthermore, the gain of methylation events are non-randomly distributed throughout the genome, most highly enriching with DMVs and regions annotated as possessing bivalent or polycomb signatures in external tissues.

4.3.3 Preferential gain of methylation over DNA methylation valleys containing bivalent enhancer marks during disease progression

The regions with the greatest recurrent gain of methylation from primary to metastasis within patients were within DNA methylation valleys (DMVs) (Figure 4.4A). While the degree of methylation change over DMVs varied across patients (Figure 4.5A), a significant increase in

DMV methylation was found from both normal to primary and primary to metastasis (Figure 4.5B and[17]). To better understand how methylation changes within these DMVs may impact metastatic progression, we characterized all genes falling within DMVs. In total, 735 genes overlapped autosomal DMVs, many of which are related to developmental and brain-related functions (Appendix C.3). The DMV with the greatest average gain in methylation from primary to metastasis overlaps *PAX6*, a gene known to play an important role in central nervous system development[27], whose increased promoter methylation status has been linked to poor outcomes in NSCLC patients[28]. We see that this DMV is relatively unmethylated across all normal lung samples (by definition), shows a gain of methylation within most of the primary samples, and a further gain across many of the metastatic samples (Figure 4.6A). NHLF DNA methylation data confirms that this region is typically unmethylated in normal lung; however, this region shows increased methylation in the lymph node-derived lung cancer cell line H1299, as well as in brain hippocampus, suggesting that perhaps methylation of this region is inherent to brain cell lineages and is mimicked during tumor progression (Figure 4.6A).

To better understand the possible functional consequences of increased methylation within DMVs throughout disease progression, we characterized these regions further by identifying average enrichment profiles of various histone modifications over DMVs and surrounding 2kb regions in normal lung, using Roadmap data for NHLF[20]. DMVs demonstrated significantly higher H3K4me1, H3K9me3, H3K4me3, H3K27ac, and H3K27me3 enrichments scores, and lower H3K36me3 enrichments scores, compared with background 5 kb genomic regions in normal lung (Figure 4.6B, Figure 4.6C). Since many of these histone modifications play opposing roles in regulating gene expression[29], we wanted to better understand the connection between the average histone modification values over a DMV in a normal lung

context and the DMV's average DNA methylation change across disease states. Our results show that the gain in methylation over DMVs throughout disease progression is driven by a gain in methylation over only a subset of the DMVs, as several remain unmethylated across samples (Figure 4.5C). Interestingly, the average change in methylation across patients from primary to metastasis for each DMV positively correlated with H3K27me3, H3K4me1, and H3K9me3 histone marks, and negatively correlated with H3K27ac and H3K4me3 marks in normal lung, revealing similar trends as those seen comparing normal lung and primary tumors (Figure 4.6D). This suggests that the DMVs that are progressively gaining methylation are not regions of active transcription in normal lung, but instead harbor repressive marks H3K27me3 and H3K9me3 as well as the active enhancer mark H3K4me1, indicative of bivalent enhancers. This is confirmed by examining the expression of genes falling within DMVs in normal lung; those with a greater gain of methylation from primary to metastasis are significantly more lowly expressed in normal lung (Figure 4.5D).

Clustering patients based on their average change in methylation over each DMV from primary to metastasis revealed that patients distinctly fell into two groups, one with a more severe gain in methylation over DMVs from primary to metastasis and one with a less severe gain (Figure 4.6E). While not meeting significance likely due to small sample sizes, patients in the greater DMV methylation gain group tended towards shorter progression-free survival times than patients demonstrating less DMV methylation gain (4 patients with less than 1 year regression free survival and 1 patient with greater than 1 year regression free survival in the larger DMV methylation gain group compared to 1 and 4 patients, respectively, in the lesser DMV methylation gain group, p=0.1, Fisher's exact test). This two group clustering partition is identical to the groupings when clustering patients based on their degree of lung bivalent

enhancer methylation change (Figure 4.5E). Mutation profiles for these patients support a correlation between increased genetic mutation burden / load on polycomb repressive complex 1 and increased DMV hypermethylation (Figure 4.6F).

Possessing both repressive and active histone modifications indicates that these regions likely contain bivalent or poised regulatory elements, which is consistent with the observation that lung bivalent enhancers exbibit increased methylation in metastasis (Figure 4.4B)[29]. Supporting this connection, we found that a significant proportion of lung bivalent enhancers were encompassed within DMVs and preferentially overlapped DMVs with a greater increase in methylation from primary to metastasis (Figure 4.5F, Figure 4.5G). Correspondingly, we also observed a high correlation (r^2=0.941) between the average change in DMV methylation and in lung bivalent enhancer methylation from primary to metastasis across patients (Figure 4.5H).

4.3.4 DNA methylation valley (DMV) gain of methylation is associated with loss of polycomb repressive complex 2 occupancy

The repressive histone modification, H3K27me3, is deposited by the multi-protein complex polycomb repressive complex 2 (PRC2)[30], and has demonstrated mutual exclusivity with DNA methylation at CpG-rich promoters[31]. To determine if increased methylation within DMVs throughout disease progression was associated with loss of polycomb occupancy, we profiled EZH2, the catalytic subunit of PRC2, binding in the lymph node-derived non-small cell lung cancer cell line H1299 and compared with EZH2 occupancy in normal lung (NHLF)[32]. All sequencing data met the Transcription Factor ChIP-Seq Data Standards put forth by ENCODE (Appendix C.4)[33]. A global reduction in EZH2 occupancy was observed in H1299 relative to NHLF, with 4448 sites (90.57%) with a loss of binding and 463 sites (9.43%) with a gain of binding (Figure 4.7A). The most significant loss of EZH2 binding in H1299 was over the promoter of *CBX8*, which demonstrated a substantial concordant increase in DNA methylation,

as well as increased expression (Figure 4.7B). In contrast, the greatest gain of EZH2 fold change resided upstream the *IGHV2OR16-5* promoter, concordantly displaying a loss of DNA methylation in H1299 (Figure 4.7C). Enriched processes of lost EZH2 peaks consisted overwhelmingly of brain and developmental terms (Figure 4.7D), absent in gain of EZH2 peak enrichment, displaying enriched terms exclusively related to the epoxygenase P450 pathway (Figure 4.7E).

We found that only a subset of DMVs demonstrated EZH2 binding enrichment in NHLF, which was significantly reduced in H1299 and corresponded with an increase in DNA methylation (Figure 4.8A, Figure 4.8B). The negative correlation between change in DNA methylation and change in EZH2 binding from NHLF to H1299 was significantly greater over DMVs than background regions (Figure 4.8C), suggesting that DMVs are locations of targeted epigenetic switching in the metastasis surrogate.

4.4 Discussion

Non-small cell lung cancer regularly claims the majority of cancer-related deaths, owing in large part to a high prevalence of metastasis, most frequently disseminating to the brain. Although significant findings have contributed to our understanding of the metastatic process, many cell-type-specific influences are currently unknown. In this study, we characterized methylation patterns for 12 patients' normal lung, primary lung tumor, and brain metastasis, and identified regions of recurrent methylation alterations during the metastatic process. We found that although methylation changes between normal and matched primary samples were more coordinated than those between matched primary and metastasis samples over various genomic features across patients, the methylation differences between matched primary and metastasis samples were just as great. This suggests that while methylation alterations are just as common

in the process of metastatic disease as in primary tumorigenesis, a greater fraction of these methylation changes during metastatic disease may be passenger events.

The most recurrent single region methylation change between primary and metastasis samples, observed in 6 out of 12 cases, was an increase in methylation within a region overlapping the *ZNF154* promoter. Consistent with previous reports associating *ZNF154* promoter methylation with various different cancers, we found that the top and bottom quantiles of TCGA lung adenocarcinoma patients, stratified by average *ZNF154* promoter methylation level, displayed a significant difference in survivability, where patients with a greater level of methylation over the promoter exhibited a worse prognosis. This data suggests that recurrent methylation patterns identified in metastatic samples may be useful biomarkers of prognosis, as well as potential contributors to disease progression, although additional functional studies are needed to determine the latter.

Identification of recurrent differentially methylated regions revealed a strong pattern of methylation gain throughout disease progression within regions defined as DNA methylation valleys (DMVs) and bivalent enhancers in normal lung tissue. Upon further investigation of the DMV methylation gain, it became clear that the gain in methylation observed over DMVs was driven by a subset of the DMVs, as many of them did not accrue methylation changes. Comparisons of the DMVs that underwent methylation changes as opposed to those that remained consistently unmethylated revealed a strong bias in histone modifications. The increase in methylation over DMVs was positively correlated with the bivalent histone modifications H3K27me3 and H3K4me1, suggesting that bivalent enhancers within DMVs may be the primary sites of methylation gain. Examining these regions in normal lung tissue revealed high H3K27me3, consistent with polycomb repression. Therefore, one possibility is that polycomb

repressive complexes repress genes within DMVs in normal lung, and through disease progression, this repression is lost. As a result of the loss of polycomb occlusion, DNA methylation accumulates. Interestingly, the degree of methylation gain over DMVs varied across patients, clearly separating patients into two groups. Although sample size limitations occlude the identification of a robust statistical difference in patient survivability within the two groups, we did find that patients with a higher degree of methylation gain within DMVs trended towards a shorter regression free survival time than those with a lower DMV methylation average. While larger sample sizes are need to verify this trend, our data are consistent with the hypothesis that increased DMV methylation accompanying loss of PRC2 binding contributes to the ability of primary lung tumor cells to successfully colonize brain tissues. Since genes within DMVs are largely related to development, we propose that loss of polycomb repression over these regions may lead to an increase in developmental gene expression, resulting in a phenotype shift towards a less differentiated state, potentially allowing tumor cells to succeed in metastatic dissemination through changes in their cellular identity.

4.5 Methods

4.5.1 Global DNA methylation profiling of samples

In order to unveil methylation changes during disease progression from primary NSCLC to brain metastasis, we obtained DNA samples of normal lung, primary lung tumor, and brain metastasis from formalin-fixed paraffin-embedded (FFPE) tissue blocks of 11 patients and fresh/frozen tissue from 1 patient. For each sample, we performed global methylation profiling using Methylation Dependent Immuno-Precipitation followed by sequencing (MeDIP-seq) and Methylation-sensitive Restriction Enzyme digestion followed by sequencing (MRE-seq) as previously described[34]. We also performed whole genome bisulfite sequencing (WGBS) for the fresh/frozen samples to validate candidate methylation alterations (data not shown). All samples passed our sequencing depth threshold requirements (Table 4.2). MRE-seq data were processed by first merging all reads from different sequencing lanes for a given sample. We estimated single CpG-resolution methylation values for all MeDIP-seq + MRE-seq samples using methylCRF[35] (where -c was either set to 0 or 3 depending on sequencing), and visualized our data on the Washington University Epigenome Browser[36,37].

4.5.2 Calculating partially methylated domains (PMDs)

Partially methylated domains (PMDs) were calculated based on a previously described method[38]. In short, the genome was divided into 10 kb bins, and the average methylation per 10 kb bin was calculated for each of the 12 normal lung samples, omitting black-listed CpGs. 10 kb bins with an average methylation less than 70% in all normal samples and containing at least 10 CpGs were called PMDs. Adjacent PMDs were merged.

4.5.3 Calculating DNA methylation valleys (DMVs)

DNA methylation valleys (DMVs) were calculated as previously described[39]. In short, the genome was divided into 1 kb bins, and the average methylation per 1 kb bin was calculated for each of the 12 normal lung samples. A sliding window was then applied to calculate the average methylation across 5 kb windows, incrementing 1 kb at a time. Average methylation across 5 consecutive 1 kb bins was calculated, and regions with less than or equal to 15% methylation were classified as DMVs. Consecutive DMVs were then merged.

4.5.4 Calculating PCA distances of samples, scaled by variance

To obtain a measurement of DNA methylation similarity between pairs of samples, we calculated the Euclidean distance between samples in PCA space, scaled by the percent variance explained in each PC:

$$D = \sqrt{\sum_{i=1}^{l} [(x_{i2} - x_{i1}) * v_i]^2}$$

Where: D = distance; l = number of PCs; x_{i2} = position in PC space of sample 1 in the ith PC; x_{i2} = position in PC space of sample 2 in the ith PC; v_i = percent variance explained in the ith PC.

4.5.5 Processing data for input to MnM and methylCRF

4.5.5.1 Processing MeDIP data

Starting with raw fastq.gz files of paired-end sequencing results for each sample, cutadapt (v:1.10) was used to remove adapter sequences (adapter 1: AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC; adapter 2: AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT), first removing all low-quality bases from the 3' and 5' read ends using a quality cutoff of 10. Resulting reads less than 20 base pairs long were discarded. Next, reads were aligned to the hg19 genome using bwa mem (v:0.7.15) (contigs and haplotypes were excluded). Resulting SAM files were converted to BAM files and coordinate-sorted using samtools (v:1.3.1). BAM files corresponding to the same sample, just from different sequencing lanes, were merged using samtools (v:1.3.1). Finally, methylQA[34] (v:0.1.6) was used to obtain further sequencing statistics and produce BigWig and bedGraph files from the processed BAM files.

4.5.5.2 Processing MRE data

Starting with raw fastq.gz files of single-end sequencing results for each sample, cutadapt (v:1.10) was used to remove adapter sequences (AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC), first removing all low-quality bases from the 3' read ends using a quality cutoff of 10. Resulting reads less than 20 base pairs long were discarded. Next, reads were aligned to the hg19 genome using bwa mem (v:0.7.15) (contigs and haplotypes were excluded). Resulting SAM files were converted to BAM files and coordinate-sorted using samtools (v:1.3.1). BAM files corresponding to the same sample, just

from different sequencing lanes, were merged using samtools (v:1.3.1). Finally, methylQA[34] (v:0.1.6) was used with the appropriate -c flag (1 or 4, depending on sequencing) to obtain further sequencing statistics and produce BigWig and bedGraph files from the processed BAM files.

4.5.6 Calculating differentially methylated regions (DMRs) using MnM

To define differentially methylated regions (DMRs) across pairs of samples, we ran MnM with default parameters, as previously described[34], using MeDIP-seq and MRE-seq data from each sample as input, excluding DAC blacklisted CpG regions defined by ENCODE (accession: wgEncodeEH001432)[32]. This allowed for the identification of 500 bp regions that were differentially methylated. MnM.selectDMR() was run 4 different times, varying the q.value parameter (0.01 (default), 0.001, 0.0001, 0.00001), producing lists of DMRs per pairwise comparison across different stringency thresholds (Figure 4.9A). DMRs were also calculated for each normal vs normal comparison across all stringency thresholds, to serve as a measurement of DMR "false-positives". For each patient, we calculated the average number of DMRs called between their normal sample and all other normal samples. We then calculated the ratio of that value to the number of normal vs. primary DMRs for that patient and to the number of normal vs. metastasis DMRs for that patient, to measure the ratio of "false-positives" to "true-positives". We found that apart from the patient 9 normal to primary comparison, all false-positive to true-positive ratios were relatively low, indicating a greater proportion of DMRs across disease statuses than across normal samples (Figure 4.9B). For patient 9, few DMRs were called between the normal and primary samples, causing the false-positive to true-positive ratio to appear large. In general, we found that as the stringency threshold for calling DMRs increased, the false-positive to true-positive ratio decreased, suggesting that at higher stringency thresholds, fewer normal vs normal DMRs are detected relative to DMRs detected across disease states (Figure 4.9B). Since the false-positive to true-positive ratio continued to fall across our thresholds, we chose to proceed with a DMR stringency threshold of q=0.00001, the smallest of the four values we tested. We did not test stringency thresholds less than q=0.00001 since the number of DMRs retained was already quite low for some samples at this threshold.

4.5.7 Survival analysis of TCGA LUAD patients

Array-based methylation data was downloaded from https://gdc.cancer.gov for The Cancer Genome Atlas (TCGA) lung adenocarcinoma (LUAD) cohort (663 patients total). The average methylation over probes was taken for any patients for which multiple datasets were collected. Patients were then stratified into four quantiles based on their average methylation over the genomic region: chr19:58220000-58220500, corresponding to the most recurrent DMR across patients when comparing paired primary and metastatic samples. The survivability of patients in the top and bottom quantiles were then compared.

4.5.8 Identification of enriched motifs

Analysis of Motif Enrichment (AME)[40], version 5.0.5 from the MEME Suite, was used to identify enriched motifs. To identify motifs enriched within shared primary to metastasis hyperDMRs, autosomal primary to metastasis hyperDMRs shared in at least 2 patients (n=3037) were used as input while autosomal primary to metastasis hypoDMRs shared in at least 2 patients (n=1869) were used as control sequences. Likewise, to identify enriched motifs within shared primary to metastasis hypoDMRs, autosomal primary to metastasis hypoDMRs shared in at least 2 patients were used as input while autosomal primary to metastasis hyperDMRs shared in at least 2 patients were used as control sequences. The HOCOMOCOv11 database was used in both cases, along with default parameters.

4.5.9 Calculating percentage methylation over various epigenetic chromatin states defined in normal lung across disease states

200 bp genomic regions were previously classified as belonging to one of 18 possible chromatin states in normal human lung fibroblasts[20]. The average methylation across all normal, primary, and metastasis samples calculated for each chromatin state defined in normal lung, and paired Wilcoxon tests were performed to determine if there were any significant changes between normals and primaries, primaries and metastases, and normal and metastases for each chromatin state. Corrected p-values (Benjamini-Hochberg) are reported.

Promoter, 5'UTRs, CGIs, all exons, coding exons, 3'UTRs, introns, intergenic regions, and repeats were defined using the NCBI RefSeq annotations for hg19 downloaded from the UCSC genome browser. Promoter regions were defined as +1.5 kb upstream the transcription start site (TSS) to -.5 kb downstream the TSS. Partially methylated domains (PMDs) and DNA methylation valleys (DMVs) were calculated as described above.

4.5.10 Calculating enrichment of DMRs within genic features

Genic features were defined using the NCBI RefSeq annotations for hg19 downloaded from the UCSC genome browser. For each pair of samples the percentage of autosomal hyperDMRs and hypoDMRs overlapping each genic feature was calculated. As long as any portion of the DMR overlapped to the feature, it was counted, allowing DMRs to belong to more than one genic feature. Background regions were defined separately for each pair of samples and consisted of autosomal 500 bp regions that contained at least 1 MeDIP-seq read and/or 1 MRE-seq read in both samples and that did not contain any blacklisted CpGs (Duke Excluded and DAC UltraHighSignal blacklist downloaded from the UCSC genome browser). The percentage of background bins that overlapped to each feature was also calculated (again, counting a bin if any portion overlapped to the feature). Enrichment scores were calculated by dividing the percentage of DMRs overlapping to the feature by the percentage of background bins overlapping the feature.

4.5.11 Calculating enrichment of DMRs within various chromatin states defined in normal lung

18 chromatin states globally defined by ChromHMM for normal human lung fibroblast tissue (E096) were obtained from the Roadmap Epigenomics Consortium[20]. For each pair of samples, the percentage of autosomal hyperDMRs and hypoDMRs overlapping each of the 18 chromatin states was calculated, again counting a DMR if any portion overlapped to the state. The percentage of background bins (defined above) that overlapped to each chromatin state was also determined, and enrichment scores were calculated by dividing the percentage of DMRs overlapping to the chromatin state by the percentage of background bins overlapping the state.

4.5.12 Calculating histone modification scores in DMVs and background regions

Whole genome maps of histone modification fold change over input signal were downloaded from Roadmap for sample E096 (normal lung), for histone modifications: H3K27ac, H3K27me3, H3K4me1, H3K4me3, H3K9me3, and H3K36me3. Background genomic bins were defined as 5 kb autosomal bins that did not contain any blacklisted CpGs and did not overlap with DMVs. The average histone modification fold change over input was then calculated for all autosomal DMVs and background bins for each specific modification.

4.5.13 H1299 ChIP-seq data generation and processing

We cultured H1299 cells in RPMI1640 medium (Gibco, 11875-085) supplemented with 10% fetal bovine serum (Corning, 30-011-CV) and 100 U ml-1 penicillin-streptomycin (Bibco, 15140-122), as previously described in[41]. To generate ChIP-seq probing EZH2 locations in H1299, we obtained the monoclonal, anti-rabbit EZH2 antibody, (5246) Ezh2 (D2C9) XP Rabbit mAb, from Cell Signaling Technology. Western blot confirmed successful enrichment of EZH2 in our H1299 cells using the specified antibody. The SimpleChIP Plus Enzymatic Chromatin IP Kit (Magnetic Beads) #9005 from Cell Signaling Technology was then used, following the manufacturer's recommendations.

Two replicates of EZH2 ChIP-seq and control (input) were generated for H1299 using Illumina paired-end sequencing. All Fastq files were individually quality-assessed using FastQC v0.11.5 and default parameters. Multiqc v1.9 was then run to generate a merged report file (Appendix C.4), which demonstrated that all samples exhibited high per-base sequencing quality. After confirming high quality reads, any adapter sequences and low quality read ends were trimmed using cutadapt v1.10 (with parameters: -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -q 20,20 --minimum-length 20). Trimmed reads were then input to ENCODE's Transcription Factor ChIP-seq Processing Pipeline v2, and resulting conserved IDR peaks and individual replicate fold enrichment files were used in downstream analyses. In addition, *z*-scored fold enrichment files were created for each sample, individually to account for differences in global binding affinity of EZH2.

4.5.14 NHLF ChIP-seq data processing

Two replicates of NHLF ChIP-seq and control were generated by the ENCODE consortium, and their aligned BAM files were downloaded from http://hgdownload.soe.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/. BAM files were then input to ENCODE's Transcription Factor ChIP-seq Processing Pipeline v2, and resulting conserved IDR peaks and individual replicate fold enrichment files were used in downstream analyses. In addition, *z*-scored fold enrichment files were created for each sample, individually to account for differences in global binding affinity of EZH2.

4.5.15 Identification of differentially bound ChIP-seq peaks

EZH2 ChIP BAM files for both replicates of H1299, NHLF, and their respective controls were used as input to DiffBind to detect differential binding locations of IDR conserved peaks, using default parameters. Differentially bound locations with an FDR-corrected value of < 0.05 were retained for downstream analyses.

4.5.16 Additional dataset sources

Exome-seq mutations for a subset (10/12) of patients and RNA-seq data for all samples were obtained from[17]. H1299 whole genome bisulfite sequencing data was obtained from[42]. NHLF whole genome bisulfite sequencing data was obtained from[32].

4.6 Figures

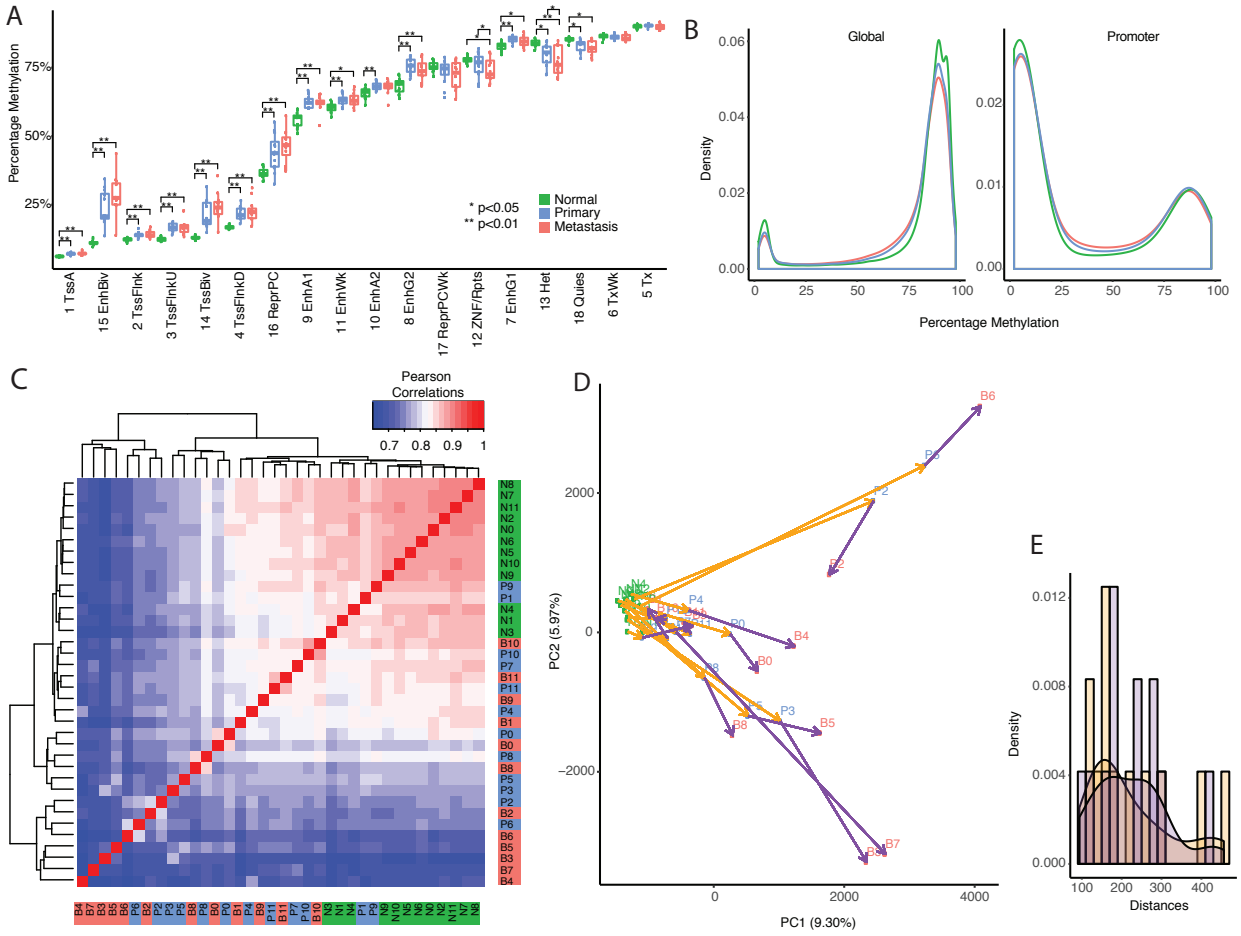


Figure 4.1. DNA methylation changes are more coordinated from normal to primary than from primary to metastasis.

A. Average DNA methylation across Roadmap-defined normal human lung fibroblast epigenetic states (18-state chromHMM model) in normal, primary, and metastasis samples. All within epi-state comparisons were made (paired Wilcoxon t-tests) and only those with a Benjamini Hochberg-corrected *p*-value of less than 0.05 are reported. **B**. Density plots depicting the methylation values across each CpG (all autosomal CpGs: left, all autosomal promoter CpGs: right), averaged and grouped by disease state. **C**. Pearson correlations of all autosomal methylation values for all pairwise combinations of samples. Disease status is indicated by color (normal lung = green, primary NSCLC = blue, brain metastasis = red). **D**. Principal component (PC) analysis of global autosomal methylation values for all samples. The first two PCs are shown. Orange arrows connect normal lung samples to matched primary NSCLC samples, and purple arrows connect primary NSCLC samples to matched brain metastases. **E**. Density plot depicting distances between normal lung and matched primary NSCLC samples (orange) and between primary NSCLC samples and matched brain metastases (purple). Distances are Euclidean distances scaled by the percent variance explained in a given PC.

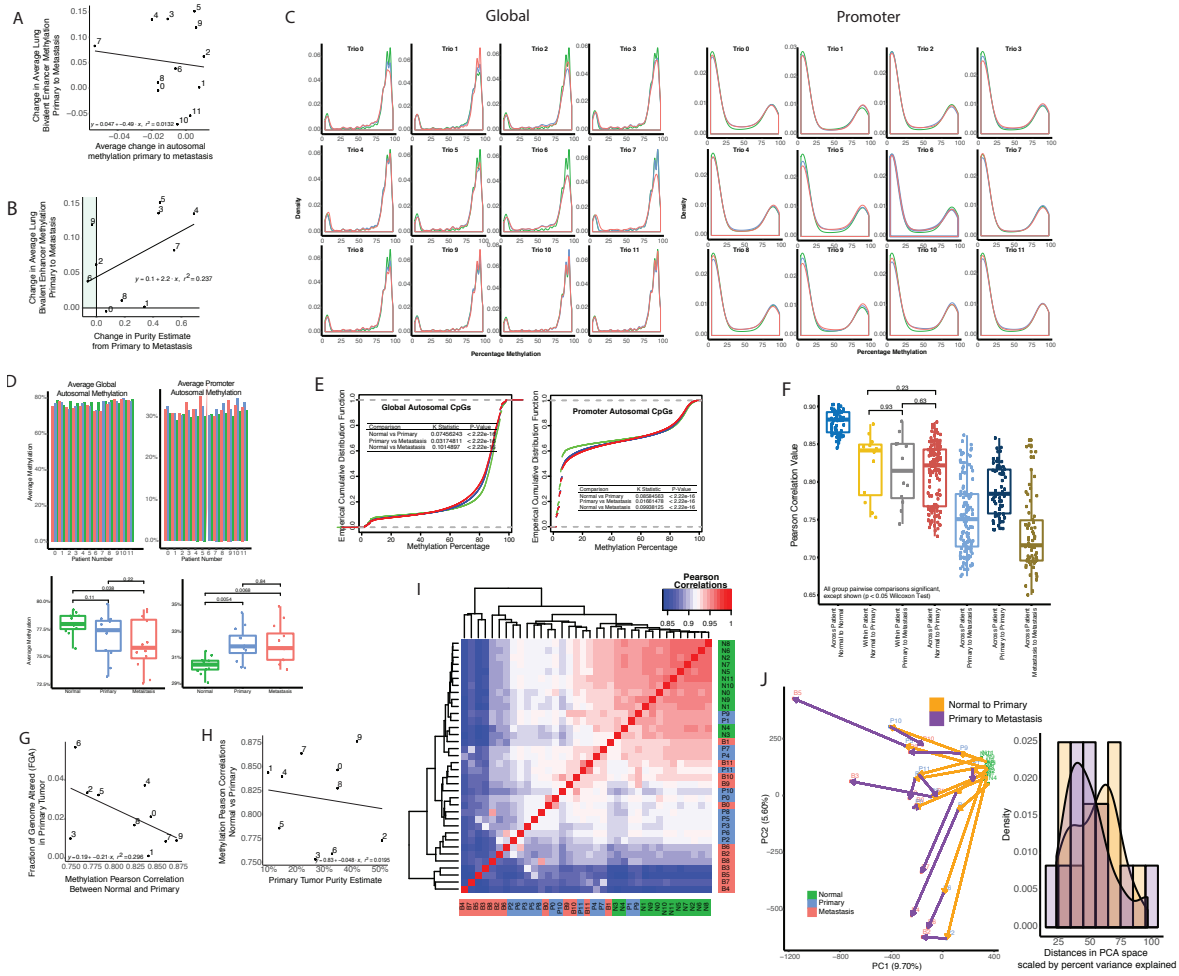


Figure 4.2. DNA methylation changes are more coordinated form normal to primary than from primary to metastasis (Supplement).

A. Average change in autosomal methylation from primary to metastasis as a function of change in average lung bivalent enhancer methylation from primary to metastasis (linear regression, p=0.7226). **B**. Average change in tumor purity (metastasis – primary) as a function of change in average lung bivalent enhancer methylation from primary to metastasis (linear regression, p=0.1534). Green vertical strip highlights samples with a loss of purity in metastasis and an increase in bivalent enhancer methylation. **C**. Methylation density profiles for each patient considering all autosomal CpGs (left) and all autosomal promoter CpGs (right). Color key is indicated in panel **J**. **D**. Top: average global autosomal methylation (left) and average promoter autosomal methylation (right) for each sample from each patient. Bottom: statistical comparison of average global autosomal methylation (left) and average promoter autosomal methylation (right) across disease states, paired-Wilcoxon test values are reported. **E**. Empirical cumulative distribution functions for average global autosomal methylation (left) and promoter autosomal methylation (right), accompanied by k-statistics and p-values from described Kolmogorov-Smirnov tests. **F**. Comparisons of Pearson correlation value distributions. All group pairwise comparisons are sigifcant, except those shown (p<0.05, Wilcoxon Tests). **G**. Fraction of genome altered (FGA) in primary tumors versus the global autosomal methylation Pearson correlation between normal and primary samples from the same patient. Primary tumors P1 and P9

correspond to IDs 19 and 24 in[17]. **H**. Global autosomal methylation Pearson correlations between paired normal and primary samples vs primary tumor purity esitmates. **I**. Pearson correlations of all promoter autosomal methylation values for all pairwise combinations of samples. Disease status is indicated by color (normal lung = green, primary NSCLC = blue, brain metastasis = red). **J**. Left: Principal component (PC) analysis of promoter autosomal methylation values for all samples. The first two PCs are shown. Orange arrows connect normal lung samples to matched primary NSCLC samples, and purple arrows connect primary NSCLC samples to matched brain metastases. Right: Density plot depicting distances between normal lung and matched primary NSCLC samples (orange) and between primary NSCLC samples and matched brain metastases (purple). Distances are Euclidean distances scaled by the percent variance explained in a given PC.

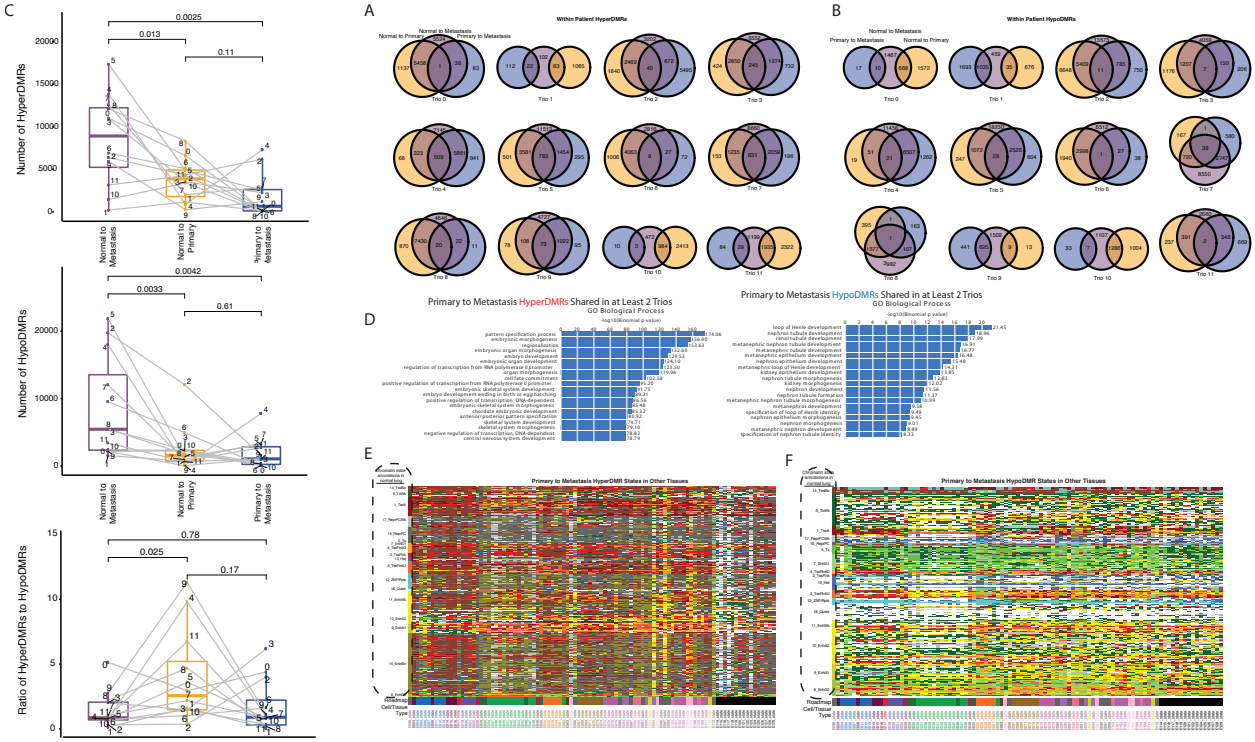
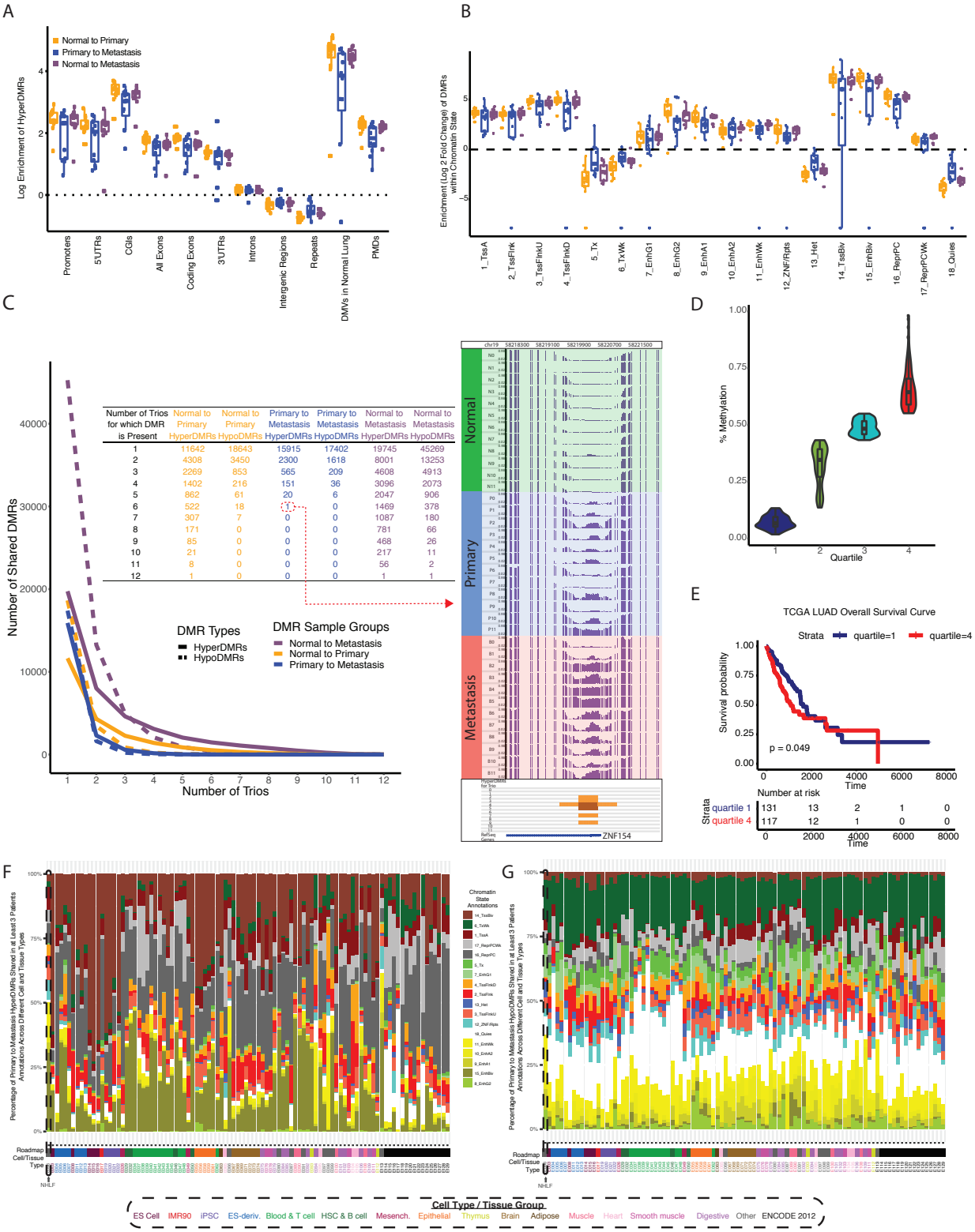


Figure 4.3. Regions with recurrent gain of DNA methylation during metastatic spread enrich within DNA methylation valleys and display bivalent and polycomb signatures across tissues (Supplement).

A. Differences in the number of DMRs across disease comparisons (top: hyperDMRs, middle: hypoDMRs, bottom: ratio of hyperDMRs to hypoDMRs). **B**. Venn diagrams displaying the overlap between within patient hyperDMRs across disease state comparisons for each patient. **C**. Venn diagrams displaying the overlap between within patient hypoDMRs across disease state comparisons for each patient. **D**. Enriched gene ontology for primary to metastasis hyperDMRs shared in at least 3 patient trios (left) and primary to metastasis hypoDMRs shared in at least 3 patient trios (right). **E**. Epigenetic state classifications of recurrent primary to metastasis hyperDMRs (shared in at least 2 patients) across all Roadmap cell and tissue types[20]. **F**. Epigenetic state classifications of recurrent primary to metastasis hypoDMRs (shared in at least 2 patients) across all Roadmap cell and tissue types[20].



Number of Trios for which DMR is Present	Normal to Primary HyperDMRs	Normal to Primary HypoDMRs	Primary to Metastasis HyperDMRs	Primary to Metastasis HypoDMRs	Normal to Metastasis HyperDMRs	Normal to Metastasis HypoDMRs
1	11642	18643	15915	17402	19745	45269
2	4308	3450	2300	1618	8001	13253
3	2269	853	565	209	4608	4913
4	1402	216	151	36	3096	2073
5	862	61	20	6	2047	906
6	522	18	7	0	1469	378
7	307	7	0	0	1087	180
8	171	0	0	0	781	66
9	85	0	0	0	468	26
10	21	0	0	0	217	11
11	8	0	0	0	56	2
12	1	0	0	0	1	1

Figure 4.4. Regions with recurrent gain of DNA methylation during metastatic spread enrich within DNA methylation valleys and display bivalent and polycomb signatures across tissues.

A. Enrichment (log) of hyperDMRs in various genomic regions, separated according to disease class. **B**. Enrichment (log2 fold change) of hyperDMRs in various lung-defined chromatin states (E096)[20], separated according to disease class. **C**. Left: Depiction of the number of DMRs shared across numbers of patient trios (left) and Epigenome Browser[36,37] view of the top shared primary to metastasis hyperDMR, overlapping the *ZNF154* promoter (shared in 6/12 patients) (right). **D**. Violin plot displaying the average methylation over the *ZNF154* promoter in a cohort of 579 TCGA LUAD patients, split into four quantiles. DNA methylation values are calculated based on available 450k and 850k array data. **E**. Kaplan Meier curve illustrating the relative survival differences in TCGA LUAD patients based on average methylation over the *ZNF154* promoter. The top and bottom quantiles are compared. **F**. Epigenetic state classification percentages of recurrent primary to metastasis hyperDMRs (shared in at least 3 patients) across all Roadmap cell and tissue types[20]. **G**. Epigenetic state classification percentages of recurrent primary to metastasis hypoDMRs (shared in at least 3 patients) across all Roadmap cell and tissue types[20].

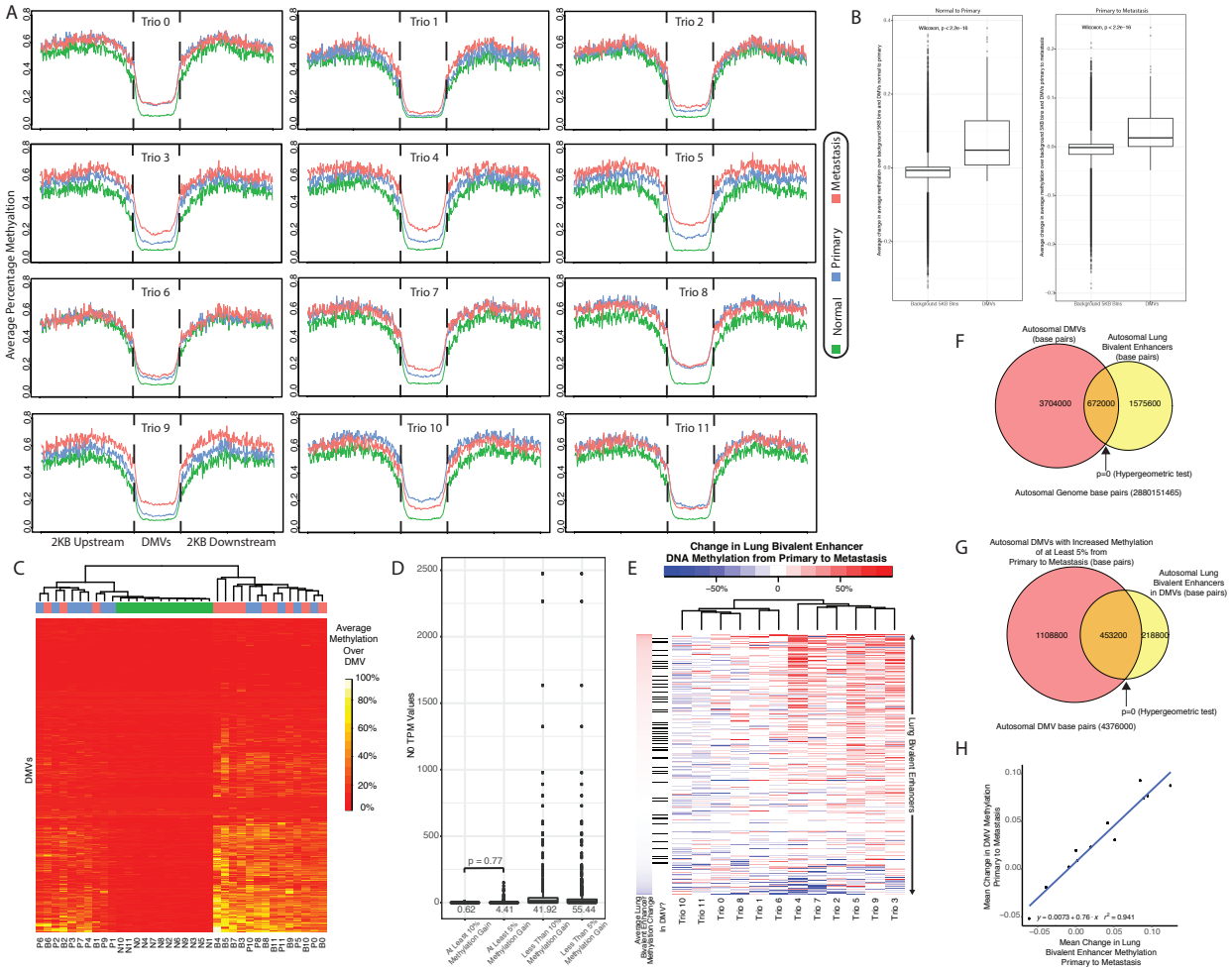


Figure 4.5. Preferential gain of methylation over DNA methylation valleys containing bivalent enhancer marks during disease progression (Supplement).

A. Aveage methylation profiles over all autosomal DMVs and surrounding 2 kb regions for each patient, seperated by disease state. **B**. Distribution comparisons between the average change in average methylation over background 5 kb regions compared to the average change in average methylation over DMVs (Left: normal to primary; Right: primary to metastasis). **C**. Average methylation across each DMV for each sample, colored at the top according to disease state. **D**. Comparison of the expression of DMV genes in NHLF, stratified according to the degree of average methylation change exhibited over the DMV from primary to metastasis samples. All pairwise comparisons are significant (p<0.05, Wilcox test) unless otherwise indicated. **E**. Heatmap displaying the average change in methylation over each lung (NHLF) bivalent enhancer (defined according to Roadmap[20]) from primary to metastasis. Patients are hierarchically clustered and lung bivalent enhancers are ordered according to their average change in methylation. An additional side bar is provided displaying whether or not each enhancer resides within a DMV. **F**. Venn diagram displaying the significant genomic overlap between autosomal DMVs and autosomal lung (NHLF) bivalent enhancers. **G**. Venn diagram displaying the significant specific overlap between autosomal DMVs with an average increased methylation of 5% or greater from primary to metastasis with autosomal lung (NHLF) bivalent enhancers,

relative to all DMVs. **H**. Scatterplot displaying the average change in DMV methylation from primary to metastasis vs the average change in lung (NHLF) bivalent enhancer methylation change from primary to metastasis for each patient.

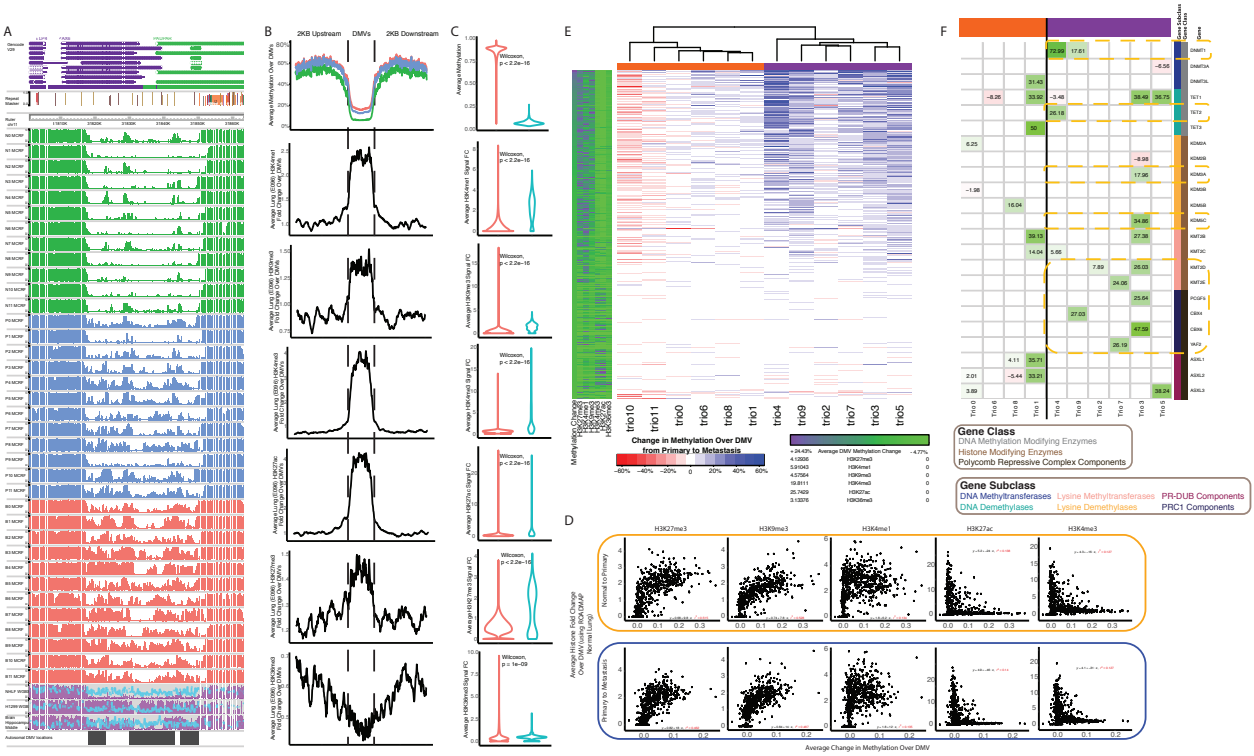


Figure 4.6. Preferential gain of methylation over DNA methylation valleys containing bivalent enhancer marks during disease progression.
A. Browser view[36,37] displaying the methylation status over the DMV with the greatest average gain in methylation from primary to metastasis samples and surrounding regions for all 12 normal samples (green), primary samples (blue), and metastasis samples (red). Also included are three tracks displaying the methylation status for normal human lung fibroblast (NHLF)[20], brain hippocampus middle[20], as well as for the cell line H1299[42]. **B**. Average signal fold change for NHLF chromatin modifications (H3K4me1, H3K9me3, H3K4me3, H3K27ac, H3K27me3, and H3K36me3) over DMVs and surrounding 2 kb regions. **C**. Distribution comparisons of average signal fold change for NHLF histone modifications over DMVs and background 5 kb regions. **D**. Scatterplots displaying the average histone modification fold change in NHLF as a function of the average change in methylation for each DMV (Top: normal to primary; Bottom: primary to metastasis). **E**. Heatmap displaying the change in average methylation over each DMV from primary to metastasis across patients. Patients are hierarchically clustered and DMVs are sorted according to their average primary to metastasis methylation change. Supplementary heatmap to the left displays average DMV methylation change, as well as average NHLF histone modification fold change values per DMV. **F**. Heatmap displaying the average change in variant allele frequency (VAF) for several epigenetic modifiers from primary to metastasis for each patient[17]. Patients are ordered according to E, and yellow boxes indicate genes with increased VAFs only identified in patients with a more sevear gain of methlyation over DMVs.

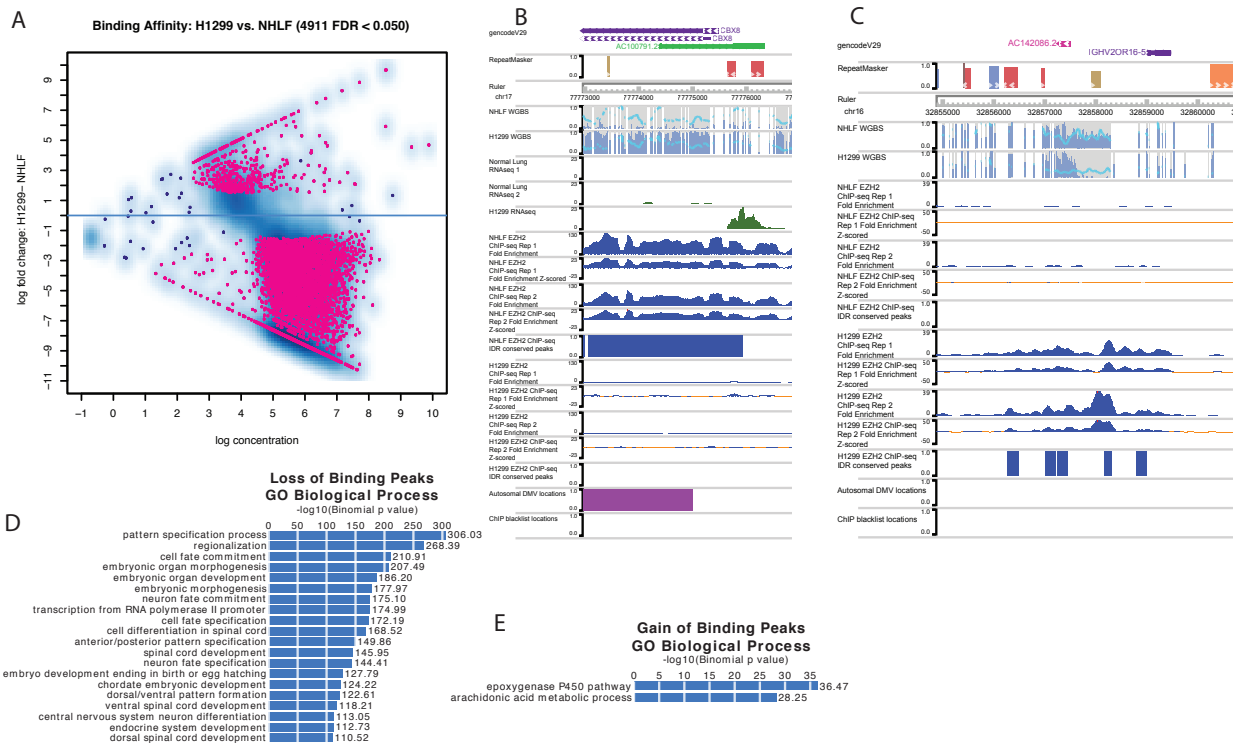


Figure 4.7. DNA methylation valley (DMV) gain of methylation is associated with loss of polycomb repressive complex 2 occupancy in the metastasis surrogate, H1299 (Supplement).

A. MA plot depicting differential EZH2 binding affinity in H1299 and normal lung (NHLF). Significantly differentially bound sites (FDR < 0.05) are highlighted in pink. **B**. Browser view[36,37] depicting the most significant loss of EZH2 binding in H1299 relative to NHLF. Tracks include (from top to bottom) gene, repeat, and chromosomal annotations, whole genome bisulfite sequencing, RNA-seq, EZH2 fold enrichment for each replicate, *z*-scored EZH2 fold enrichment for each replicate, locations of IDR conserved peaks, as well as autosomal DMV locations and blacklisted regions. **C**. Browser view[36,37] depicting the greatest fold change increase of EZH2 binding in H1299 relative to NHLF. Tracks include (from top to bottom) gene, repeat, and chromosomal annotations, whole genome bisulfite sequencing, EZH2 fold enrichment for each replicate, *z*-scored EZH2 fold enrichment for each replicate, locations of IDR conserved peaks, as well as autosomal DMV locations and blacklisted regions. **D**. Top 20 enriched GO Biological Process terms[25] for NHLF EZH2 IDR peaks lost in H1299. **E**. Only 2 enriched GO Biological Process terms[25] for H1299 EZH2 IDR peaks not present in NHLF.

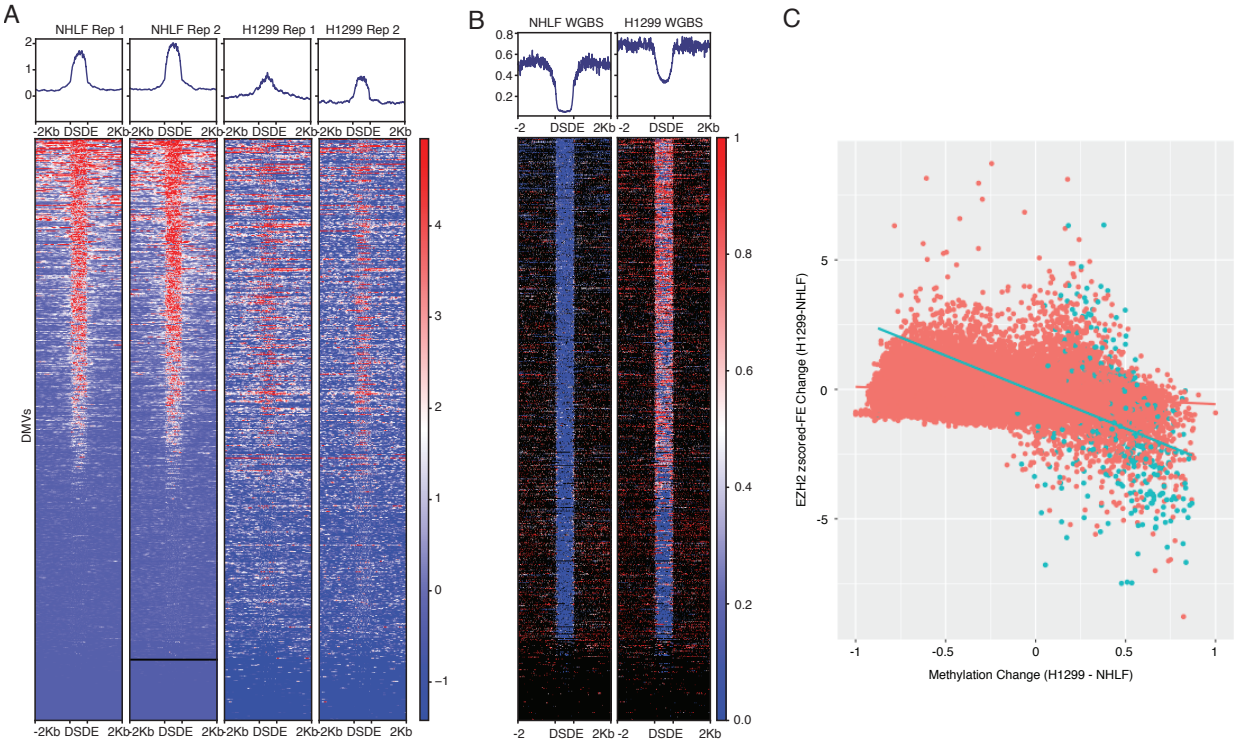


Figure 4.8. DNA methylation valley (DMV) gain of methylation is associated with loss of polycomb repressive complex 2 occupancy in the metastasis surrogate, H1299.
A. Heatmap displaying EZH2 *z*-scored fold enrichment (signal over input) in NHLF (left 2 replicates) and H1299 (right two replicates) for all autosomal DMVs and surrounding 2 kb regions, ordered from greatest average EZH2 enrichment in NHLF to least. DS: DMV start, DE: DMV end. **B**. Heatmap displaying the average DNA methylation over each DMV and surrounding 2 kb regions in NHLF and H1299, ordered according to **A**. **C**. Scatterplot showing the correlation between EZH2 *z*-scored fold enrichment change (from NHLF to H1299) and the average methylation change (form NHLF to H1299) over genomic background 5 kb bins (red) and DMVs (blue).

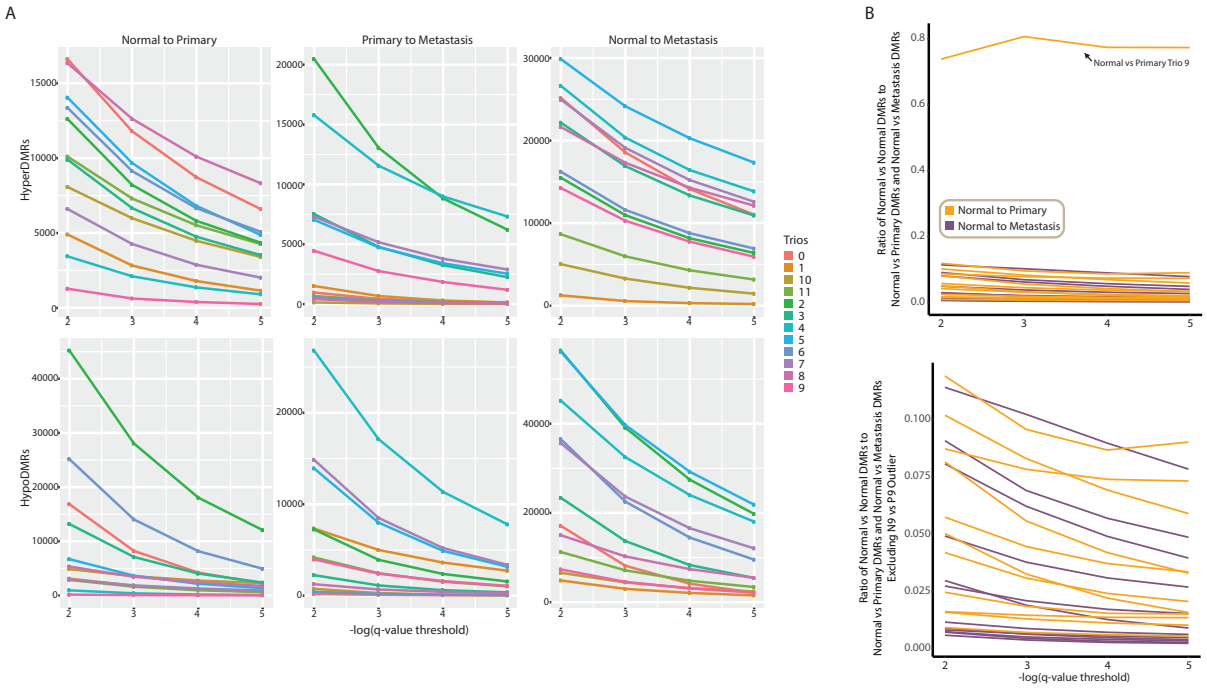


Figure 4.9. Differentially methylated region stringency threshold comparison.
A. Number of differentially methylated regions (DMRs) called for each trio when comparing normal to primary (left), primary to metastasis (middle), and normal to metastasis (right) within a pateint, across various stringency thresholds. Top: hyperDMRs, Bottom: hypoDMRs. **B**. Ratio of normal vs. normal DMRs (across patients) to normal vs. primary DMRs and normal vs. metastasis DMRs (within patients). Top: all trios, Bottom: excluding trio 9 normal vs. primary comparison.

4.7 Tables

Table 4.1 Clinical patient data

Patient ID	ID from Reference #17	Metastasis Type	Gender	Race	Age at Diagnosis	Smoking History	Histology	Chemotherapy	Radiotherapy	Vital Status	Age at Death	Months of Follow-Up	Months of Survival	Tumor Max Cellularity (Sequenza)	Tumor Max Ploidy (Sequenza)	Metastasis Max Cellularity (Sequenza)	Metastasis Max Ploidy (Sequenza)	TMB (Met)	TMB (Tumor)	FGA (Met)	FGA (Tumor)	Jaccard
0	34	ND	Female	White	56	Smoker	Adenocarcinoma	Yes - BCNU wafers; Carboplatin & Pemetrexed-1st line; Erlotinib-2nd line after RLL relapse; Taxol-3rd line	Yes - L scapula mets	Deceased	57	16	16	0.36	2.4	0.43	2.4	13.78	13.76	0.021088	0.021119	0.902271
1	19	Metachronous	Male	White	54	Smoker	Adenocarcinoma	ND	ND	Deceased	56	20	20	0.11	2	0.45	4.8	14.86	2.62	0.048191	0.001045	0.128778
2	4	Synchronous	Male	White	61	Smoker	Adenocarcinoma	Yes - Adjuvant (Gemcitabine/Carboplatin)	Yes - Chest	Alive	NA	127	127	0.52	3.3	0.52	3.3	7.2	12.18	0.049276	0.033421	0.405224
3	20	Metachronous	Female	African American	50	Smoker	Adenocarcinoma	Yes - Adjuvant (Cisplatin/Taxotere), Carboplatin/Taxotere	Yes - WBRT	Deceased	NA	89	89	0.28	2.6	0.72	2.3	14.08	12.12	0.010823	0.009957	0.775339
4	5	Synchronous (LCNEC brain)	Female	White	46	Smoker	Adenocarcinoma	Yes - Adjuvant (Cisplatin/Pemetrexed)	Yes - WBRT	Alive	NA	57	57	0.16	2.8	0.85	3.6	7.36	5.84	0.052445	0.037222	0.585938
5	21	Metachronous	Female	White	57	Smoker	Adenocarcinoma, poorly differentiated	No	Yes - WBRT	Deceased	59	14	14	0.15	5.4	0.6	3.7	9	6.94	0.059845	0.032256	0.621094
6	22	Metachronous	Female	White	55	Smoker	Adenocarcinoma	No	No	Alive	NA	176	176	0.34	5.3	0.28	4.9	7.02	7.32	0.033811	0.056735	0.730676
7	23	Metachronous	Female	Asian	73	Nonsmoker	Adenocarcinoma	ND	ND	Deceased	74	17	17	0.23	3.1	0.78	4.1	2.82	3.12	0.070593	0.008519	0.167241
8	6	Synchronous	Female	White	69	Smoker	Squamous cell carcinoma	Yes - Adjuvant (Cicplatin/Docetaxel), Erlotinib	No	Alive	NA	6	6	0.36	3.4	0.54	1.8	6.06	5.56	0.020177	0.016849	0.629776
9	24	Metachronous	Female	White	ND	ND	Carcinoma, poorly differentiated	ND	ND	Deceased	51	43	43	0.43	2.1	0.4	3.6	5.64	2.6	0.024224	0.008948	0.223859
10	28	ND	Male	ND	51	ND	Adenocarcinoma	ND	ND	Deceased	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
11	39	ND	Female	ND	50	ND	Adenocarcinoma	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND

NA: Not applicable; ND: Not determined/no data

Table 4.2 MeDIP and MRE sequencing statistics

Sample	Number of Non-Redundant, Uniquely Mapped MeDIP Fragments	Number of MRE Filtered Reads	Number of CpG Sites Sampled (MRE)
N0	45655695	20508236	1789400
P0	44520008	20462172	2470336
B0	36534456	19348471	2314652
N1	41160590	55189685	1905327
P1	45966175	24589542	1512191
B1	27680417	27467013	1376658
N2	73416623	48367431	1594200
P2	45759761	27440261	1931307
B2	38601951	31439552	1735960
N3	28508944	25769771	1242424
P3	32292234	13236869	1435900
B3	30494631	23454005	1394917
N4	36628791	37860547	1533749
P4	36874825	26780372	1620655
B4	39390251	33124201	1484732
N5	42036159	22152065	1270311
P5	31945145	20037171	1450734
B5	49004874	21784884	1535680
N6	42272773	26487553	1461826
P6	39707432	36042971	2092491
B6	34312156	23980378	1900609
N7	47756141	32837832	1523850
P7	41880806	11423436	1151597
B7	33637329	25264348	1506970
N8	47955480	34098626	1340320
P8	43943922	27308864	1478275
B8	40680804	27606556	1436488
N9	48839534	23495002	1504045
P9	41818421	34640997	1420906
B9	40003428	27713864	1292284
N10	39964554	54108608	1399785
P10	36783513	19291825	1130174
B10	44585459	43244090	1517505
N11	43487063	40791564	1475551
P11	39159205	21208134	1442202
B11	45807803	30846187	1463536

4.8 References

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
2. Duma, N., Santana-Davila, R. & Molina, J. R. Non–Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. *Mayo Clin. Proc.* **94**, 1623–1640 (2019).
3. Goldstraw, P. *et al.* The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM Classification for lung cancer. *J. Thorac. Oncol.* **11**, 39–51 (2016).
4. Walters, S. *et al.* Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: A population-based study, 2004-2007. *Thorax* **68**, 551–564 (2013).
5. Christofori, G. New signals from the invasive front. *Nature* **44**, 444–450 (2006).
6. Vanharanta, S. & Massagué, J. Origins of Metastatic Traits. *Cancer Cell* **24**, 410–421 (2013).
7. Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging Biological Principles of Metastasis. *Cell* **168**, 670–691 (2017).
8. Cameron, M. D. *et al.* Temporal progression of metastasis in lung: Cell survival, dormancy, and location dependence of metastatic inefficiency. *Cancer Res.* **60**, 2541–2546 (2000).
9. FIDLER, I. Metastasis : Quantitative Analysis of Dis- tribution and Fate of Tumor Emboli Labeled. *J Nat Cancer Inst* **45**, 773–782 (1970).
10. Fidler, I. J. & Nicolson, G. L. Fate of recirculating b16 melanoma metastatic variant cells in parabiotic syngeneic recipients brief communication. *J. Natl. Cancer Inst.* **58**, 1867–1872 (1977).
11. Liotta, L. A., Vembu, D., Saini, R. K. & Boone, C. In Vivo Monitoring of the Death Rate of Artificial Murine Pulmonary Micrometastases. *Cancer Res.* **38**, 1231–1236 (1978).
12. Varani, J., Lovett, E. J., Elgebaly, S., Lundy, J. & Ward, P. A. In Vitro and In Vivo Adherence of Tumor Cell Variants Correlated With Tumor Formation. *Am J Pathol.* **101**, 345–352 (1980).
13. Nguyen, D. X., Bos, P. D. & Massagué, J. Metastasis: From dissemination to organ-specific colonization. *Nat. Rev. Cancer* **9**, 274–284 (2009).
14. Brastianos, P. K. *et al.* Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov.* **5**, 1164–1177 (2015).
15. Chen, G. *et al.* Molecular profiling of patient-matched brain and extracranial melanoma metastases implicates the PI3K pathway as a therapeutic target. *Clin. Cancer Res.* **20**, 5537–5546 (2014).
16. Chatterjee, A., Rodger, E. J. & Eccles, M. R. Epigenetic drivers of tumourigenesis and cancer metastasis. *Semin. Cancer Biol.* **51**, 149–159 (2018).
17. Karlow, J. A., Devarakonda, S., Sankararaman, S. & Kushal, M. Multiomic comparison of paired primary and brain metastatic non-small cell lung cancer suggests tumor cell reprogramming toward a glial cell phenotype. *prep*
18. Brinkman, A. B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22**, 1128–1138 (2012).
19. Murphy, P. J. *et al.* Single-molecule analysis of combinatorial epigenomic states in normal and tumor cells. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7772–7777 (2013).
20. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
21. Gama-Sosa, M. A. *et al.* The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res.* **11**, 6883–6894 (1983).
22. Herman, J. G. *et al.* Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6870–6875 (1998).
23. Roman-Gomez, J. *et al.* Promoter hypermethylation and global hypomethylation are independent epigenetic events in lymphoid leukemogenesis with opposing effects on clinical outcome [5]. *Leukemia* **20**, 1445–1448 (2006).
24. Hu, Y. *et al.* Candidate tumor suppressor ZNF154 suppresses invasion and metastasis in NPC by inhibiting the EMT via Wnt/B-catenin signalling. *Oncotarget* **8**, 85749–85758 (2017).

25. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
26. Takebayashi-Suzuki, K. *et al.* Coordinated regulation of the dorsal-ventral and anterior-posterior patterning of Xenopus embryos by the BTB/POZ zinc finger protein Zbtb14. *Dev. Growth Differ.* **60**, 158–173 (2018).
27. Zhang, X. *et al.* Pax6 is a human neuroectoderm cell fate determinant. *Cell Stem Cell* **7**, 90–100 (2010).
28. Zhang, X. *et al.* Down-regulation of PAX6 by promoter methylation is associated with poor prognosis in non small cell lung cancer. *Int. J. Clin. Exp. Pathol.* **8**, 11452–11457 (2015).
29. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12**, 7–18 (2011).
30. Piunti, A. & Shilatifard, A. The roles of Polycomb repressive complexes in mammalian development and cancer. *Nat. Rev. Mol. Cell Biol.* **22**, 326–345 (2021).
31. Atlasi, Y. & Stunnenberg, H. G. The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* **18**, 643–658 (2017).
32. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
33. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
34. Li, D., Zhang, B., Xing, X. & Wang, T. Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. *Methods* **72**, 29–40 (2015).
35. Stevens, M. *et al.* Estimating absolute methylation levels at single-CpG resolution from methylation enrichment and restriction enzyme sequencing methods. *Genome Res.* **23**, 1541–1553 (2013).
36. Li, D., Hsu, S., Purushotham, D., Sears, R. L. & Wang, T. WashU Epigenome Browser update 2019. *Nucleic Acids Res.* **47**, W158–W165 (2019).
37. Zhou, X. *et al.* The human epigenome browser at Washington University. *Nat. Methods* **8**, 989–990 (2011).
38. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
39. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
40. McLeay, R. C. & Bailey, T. L. Motif Enrichment Analysis : a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**, (2010).
41. Jang, H. S. *et al.* Transposable elements drive widespread expression of oncogenes in human cancers. *Nat. Genet.* **51**, 611-617 (2019).
42. Brocks, D. *et al*. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat. Genet*. **49**, 1052-1060 (2017).

Chapter 5: Capture Methylation-Sensitive Restriction Enzyme Sequencing (Capture MRE-Seq) for Methylation Analysis of Highly Degraded DNA Samples

The contents of this chapter have been adapted from the submitted publication:

"Xing, X.*, Karlow, J.A.*, Li, D., Jang, H.S., Lee, H.J., Wang, T. Capture Methylation-Sensitive Restriction Enzyme Sequencing (Capture MRE-Seq) for methylation analysis of highly degraded DNA samples" *These authors contributed equally to this work.

5.1 Abstract

Understanding the impact of DNA methylation within different disease contexts often requires accurate assessment of these modifications in a genome-wide fashion. Frequently, patient-derived tissue stored in long-term hospital tissue banks have been preserved using formalin-fixation paraffin-embedding (FFPE). While these samples can comprise valuable resources for studying disease, the fixation process ultimately compromises the DNA's integrity and leads to degradation. Degraded DNA can complicate CpG methylome profiling using traditional techniques, particularly when performing methylation-sensitive restriction enzyme sequencing (MRE-seq), yielding high backgrounds and resulting in lowered library complexity. Here, we describe Capture MRE-seq, a new MRE-seq protocol tailored to preserving unmethylated CpG information when using samples with highly degraded DNA. The results

using Capture MRE-seq correlate well (0.92) with traditional MRE-seq calls when profiling non-degraded samples, and can recover unmethylated regions in highly degraded samples when traditional MRE-seq fails, which we validate using bisulfite sequencing-based data (WGBS) as well as methylated DNA immunoprecipitation followed by sequencing (MeDIP-seq).

5.2 Introduction

DNA methylation is a common epigenetic mark in both humans and other vertebrates and is characterized by the addition of a methyl group to the fifth position of a cytosine[1,2]. In humans, this epigenetic modification is most commonly observed within CpG dinucleotides[3–5], but has also been observed in other contexts, particularly within neuronal and developmental cells[6–10]. DNA methylation has been widely studied and found to play an important role in X chromosome inactivation[11,12], cellular differentiation[11,13–15], transcriptional regulation[16–18], and repression of transposable elements[19,20]. In addition to aiding in normal cellular function, aberrant methylation alterations can contribute important functional changes, many of which are likely to assist in tumorigenesis[21–27].

Sequencing-based DNA methylation profiling methods, such as methylated DNA immunoprecipitation plus sequencing (MeDIP-seq)[28–30] and methylation-sensitive restriction enzyme sequencing (MRE-seq)[30] are often jointly applied to assess DNA methylation by enriching for methylated and unmethylated CpGs, respectively[30–32]. Both methods require high quality, high molecular weight, double-stranded genomic DNA (dsDNA), free from fragmentation in order to avoid high background signals. Traditional MRE-seq relies on a collection of restriction enzymes that recognize CpG-containing sequence motifs. Each enzyme cuts DNA at its recognized motif when the CpG within the motif is unmethylated. When using

high quality DNA samples, massively paralleled sequencing reveals unmethylated CpG sites enriched at fragment ends. However, if DNA samples are degraded, roughly 80%-90% of fragment ends of resulting MRE-seq libraries do not represent truly unmethylated CpGs but are instead random breakpoints from shearing (Figure 5.1 and Figure 5.2).

Due to the relative ease of long-term maintenance, many clinically relevant tissue samples collected over the years have been preserved through a process known as formalin fixation paraffin embedding (FFPE). During the fixation process, exposure to formaldehyde, the main component of formalin, causes DNA fragmentation and results in DNA-protein crosslinks[33,34]. In addition, many factors such as duration of storage, pH, and temperature have also been shown to correlate with DNA quality[33,35]. Resulting DNA damage significantly complicates downstream analyses; however, working with samples preserved in this way is often unavoidable, necessitating modifications to standard techniques in order to analyze the data.

Here we present Capture MRE-seq, a modified version of MRE-seq, specifically tailored to preserving unmethylated CpG information in highly degraded DNA samples. Traditionally, DNA is heated at high temperatures in many commercial DNA extraction kits to remove formalin-induced crosslinks, resulting in increased fragmentation and DNA denaturation. Unlike this traditional step for processing FFPE samples, our protocol skips heating at high temperatures since denatured, single-stranded DNA (ssDNA) cannot be used in the restriction enzyme digestion. Compared to the traditional MRE-seq protocol, the main novelty of Capture MRE-seq is in the adapter ligation step. Instead of using traditional adapters, we incubate our enzyme-digested DNA fragments with adapters containing a CG-overhang, causing these adapters to capture and enrich for fragments of DNA that have complementary CG-overhangs produced by

restriction enzyme cutting, ultimately increasing the effective complexity of our MRE-seq library (Figure 5.1).

5.3 Results

When used on non-degraded DNA samples, our new Capture MRE-seq method produces highly correlated results with those produced using traditional MRE-seq. In addition, Capture MRE-seq can produce results for degraded DNA samples when traditional MRE-seq protocols fail. To demonstrate the effectiveness of our new method for working with degraded samples, we first showed that DNA extracted from a human lung tissue FFPE block was highly fragmented and produced a wide smear on an agarose gel (Figure 5.3A) whereas non-degraded DNA produced a single, sharp band localized towards the top of the gel, indicative of intact and high molecular weight DNA (Figure 5.3B). We also demonstrated that we could effectively recapitulate an FFPE genomic DNA fragment distribution by sonicating a non-degraded GM12878 sample (Figure 5.3C). We found that traditional MRE-seq and our new Capture MRE-seq method produced libraries with similar fragment size distributions when using intact DNA, with an added small band (120bp) representing adapter dimers in the Capture MRE-seq method (Figure 5.4A, lanes 1&2). Additionally, we found that Capture MRE-seq produced consistent library fragment size distributions across different degrees of degradation (Figure 5.4B, lanes 1&2), which were also similar to the resulting distributions produced for both MRE protocols on non-degraded DNA (Figure 5.4A, lanes 1&2). We observed that the library fragment size distribution for the FFPE sample using Capture MRE-seq was comparable to those produced using either degraded or intact DNA from either method (Figure 5.4A, lanes 1&2; Fig 4B, lanes 1&2; Figure 5.4C, lane 1). As expected, the fragment size distributions for all libraries shifted

upward on the gel after removal of adapter dimers (Figure 5.4A, lane 3; Figure 5.4B, lanes 3&4; Figure 5.4C, lane 2), due to the loss of small DNA fragments.

Resulting methylation scores using Capture MRE-seq correlated well (0.92) with results from traditional MRE-seq for non-degraded samples, and Capture MRE-seq results using non-degraded and degraded (1 min and 10 min sonicated) DNA also exhibited high degrees of correlation (0.89 and 0.86, respectively) (Figure 5.5A). We found that a similar proportion of reads with ends mapping to MRE-filtered cut sites were retained when using Capture MRE-seq on degraded samples as compared to using either the traditional MRE-seq protocol or Capture MRE-seq on intact DNA (Figure 5.5B). Likewise, the number of unique CpG sites sampled was proportional to the number of MRE-filtered reads when using either the traditional or Capture MRE-seq method on intact DNA, and when using Capture MRE-seq on 1 min sonicated DNA. The proportion of uniquely sampled CpG sites relative to the number of MRE-filtered reads was slightly lower for the 10 min sonicated DNA processed using the Capture MRE-seq protocol (Figure 5.5C). External validation using WGBS data[36] and MeDIP-seq data corroborated our Capture MRE-seq calls on degraded samples, as demonstrated by a browser view highlighting the clinically relevant gene TP53 (Figure 5.6A). MeDIP-seq data for the same FFPE sample also processed using the Capture MRE-seq protocol complemented as well, highlighting the use of this protocol for FFPE samples (Figure 5.6B). Our findings demonstrate both the feasibility and the utility of our new Capture MRE-seq protocol for use particularly with highly degraded samples, such as those produced through formalin fixation.

5.4 Materials

5.4.1 Genomic DNA extraction

1. GM12878 cell line
2. Human lung FFPE tissue block
3. Quick-DNA Plus Kit (Zymo Research)
4. Quick-DNA FFPE Kit (Zymo Research)
5. Ethanol
6. 2-Mercaptoethanol
7. Molecular Biology Grade Water
8. 1M Tris-HCl, pH 8.0
9. 10mM Tris-HCl, pH 8.0: Dilute 1M Tris, pH 8.0 in molecular biology grade water and store at room temperature.
10. NanoVue spectrophotometer
11. Qubit fluorometer for quantification
12. Qubit dsDNA HS Assay Kit
13. Qubit assay tubes
14. Agarose, Molecular Biology Grade
15. Ethidium bromide
16. Tris-acetate-EDTA (TAE) buffer, 50x solution
17. Orange DNA loading dye, 6x
18. 1kb plus DNA ladder
19. Gel electrophoresis system
20. Get Doc Imaging System
21. Bioruptor Pico system for DNA sonication

5.4.2 DNA fragmentation

1. Restriction enzymes: HpaII, HinP1I, AciI, HpyCH4IV
2. Thermal cycler for incubation

5.4.3 Size selection of fragmented DNA

1. Agencourt AMPure XP beads

5.4.4 Library preparation

1. T4 ligase reaction buffer for adapter annealing, 10x (NEB; cat. no. 10B0202S)
2. Capture MRE-seq adapter oligos (5'-ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT-3' and 5'-/5Phos/CGA GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC-3') (IDT) (*see* 5.6.1): Spin down lyophilized adapter oligos and resuspend each oligo in 1× T4 ligase reaction to 20 μM. Add equimolar quantities of each adapter into a 1.5-ml microcentrifuge tube. Divide into small aliquots (50 μl) and incubate at 95 °C for 10 min, and then leave to cool down to room temperature. Store aliquots at − 20 °C until required. The final concentration should be 10 μM.
3. Quick Ligation Kit (NEB)
4. 2x NEBNext Master Mix (NEB)
5. Indexed PCR primers PE1 and PE2 (5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T-3' and 5'-CAA GCA GAA GAC GGC ATA CGA GAT NNNNNNN GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG A-3') (IDT) (*see* 5.6.2): Spin down primer oligos and resuspend each oligo in molecular biology grade water to create 100 μM stocks. Dilute an aliquot tenfold in molecular biology grade water to create 10 μM dilutions. Store dilutions at -20°C until needed.

5.4.5 Library QC

1. TapeStation System (*see* 5.6.3)
2. High Sensitivity D1000 ScreenTape
3. High Sensitivity D1000 ScreenTape assay sample buffer
4. Qubit fluorometer for quantification (5.4.1.11)
5. Illumina NextSeq 500 Sequencer

5.4.6 MRE-seq data processing

1. Standard computer
2. FastQC software[37] (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), or equivalent
3. Cutadapt software[38] (https://cutadapt.readthedocs.io/en/stable/installation.html) or equivalent
4. Burrows-Wheeler Aligner (BWA) software[39] (http://bio-bwa.sourceforge.net/) or equivalent
5. Samtools software[40] (http://www.htslib.org/) or equivalent
6. MethylQA software[31] (http://methylqa.sourceforge.net/)
7. WashU Epigenome Browser[41,42] (https://epigenomegateway.wustl.edu/)

5.5 Methods

5.5.1 Genomic DNA extraction

1. Extract genomic DNA from the GM12878 cell line following the manufacturer's instructions from the Quick-DNA plus kit.
2. Sonicate the genomic DNA of GM12878 for 1 minute and 10 minutes with 30 seconds on/off using the Bioruptor Pico system to generate two different degradation levels of DNA. Check the resulting library size distributions using the High Sensitivity D1000 ScreenTape and the sample buffer on the TapeStation System (Figure 5.3C).
3. Extract genomic DNA from the FFPE tissue following the manufacturer's instructions from the Quick-DNA FFPE Kit. Skip the incubation step at 94 °C for 20 minutes after proteinase K digestion (*see* 5.6.4). Skip the addition of isopropanol after adding the genomic lysis buffer to the sample (*see* 5.6.5). Save the first flow-through from the column (*see* 5.6.6).
4. Purified genomic DNA can be stored at -20 °C long-term.
5. Quantitate DNA by NanoVue Spectrophotometer and Qubit fluorometer (*see* 5.6.7).
6. Validate DNA quality by running it on 1% TAE agarose gel (*see* 5.6.8) (Figure 5.3A and Figure 5.3B).

5.5.2 DNA fragmentation

1. Transfer 50-100 ng of genomic DNA to a PCR tube for each reaction mix. A pre-mix is set up in a 20 µl volume with 1x restriction enzyme buffer and 1 U of restriction enzyme (HpaII, HinP1I, AciI, and HpyCH4IV). Keep the mixture on ice. Incubate the reaction in a thermocycler at 37 °C for 3 hours. Four digests for each sample can be set up in parallel (*see* 5.6.9).
2. Add an additional 1 U of enzymes to each reaction after 3 hours of incubation. Mix and incubate for another 3 hours at 37 °C (*see* 5.6.10).

3. After a total digestion time of 6 hours, incubate the reaction at 65 ºC for 20 minutes to deactivate all enzymes except for HpaII, which is deactivated at 80 ºC.
4. Combine the four reactions. Each sample should have a total volume of 80 µl and can be stored at -20 ºC until needed.

5.5.3 Size selection of fragmented DNA

1. Perform a gel-free selection using Agencourt AMPure XP beads. Use 0.5x and 1.5x ratios of AMPure XP beads to dual size-select DNA with fragment sizes between 100-2000 bp.
2. Mix thoroughly to resuspend beads before use.
3. Add 40 µl (0.5x) of the resuspended beads to 80 µl of the fragmented sample (5.5.2.4). Mix thoroughly by vortexing and incubate at room temperature for 5 minutes.
4. Place the sample tube on a magnetic separator until the solution is clear. Transfer the clear supernatant, which contains the size-selected sample (120 µl), to a new tube. The tube with the remaining beads can be discarded.
5. Add 120 µl (1.5x) of the resuspended beads to the 120 µl of the supernatant. Mix thoroughly by vortexing and incubate at room temperature for 5 minutes.
6. Place the sample tube on a magnetic separator until the solution is clear. Remove and discard the supernatant. The size-selected sample is now bound to the beads.
7. Wash the beads twice with 200 µl of 85% ethanol. Incubate at room temperature for 30 seconds. Remove and discard the ethanol.
8. Quickly spin the beads and remove any remaining liquid. Place the tube on a heat block at 37 ºC for about 2 minutes with the cap open. Cracks will be observed in the bead pellets when drying is complete.
9. Add 23 µl of 10 mM Tris (pH 8.0) to the dried beads and pipet up and down 10 times to mix thoroughly. Place the sample tube on a magnetic separator until the solution is clear. Transfer the liquid to a fresh tube. The liquid contains the desired size-selected sample and can be stored at -20 ºC until needed.

5.5.4 Library preparation

1. Adapter ligation is set up in a 50 µl volume with 1x Quick ligation reaction buffer, digested, size-selected DNA (5.5.3.9), 1 µl of 0.01 µM pre-annealed adapter oligo mix (*see* 5.6.11), and 1 µl of T4 DNA ligase. Incubate the mixture at room temperature for 15 minutes.
2. Perform a gel-free purification using AMPure XP beads. A 1.5x ratio of AMPure XP beads can be used to remove free adapters from the adapter-ligated DNA (*see* 5.6.12).
3. Add 75 µl (1.5x) of the resuspended beads to the adapter ligated sample (step 1 above). Mix thoroughly by vortexing and incubate at room temperature for 5 minutes.
4. Repeat steps 6-8 in section 5.5.3 to wash the beads.
5. Add 20 µl of 10 µM Tris (pH 8.0) to the dried beads and pipet up and down 10 times to mix thoroughly. Place the sample tube on a magnetic separator until the solution is clear. Transfer the liquid to a fresh tube. The liquid contains the purified sample and can be stored at -20 ºC until needed.
6. An adapter-modified PCR should be carried out on ice in a 50 µl volume with 25 µl of 2x NEBNext Master Mix, 20 µl of purified, adapter-ligated DNA (step 5 above), and 2.5 µl of each 10 µM PCR primer PE1 and PE2.
7. Perform PCR in a preheated thermocycler with denaturation conditions set for 30 seconds at 98 ºC, followed by 15 cycles of 10 seconds at 98 ºC, 30 seconds at 64 ºC, and 30 seconds at 72 ºC, with a final extension of 5 minutes at 72 ºC.

5.5.5 Library size selection

1. Use 0.55x and 1x ratios of AMPure XP beads to select DNA within the range of 200-600 bp (*see* 5.6.13) (Figure 5.4A, lane 2; Figure 5.4B, lanes 1&2; Figure 5.4C, lane 1).
2. Add 27.5 µl (0.55x) of the resuspended beads to 50 µl of the PCR sample. Mix thoroughly by vortexing and incubate at room temperature for 5 minutes.
3. Place the sample tube on a magnetic separator until the solution is clear. Transfer the clear supernatant, which contains the size-selected sample (77.5 µl), to a new tube. The tube with the remaining beads can be discarded.
4. Add 50 µl (1x) of the resuspended beads to 77.5 µl of the supernatant. Mix thoroughly by vortexing and incubate at room temperature for 5 minutes.
5. Repeat steps 6-8 in section 5.5.3.
6. Add 15 µl of 10 µM Tris (pH 8.0) to the dried beads and pipet up and down 10 times to mix thoroughly. Place the sample tube on a magnetic separator until the solution is clear. Transfer the liquid to a fresh tube. The liquid contains the purified sample and can be stored at -20 °C until needed.
7. Run the library on the TapeStation (*see* 5.6.14).
8. Quantitate each library using the Qubit fluorometer.
9. If considering multiple libraries, pool the indexed MRE-seq libraries by equal molar concentrations and fill to 50 µl with 10 µM Tris (pH 8.0).
10. Select the library or library pool using a 0.8x ratio of AMPure XP beads to remove adapter dimers. Add 40 µl (0.8x) of the resuspended beads to 50 µl of the library sample (step 9 above). Mix thoroughly by vortexing and incubate at room temperature for 5 minutes.
11. Repeat steps 6-8 in section 5.5.3.
12. Add 15 µl of 10 µM Tris (pH 8.0) to the dried beads and pipet up and down 10 times to mix thoroughly. Place the sample tube on a magnetic separator until the solution is clear. Transfer the liquid to a fresh tube. The liquid contains the purified sample and can be stored at -20 °C until needed.
13. Continue the selection using a 0.8x ratio of AMPure XP beads until no adapter dimers can be found on the TapeStation (Figure 5.4A, lane 3; Figure 5.4B, lanes 3&4; Figure 5.4C, lane 2) (*see* 5.6.15).

5.5.6 Library QC

1. Quantitate the final library using the Qubit fluorometer.
2. Run the library on the TapeStation (*see* 5.6.3).

5.5.7 Next-generation sequencing

1. Run 75-bp single-end sequencing on a NextSeq 500 sequencer following Illumina's standard protocol but calling from the 4th base (*see* 5.6.16). On an Illumina NextSeq 500 platform, typically ~400 million raw reads can be obtained.

5.5.8 MRE-seq data processing

1. Determine the sequencing library quality using FastQC[37] (optional).
2. Trim adapter sequences and low-quality ends from the sequencing reads using Cutadapt[38] (or equivalent).
3. Align the trimmed MRE-seq reads using BWA-MEM[39] (or equivalent).
4. Sort and merge the BAM files from different sequencing lanes for the same sample using SAMtools[40] (or equivalent).

5. Obtain MRE-seq statistics using MethylQA[31] (optional) (Figure 5.5).
6. Upload MRE-seq results to the Washington University Epigenome Browser[41,42] for visualization (or equivalent; optional) (Figure 5.6).

5.6 Notes

1. The 5-terminal phosphate (5Phos) modification is added in the adapter oligo to facilitate the ligation of the Capture MRE-seq adapters to the DNA digests.
2. "N" indicates the index code. Multiple libraries can be indexed with different codes so that they can be pooled together for sequencing.
3. TapeStation or an equivalent bioanalyzer may be used.
4. Incubation at 94 °C for 20 minutes may denature the DNA to ssDNA, which cannot be used in restriction enzyme digestions. We therefore skip this high temperature incubation to avoid generating ssDNA.
5. ssDNA will also be purified if present in the sample upon the addition of isopropanol. To isolate more dsDNA, we skip the addition of isopropanol to obtain DNA > 500 bp.
6. In some rare cases, FFPE DNA may be extremely degraded and DNA > 500 bp may not be present in the sample. To get DNA back in these cases, isopropanol has to be used in the saved flow-through from the column.
7. The NanoVue Spectrophotometer quantifies both ssDNA and dsDNA. The Qubit fluorometer quantifies only dsDNA when using the Qubit dsDNA HS Assay Kit and Qubit assay tubes. Qubit readings of dsDNA should be used in the later restriction enzyme digestion step.
8. The agarose gel electrophoresis is optional if the DNA quantity is low. Instead, the genomic DNA reagents and the ScreenTape from Agilent can be used to check the DNA profile.
9. HpaII, HinP1I, AciI, and HpyCH4IV all cut unmethylated CpGs and produce 5' end CG-overhang fragments which can be ligated to the 5' end of Capture MRE-seq adapters to identify unmethylated regions at a single CpG resolution.
10. Extending the incubation time can help complete the enzyme digestion, but we do not suggest incubating overnight since doing so may induce enzyme star activity.
11. To get 0.01 μM adapters, a serial dilution of 1:100 and 1:10 needs to be made from the 10 μM adapter concentration using cold molecular biology grade water (always keeping the adapters on ice).
12. Free adapters have to be removed to avoid the formation of additional dimers during PCR.
13. All Capture MRE-seq adapters have CG-overhang ends which can easily form dimers in PCR. These dimers are difficult to remove and may result in an extra band (Figure 5.4A, lane 2; Figure 5.4B, lanes 1&2; Figure 5.4C, lane 1).
14. A 120-bp adapter dimer band can often still be seen after library size selection.
15. Since the sequencer will preferentially sequence smaller fragments, the 120 bp dimers need to be completely removed prior to sequencing in order to avoid large quantities of dimer reads. If 120 bp adapter dimers can still be seen on the TapeStation, the library pool needs to be re-selected again.
16. DNA fragments digested by restriction enzymes have CG-overhang ends which form chimeric fragments that cannot be properly paired when processing paired-end reads. We suggest running single-end sequencing or processing paired-end reads as single-end

reads. If running single-end sequencing, set up read calling beginning with the 4th base. If running paired-end sequencing which begins with the 1st base, mix MRE libraries (no more than 30%) with additional libraries with complex ends. This is necessary for an MRE library since all MRE fragments start with the same sequence after restriction enzyme digestion. Beginning with 3 dark cycles or mixing at a low percentage with other libraries allows the sequencer to efficiently map the clusters and make the necessary calculations needed to interpret the images and generate accurate sequencing data[43].

5.7 Data Availability

An epigenomic data hub has been created to display all data used in this study, and is available at https://epigenomegateway.wustl.edu/browser/ with the accompanying session bundle ID: 5d37dcb0-ed9a-11ea-a6fa-9fc16c0f334f.
Sequencing data for GM12878 traditional MRE-seq, Capture MRE-seq, 1 minute sonicated Capture MRE-seq, 10 minutes sonicated MRE-seq, MeDIP-seq, Lung FFPE Capture MRE-seq and MeDIP-seq have been deposited in Gene Expression Omnibus under the BioProject ID "PRJNA656241".

5.8 Acknowledgements

We would like to thank Jess Hoisington-Lopez and Maria LynnJaeger from the Edison Family Center for Genome Sciences and Systems Biology at Washington University for assisting in sequencing our libraries. We would also like to thank Dr. Mark Watson and Dr. Ramaswamy Govindan for generously providing the FFPE block used in this study.

5.9 Figures

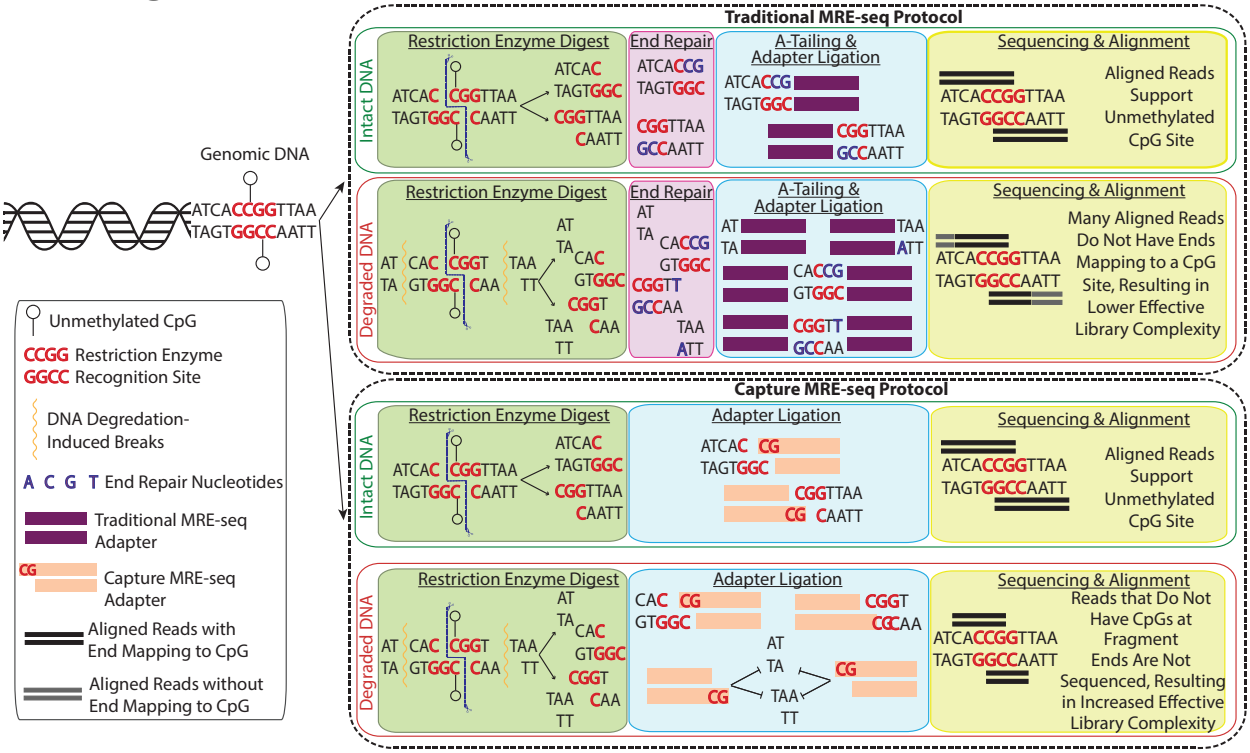


Figure 5.1: Schematic overview of traditional MRE-seq and the new Capture MRE-seq protocol run on both intact DNA and degraded DNA.
Top panel contrasts the traditional MRE-seq protocol run with intact (top) vs. degraded (bottom) DNA. Bottom panel contrasts the new Capture MRE-seq protocol run with intact (top) vs degraded (bottom) DNA.

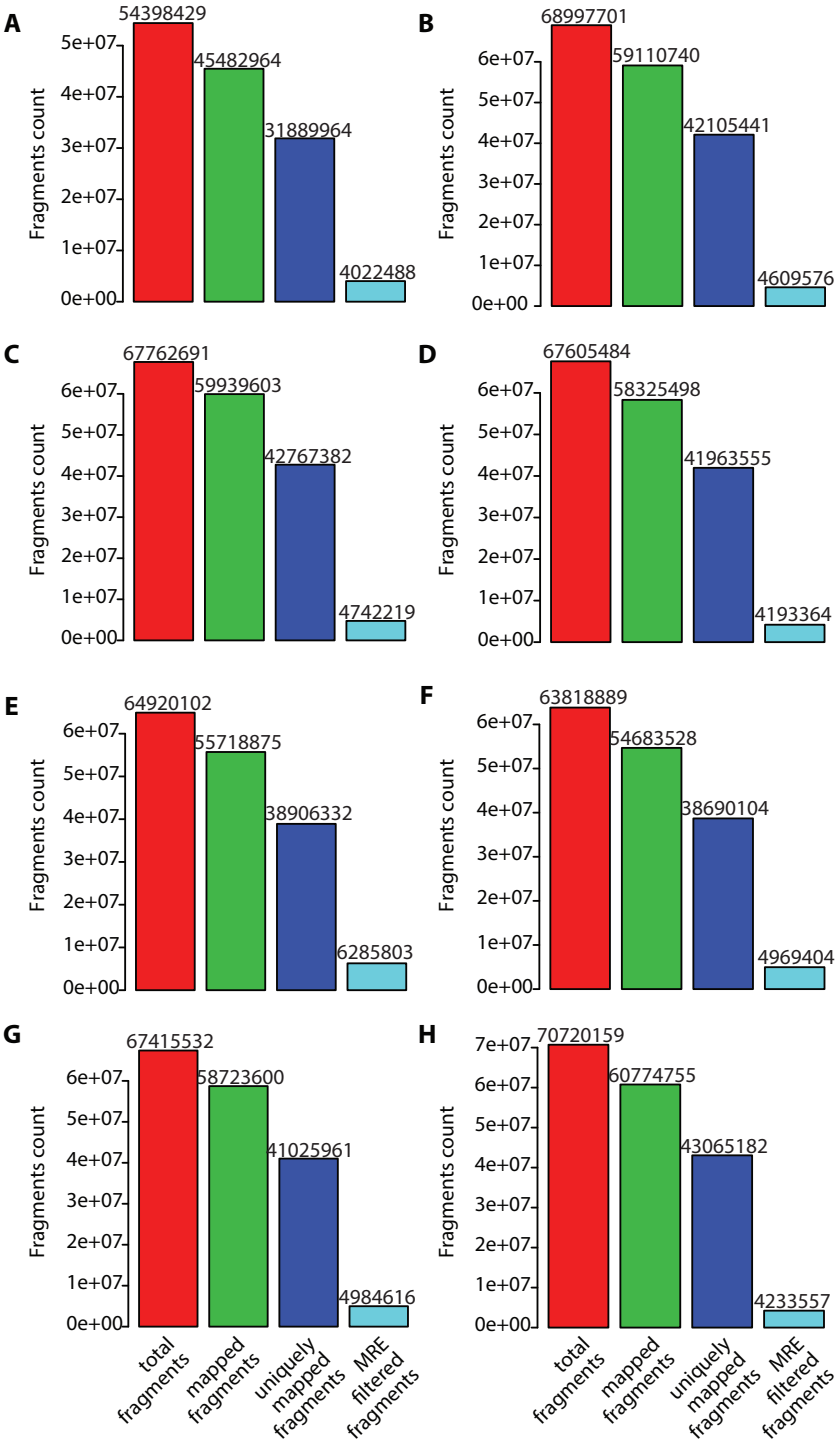


Figure 5.2: Sequencing fragment counts (as output by MethylQA [31]) throughout traditional MRE-seq processing.

(A-H) indicate various degraded DNA samples of neuronal cells from Swiss Webster mouse cortex.

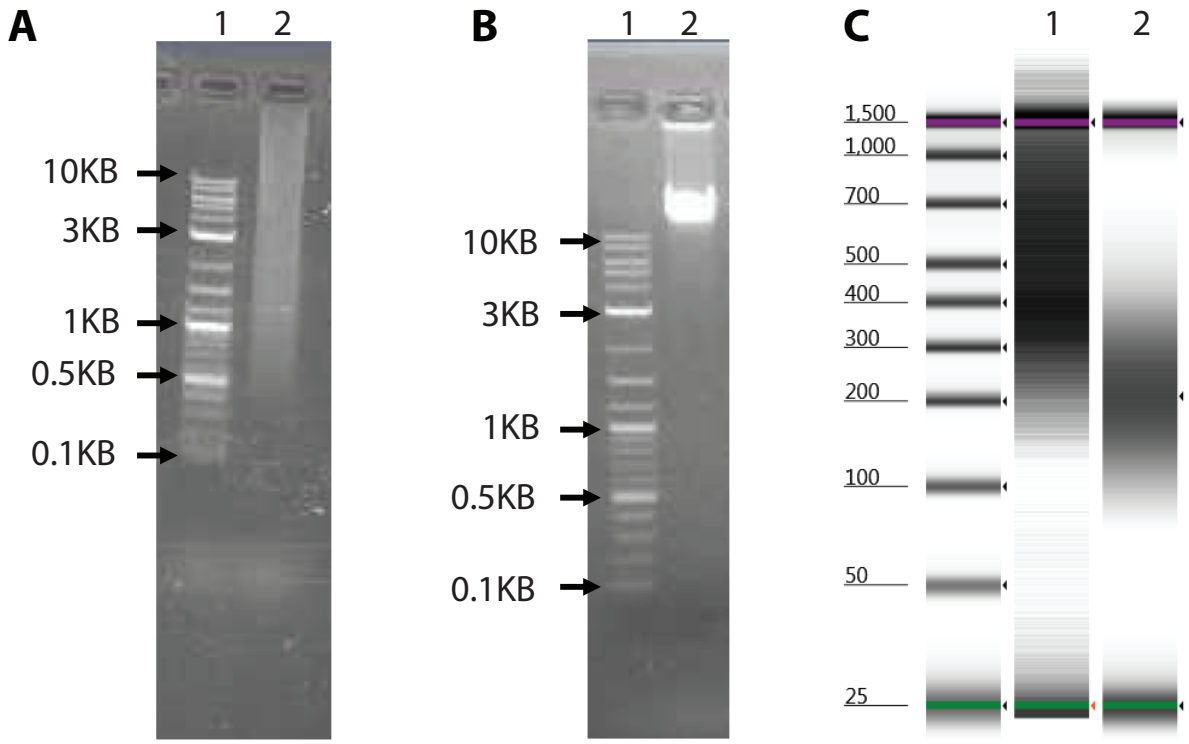


Figure 5.3: Gel images of genomic DNA (gDNA) with varying degrees of degradation. (A) gDNA extracted from an FFPE slice from human lung using the described protocol on a 1% agarose gel. (B) non-degraded gDNA from the GM12878 cell line on a 1% agarose gel. (C) gDNA of GM12878 sonicated for 1 minute (lane 1) and 10 minutes (lane 2) on the HS TapeStation.

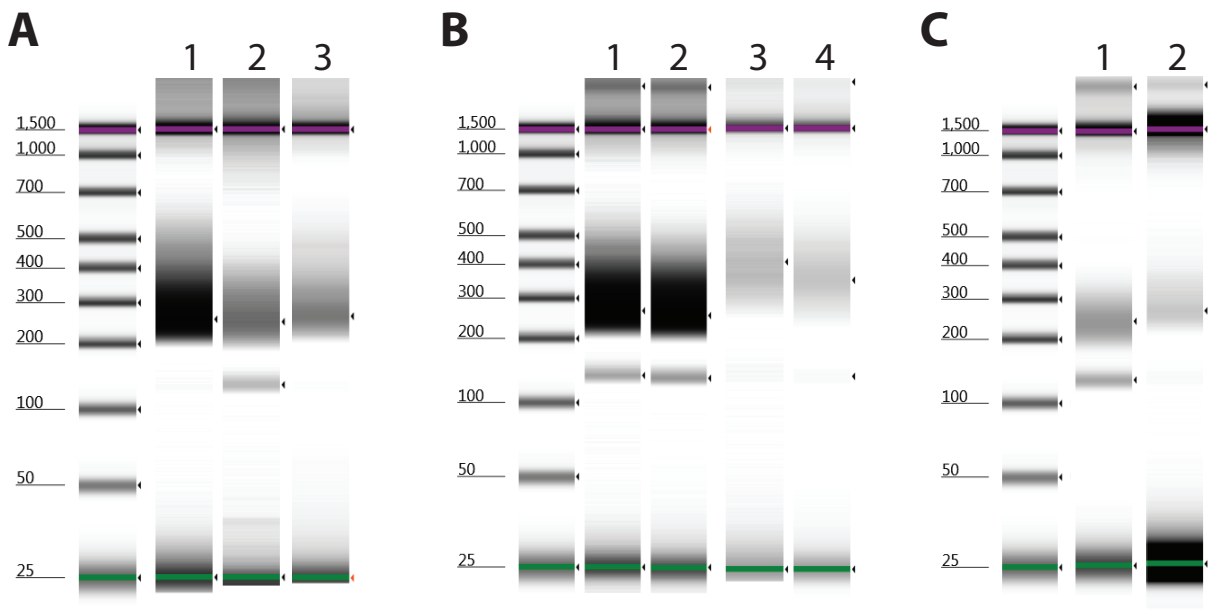


Figure 5.4: Gel images of MRE-seq libraries using the HS TapeStation.
(A) MRE-seq library of non-degraded GM12878 using traditional MRE-seq (lane 1), Capture MRE-seq (lane 2), and Capture MRE-seq post adapter removal (lane 3). (B) MRE-seq libraries of 1 minute (lanes 1&3) and 10 minutes (lanes 2&4) sonicated GM12878 using the Capture MRE-seq method, pre- and post-adapter removal, respectively. (C) Capture MRE-seq library of the FFPE sample pre- and post-adapter dimer removal, respectively (lanes 1& 2).

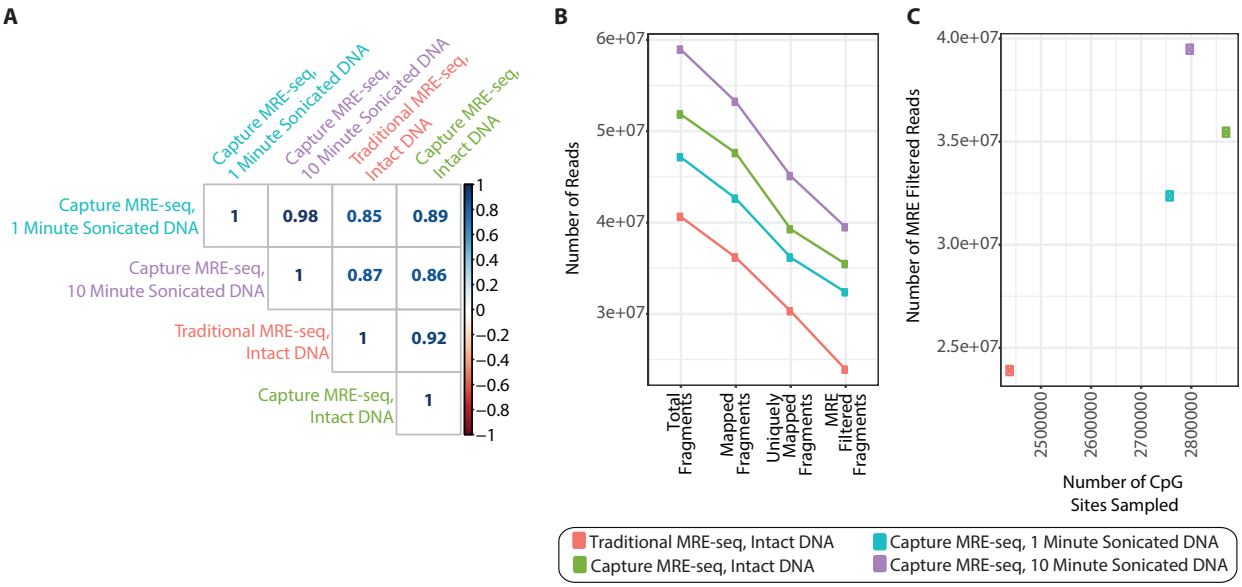


Figure 5.5: Comparison of results from MRE-seq methods for non-degraded and degraded DNA samples.

(A) Correlation analysis of all possible MRE-seq cut site MRE scores from different libraries (theoretical cut sites whose closest neighbour was greater than 500 bp away were excluded). (B) Comparison of the number of reads retained after various filtering conditions across MRE-seq libraries. (C) Comparison of the number of uniquely sampled CpGs with respect to the number of MRE-seq filtered reads for each MRE-seq library.

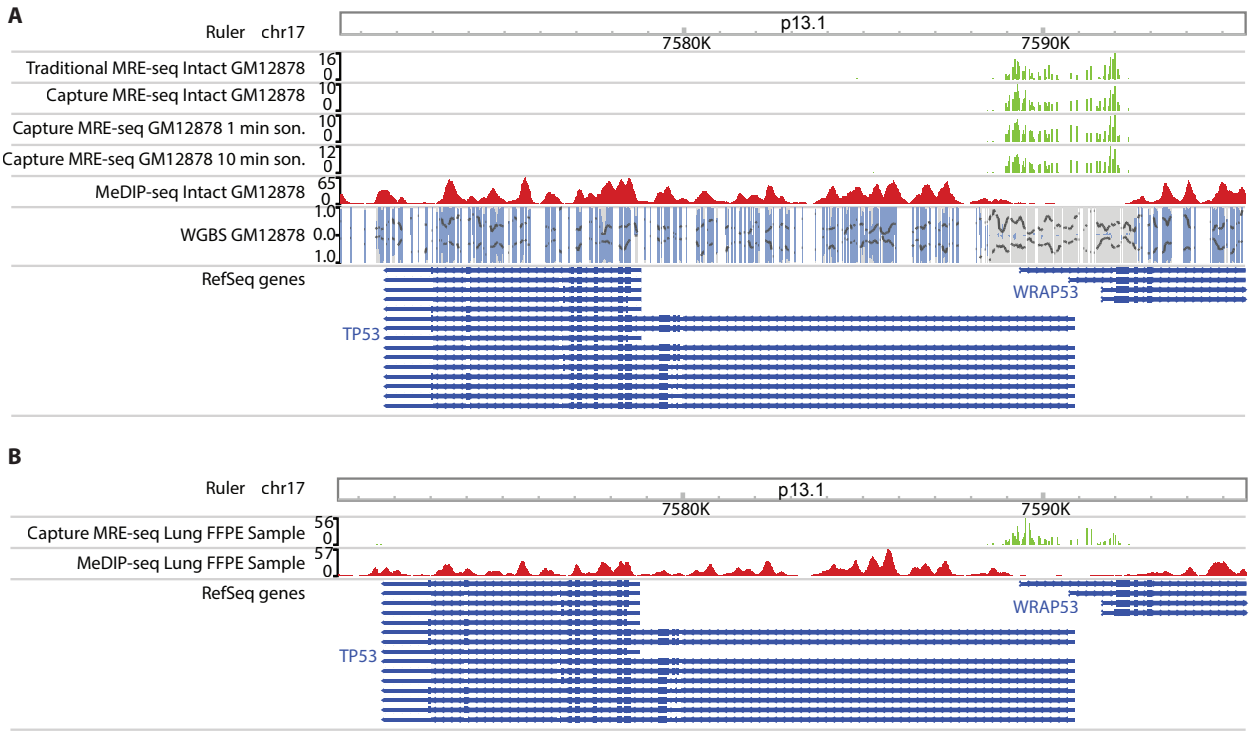


Figure 5.6: Capture MRE-seq validation over *TP53*.
(A) Browser view of *TP53* showing a high correspondence of MRE-seq calls across different MRE-seq libraries. WGBS data for GM12878[36] and MeDIP-seq data support the MRE-seq data, showing a high degree of methylation over the gene body but not over the promoter. Regarding the WGBS track: gray bars represent CpG locations, the height of blue bars indicates the percentage of reads supporting methylated CpG sites, and the black marks indicate the sequencing depth at each position. (B) Browser view of *TP53* showing Capture MRE-seq calls on the FFPE sample and complementary MeDIP-seq calls for the same sample.

5.10 References

1. SINSHEIMER, R. L. The action of pancreatic deoxyribonuclease. II. Isomeric dinucleotides. *J. Biol. Chem.* **215**, 579–583 (1955).
2. Smith, Z. D. & Meissner, A. DNA methylation: Roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
3. Ehrlich, M. & Wang, R. Y. H. 5-Methylcytosine in Eukaryotic DNA. *Science (80-.).* **212**, 1350–1357 (1981).
4. Gruenbaum, Y., Stein, R., Cedar, H. & Razin, A. Methylation of CpG sequences in eukaryotic DNA. *FEBS Lett.* **124**, 67–71 (1981).
5. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).
6. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
7. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science (80-.).* **341**, (2013).
8. Grafstrom, R. H., Yuan, R. & Hamilton, D. L. Charcteristics of DNA Meth from mouse fibro. *Nucleic Acids Res.* **13**, 2827–2842 (1985).
9. Ramsahoye, B. H. *et al.* Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5237–5242 (2000).
10. Ziller, M. J. *et al.* Genomic distribution and Inter-Sample variation of Non-CpG methylation across human cell types. *PLoS Genet.* **7**, (2011).
11. RIggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet.* **14**, 9–25 (1975).
12. Patrat, C., Ouimette, J. F. & Rougeulle, C. X chromosome inactivation in human development. *Development* **147**, (2020).
13. Jones, P. A. & Taylor, S. M. Cellular differentiation, cytidine analogs and DNA methylation. *Cell* **20**, 85–93 (1980).
14. Parry, A., Rulands, S. & Reik, W. Active turnover of DNA methylation during cell fate decisions. *Nat. Rev. Genet.* **22**, 59–66 (2021).
15. Atlasi, Y. & Stunnenberg, H. G. The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* **18**, 643–658 (2017).
16. Holliday, R. & Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science (80-.).* **187**, 226–232 (1975).
17. Busslinger, M., Hurst, J. & Flavell, R. DNA methylation and the regulation of globin gene expression. *Cell* **34**, 197–206 (1983).
18. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science (80-.).* **356**, (2017).
19. Schwartz, D. & Dennis, E. Transposase activity of the Ac controlling element in maize is regulated by its degree of methylation. *MGG Mol. Gen. Genet.* **205**, 476–482 (1986).
20. Deniz, Ö., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* **20**, 417–431 (2019).
21. Feinberg, A. P. & Vogelstein, B. Hypomethylation distinguishes genes of some human

cancers from their normal counterparts. *Nature* **301**, 89–92 (1983).
22. Gama-Sosa, M. A. *et al.* The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res.* **11**, 6883–6894 (1983).
23. Greger, V., Passarge, E., W, H., E, M. & Horsthemke, B. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum Genet.* **83**, 155–158 (1989).
24. Sakai, T. *et al.* Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *Am. J. Hum. Genet.* **48**, 880–888 (1991).
25. Herman, J. G. *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 9700–9704 (1994).
26. Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome-biological and translational implications. *Nat. Rev. Cancer* **11**, 726–734 (2011).
27. Klutstein, M., Nejman, D., Greenfield, R. & Cedar, H. DNA methylation in cancer and aging. *Cancer Res.* **76**, 3446–3450 (2016).
28. Weber, M. *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* **37**, 853–862 (2005).
29. Taiwo, O. *et al.* Methylome analysis using MeDIP-seq with low DNA concentrations. *Nat. Protoc.* **7**, 617–636 (2012).
30. Maunakea, A. K. *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253–257 (2010).
31. Li, D., Zhang, B., Xing, X. & Wang, T. Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. *Methods* **72**, 29–40 (2015).
32. Xing, X., Zhang, B., Li, D. & Wang, T. Comprehensive Whole DNA Methylome Analysis by Integrating MeDIP-seq and MRE-seq. in *DNA MEthylation Protocols* (ed. Tost, J.) 209–246 (Humana Press, New York, NY, 2018). doi:https://doi.org/10.1007/978-1-4939-7481-8_12
33. Srinivasan, M., Sedmak, D. & Jewell, S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am. J. Pathol.* **161**, 1961–1971 (2002).
34. Dietrich, D. *et al.* Improved PCR Performance Using Template DNA from Formalin-Fixed and Paraffin-Embedded Tissues by Overcoming PCR Inhibition. *PLoS One* **8**, 1–10 (2013).
35. Guyard, A. *et al.* DNA degrades during storage in formalin-fixed and paraffin-embedded tissue blocks. *Virchows Arch.* **471**, 491–500 (2017).
36. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
37. Andrews, S. FastQC: a quality control tool for high throughput sequencing data. (2010).
38. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **7**, 2803–2809 (2011).
39. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
40. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
41. Zhou, X. *et al.* The human epigenome browser at Washington University. *Nat. Methods* **8**, 989–990 (2011).
42. Li, D., Hsu, S., Purushotham, D., Sears, R. L. & Wang, T. WashU Epigenome Browser

update 2019. *Nucleic Acids Res.* **47**, W158–W165 (2019).
43. X, W. Epigenomic Analysis of Chromatin Organization and DNA Methylation. in *Computational Biology and Bioinformatics. Gene Regulation* (ed. K, W.) (CRC Press, 2016).

Chapter 6: Exploring the coronavirus pandemic with the WashU Virus Genome Browser

The contents of this chapter have been adapted from the published manuscript:

"Flynn, J.A.*, Purushotham, D.*, Choudhary, M.N.K.*, Zhuo, X., Fan, C.*, Matt, G.*, Li, D.†, Wang, T.†, *These authors contributed equally to this work. †These authors jointly supervised this work.

6.1 Abstract

The WashU Virus Genome Browser is a web-based portal for efficient visualization of viral 'omics' data in the context of a variety of annotation tracks and host infection responses. The browser features both a phylogenetic-tree-based view and a genomic-coordinate, track-based view in which users can analyze the sequence features of viral genomes, sequence diversity among viral strains, genomic sites of diagnostic tests, predicted immunogenic epitopes and a continuously updated repository of publicly available genomic datasets.

6.2 Introduction

Coronavirus disease 2019 (COVID-19) is a rapidly spreading viral disease that has become a global health crisis. The first case of COVID-19 was reported on 12 December 2019, in Wuhan, China; by 2 August 2020, the disease had spread to more than 215 countries, territories and areas, resulting in at least 17,660,523 cases and 680,894 deaths

(https://covid19.who.int/). COVID-19 is caused by a virus called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is a zoonotic, enveloped virus containing a positive single-stranded RNA genome 29,903 base pairs in size. The virus is one of seven coronaviruses known to infect humans, along with severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV)[1,2]. To better understand the evolutionary history and pathogenesis of this new virus, large amounts of genomic data have been generated for both SARS-CoV-2 and human host cells. The genomes of thousands of SARS-CoV-2 strains have now been fully sequenced, the SARS-CoV-2 genome and transcriptome have been functionally annotated[3,4], and the genomic activity of host cells in response to viral infection is beginning to be elucidated[5,6]. This explosion of omics data has created a need for platforms to store, process, analyze and visualize these data, to gain insights into the genomic basis of viral infection. Sequencing databases, such as National Center for Biotechnology Information (NCBI) GenBank[7], and Global Initiative on Sharing All Influenza Data (GISAID)[8], currently store most of the known genome sequences of individual SARS-CoV-2 strains. The pathogen genomics platform Nextstrain currently analyzes the genomic diversity and epidemiology of a subset of these strains and provides a useful overview of the phylogeography of SARS-CoV-2 transmission[9]. Although these databases have amassed a rich source of genomic and phylogenetic information for SARS-CoV-2, comprehensive analysis of the SARS-CoV-2 genome will require the use of a high-performance genome browser designed for storing and visualizing various viral and host omic datasets.

6.3 Results

To address this need, we created the WashU Virus Genome Browser (https://virusgateway.wustl.edu), a web-based portal adapted from the WashU Epigenome

Browser[10–13], which is specifically designed for efficient visualization of viral genomic sequencing data (Appendix D.1). The browser contains the genomes of four different pathogenic virus species: SARS-CoV-2, SARS-CoV, MERS-CoV and Ebola. A reference-genome sequence is provided for each species, as well as a comprehensive collection of genome sequences of individual strains that have been isolated from patients from different geographical regions and time periods. In total, we have collected genomes of >80,000 SARS-CoV-2 strains, 332 SARS-CoV strains, 551 MERS-CoV strains and 1,574 Ebola strains, and we are continuously updating our database (Table 6.1). The genome of each strain is automatically aligned in pairwise fashion to the reference genome, and sequence variants relative to the reference are visualized through single-nucleotide variant (SNV) tracks, which provide a simple and effective way to visualize sequence variation across multiple viral strains (Figure 6.1b, Figure 6.2, Figure 6.3b, Appendix D.2). Users can search for viral strains in the browser by using the data table (Figure 6.4, Appendix D.1). Here, strains can be filtered by several metadata features, including country, continent, data source, collection date, tree-view availability and clade. Additionally, the data table's search feature allows users to locate strains of interest by querying individual accession IDs or mutations. Strains of interest can be added to the user's cart and displayed in the form of SNV tracks in the track-based browser view or highlighted in the tree-based view (Appendix D.1). To accommodate users who wish to visualize sequence variation within their own viral strains, the browser supports user upload of viral sequences in SNV format. Users are provided with easy-to-use scripts to convert their alignment results into SNV format for visualization (Appendix D.2). Additionally, the browser offers various precomputed genome-annotation tracks, such as gene annotations, protein annotations, recombination sites, sequence diversity, mutation frequencies, comparative tracks between species and GC density. Along with the

traditional track-based view, our genome browser also features a phylogenetic-tree-based view, which allows users to analyze the evolutionary relationships and history of viral strains (Figure 6.5, Appendix D.1). In this view, users can annotate strains on the phylogenetic tree with different metadata categories (for example, country of origin, collection date or Nextstrain clade designations) and can also highlight strains of interest within the tree to determine their relative relationships to other strains or clades (Appendix D.1).

The WashU Virus Genome Browser also hosts a large set of publicly available genomic datasets for both SARS-CoV-2 and human host cells. These datasets are organized as data hubs consisting of one or more tracks of data related to a specific aspect of viral biology, host biology, disease diagnosis or disease therapeutics. The browser currently hosts data hubs for viral transcription, viral recombination sites, viral RNA modifications, host transcriptional responses to infection, predicted antigenic epitopes, binding sites of diagnostic primers and genomic targets of CRISPR-based diagnostic tests (Appendix D.3, Appendix D.5). These data hubs can be easily loaded into the genome browser, providing users with an efficient way to analyze multiple data tracks of interest in tandem. As new omics data become publicly available, these datasets are promptly uploaded to the browser in data hubs. Emerging evidence now links host genomic variation with disease severity[14], and further research efforts are focused on understanding the host response to viral infection, thus necessitating a platform for visualizing cross-species genomic data. The compatibility between the WashU Virus Genome Browser and the WashU Epigenome Browser provides a platform for visualizing both viral and host genomic data. Our browser provides the unique capability of seamlessly visualizing viral genomics data and the corresponding host genomic response data, also in a modular data-hub format (Figure 6.6).

As the SARS-CoV-2 virus continues to evolve, one major task is studying how mutations accumulate within diagnostic PCR-primer-binding sites, which could potentially decrease test efficacy. To efficiently track mutation hotspots within a viral species, we developed two annotation tracks: a mutation-alert track, which displays the number of strains with a mutation at each genomic position, and a sequence-diversity (Shannon entropy) track, which displays the Shannon entropy (variation) at each genomic position across strains (Figure 6.1a). We also provide a data hub of the genomic binding sites of diagnostic PCR primers from the Centers for Disease Control and Prevention (CDC) and World Health Organization, which users can load into the genome-browser view to determine whether primers overlap with mutation hotspots or mutations in certain strains (Appendix D.3). As shown in Figure 6.1a, after loading the track for China-CDC detection primers, users can readily see a mutation hotspot overlapping primer locations within the *N* gene. Zooming into the region of interest and adding SNV tracks for individual strains offers a color-coded display of individual mutations and accession IDs of strains with a mutation at the given position (Figure 6.1b). In addition to the preloaded primer-location tracks, the browser supports user upload of novel primer locations in standard bed format[15], thereby allowing users to determine whether their primers overlap with mutations.

The WashU Virus Genome Browser can also be used to identify antigenic epitopes that are conserved across viral species. We demonstrate this utility by performing a similar analysis to that in Extended Data Fig. 5 in ref.[3], showing a genomic alignment of two SARS-CoV strains and five SARS-CoV-2 strains to the reference SARS-CoV-2 *N* gene, along with a track of putative immune epitopes identified in SARS-CoV (Figure 6.2). Peptides in SARS-CoV-2 that are homologous to putative antigenic epitopes in SARS-CoV serve as useful targets for SARS-CoV-2 vaccine development. However, the presence of sequence variants within these peptides

may compromise vaccine effectiveness (Figure 6.2b, Figure 6.2c). Motivated by the overall high sequence similarity between the SARS-CoV and SARS-CoV-2 genomes, we analyzed the full list of experimentally identified linear epitopes from the Immune Epitope Database and Analysis Resource (IEDB)[16], and identified a list of 320 high-confidence linear epitopes whose amino acid sequences are identical to those of the predicted translated products from the SARS-CoV-2 reference strain (Appendix D.6). This list provides a catalog of epitopes for researchers testing immunological targets that can potentially elicit T-cell and B-cell responses to SARS-CoV-2.

The browser supports multiple file formats, thus allowing users to visualize sequencing data in a variety of ways to better understand viral sequence evolution. Recent work suggests that the amino acid change from aspartate to glycine at position 614 within the spike (S) protein of SARS-CoV-2 has become dominant in non-Asian countries[17]. Although this mutation is not located within the core receptor-binding domain of the protein, it is nonetheless thought to contribute to the transmissibility of the virus[17]. In the browser track view, the prevalence of the p.Asp614Gly alteration among SARS-CoV-2 strains is immediately evident when viewing the preloaded sequence-diversity (Shannon entropy) track and the mutation-alert track, which highlight the entropy across strains and the number of strains with mutations, respectively (Figure 6.3a). A finer-scale visualization of the p.Asp614 codon within the browser shows that some accessions of non-Asian origin contain the p.Asp614Gly alteration, and a dynamic track of the mutation's weekly prevalence shows that more than 90% of strains isolated during the week of 27 April 2020 contained the p.Asp614Gly alteration (Figure 6.3b). Further characterization of the prevalence confirmed the higher frequency of the p.Asp614Gly alteration in non-Asian accessions (Figure 6.3c) and the increase in prevalence of the p.Asp614Gly alteration over time (Figure 6.3d).

Using a similar approach, we were also able to confirm the recent observation of a heightened mutation rate at the amino terminus of the S protein among betacoronaviruses[3]. The genomes of pangolin and bat coronaviruses were aligned to the SARS-CoV-2 reference genome (Appendix D.4), and sequence variation was displayed in the browser in the form of SNV tracks and genome-comparison tracks (Figure 6.7, Appendix D.2). This view showed that the 5′ end of the *S* gene is highly divergent between the pangolin coronavirus and SARS-CoV-2 (Figure 6.7). This finding was further supported by displaying raw next-generation sequence reads in bam format in the browser. Coronavirus reads from a pangolin viral metagenomic dataset[18] were extracted and aligned to the SARS-CoV-2 reference genome (Appendix D.4). Most pangolin reads (1,468 of 2,288) aligned to the SARS-CoV-2 genome, and the average mismatch rate between pangolin-CoV and the SARS-CoV-2 sequences was only ~7% (Appendix D.4), thus indicating that the pangolin-CoV dataset is closely related to the SARS-CoV-2 dataset. Nevertheless, we found that no reads from the pangolin library aligned to the 5′ end of the *S* gene of SARS-CoV-2 (Figure 6.7), in agreement with the observation that the N-terminal end of the S protein is one of the most divergent regions among betacoronaviruses[3].

All analyzed viral sequences are available from the NCBI GenBank[7] (https://www.ncbi.nlm.nih.gov/nuccore), GISAID[8] (https://www.gisaid.org/) and Nextstrain[9] (https://nextstrain.org/sars-cov-2) public repositories, with the exception of the pangolin viral metagenomic dataset[18], which is available at NCBI BioProject PRJNA573298). Additional datasets used to create public data hubs hosted in the browser are listed in Appendix D.5.

Notably, the UCSC SARS-CoV-2 Genome Browser has recently been developed in parallel to the work described here[19], highlighting the need for comprehensive omic visualization resources as well as community interest and contribution. We hope that the WashU Virus

Genome Browser will enable rapid sharing of processed data, facilitate collaboration and accelerate research on existing and novel pathogenic viruses. Moreover, the portable nature of the underlying technology enables us to swiftly spin up viral browser instances in response to other emerging zoonotic viruses. Our browser portal can be accessed at https://virusgateway.wustl.edu; documentation is available at https://virusgateway.readthedocs.io/; and general feedback, suggestions and bug reports may be sent to https://github.com/twlab/virusbrowser/issues.

6.4 Acknowledgements

J.A.F. is supported in part by the Siteman Cancer Center Precision Medicine Pathway (T32CA113275). X.Z. is supported in part by NIH 5R25DA027995. T.W. is supported by NIH grants R01HG007175, U24ES026699, U01CA200060, U01HG009391 and U41HG010972, and by American Cancer Society Research Scholar grant RSG-14-049-01-DMC.

6.5 Author Information

These authors contributed equally: Jennifer A. Flynn, Deepak Purushotham, Mayank N. K. Choudhary, Xiaoyu Zhuo, Changxu Fan, Gavriel Matt.
These authors jointly supervised this work: Daofeng Li, Ting Wang.

6.5.1 Affiliations

The Edison Family Center for Genome Sciences & Systems Biology, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA

- Jennifer A. Flynn, Deepak Purushotham, Mayank N. K. Choudhary, Xiaoyu Zhuo, Changxu Fan, Gavriel Matt, Daofeng Li & Ting Wang

McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA

- Ting Wang

6.5.2 Contributions

Conceptualization, T.W.; web development, D.L. and D.P.; SNV track development, J.F. and C.F.; dynamic track development, M.N.K.C. and D.L.; immune-epitope analysis, M.N.K.C.; data download, metadata generation and annotation, G.M., D.L. and C.F.; data-hub preparation, J.F., C.F., M.N.K.C. and G.M.; sequence alignment and tree generation, C.F. and X.Z.; manuscript preparation, J.F., C.F., M.N.K.C., G.M., D.L. and T.W.

6.5.3 Corresponding authors

Correspondence to Daofeng Li or Ting Wang.

6.6 Figures

Figure 6.1 Evaluating PCR primers for accumulating mutations with the WashU Virus Genome Browser.

a. Genome-browser view of the entire SARS-CoV-2 genome with the following tracks loaded (from top to bottom: gene annotations, ruler, sequence diversity, mutation alert and China CDC primer-binding sites). The sequence-diversity and mutation-alert tracks are two different representations of sequence diversity among SARS-CoV-2 strains: the former shows the Shannon entropy at each nucleotide position, and the latter shows the number of strains with a mutation at each nucleotide position. Primer-binding sites (red bars) within the *N* gene appear to overlap a region with a moderate level of sequence diversity among SARS-CoV-2 strains. The browser's magnification tool (circled in red) can be used to quickly zoom in to the region of interest. b. A zoomed-in view of the region described in a, showing that the binding site of the forward primer ('F-N') is indeed highly variable among SARS-CoV-2 strains. SNV tracks of

SARS-CoV-2 strains show three strains with mutations and three strains lacking mutations in the primer-binding site. Clicking on mutations in the mutation-alert track shows the total number of strains mutated at the selected position and the accession numbers of these strains (pop-up display window).

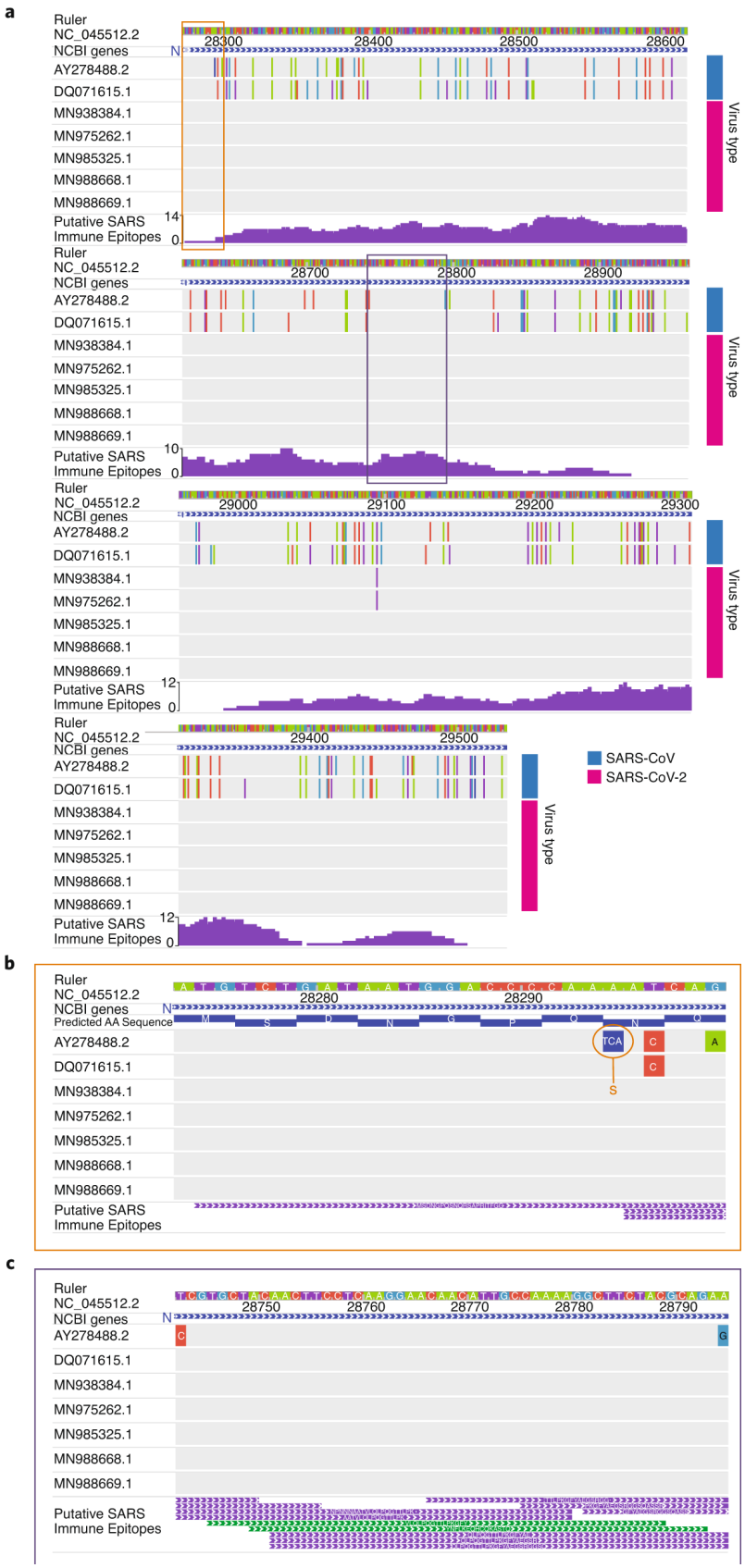

a
Ruler
NC_045512.2
NCBI genes
AY278488.2
DQ071615.1
MN938384.1
MN975262.1
MN985325.1
MN988668.1
MN988669.1
Putative SARS Immune Epitopes
Virus type
SARS-CoV
SARS-CoV-2
b
Predicted AA Sequence
c

Figure 6.2 Using the WashU Virus Genome Browser to discover conserved immune-epitopes.
a. Browser view of the SARS-CoV-2 gene encoding the nucleocapsid (N) protein with SNV tracks of two SARS-CoV strains (DQ071615.1 and AY278488.2) and five SARS-CoV-2 strains (MN938384.1, MN975262.1, MN985325.1, MN988668.1 and MN988669.1) loaded in the view. The track of putative SARS-CoV immune epitopes is also displayed in density mode. b. A zoomed-in view of the orange box in a, displaying the first 9 amino acids of the SARS-CoV-2 N protein. SNVs at positions 28296 (T>C) and 28299 (G>A) are silent mutations; however, the 'TCA' insertion at position 28294 in SARS-CoV accession AY278488.2 (BJ01) results in the insertion of a serine residue in the SARS-CoV N protein relative to the SARS-CoV-2 N protein. Of note, this insertion is not present in the other SARS-CoV accession (DQ071615.1). Owing to the variability in the amino acid sequence within this region across SARS-CoV and SARS-CoV-2 strains, this region is unlikely to be a good candidate for epitope design. c. A zoomed-in view of the purple box in a, displaying a region that is likely to be a good candidate for epitope design because it is fully conserved across SARS-CoV and SARS-CoV-2 strains, and it also encodes several putative antigenic epitopes.

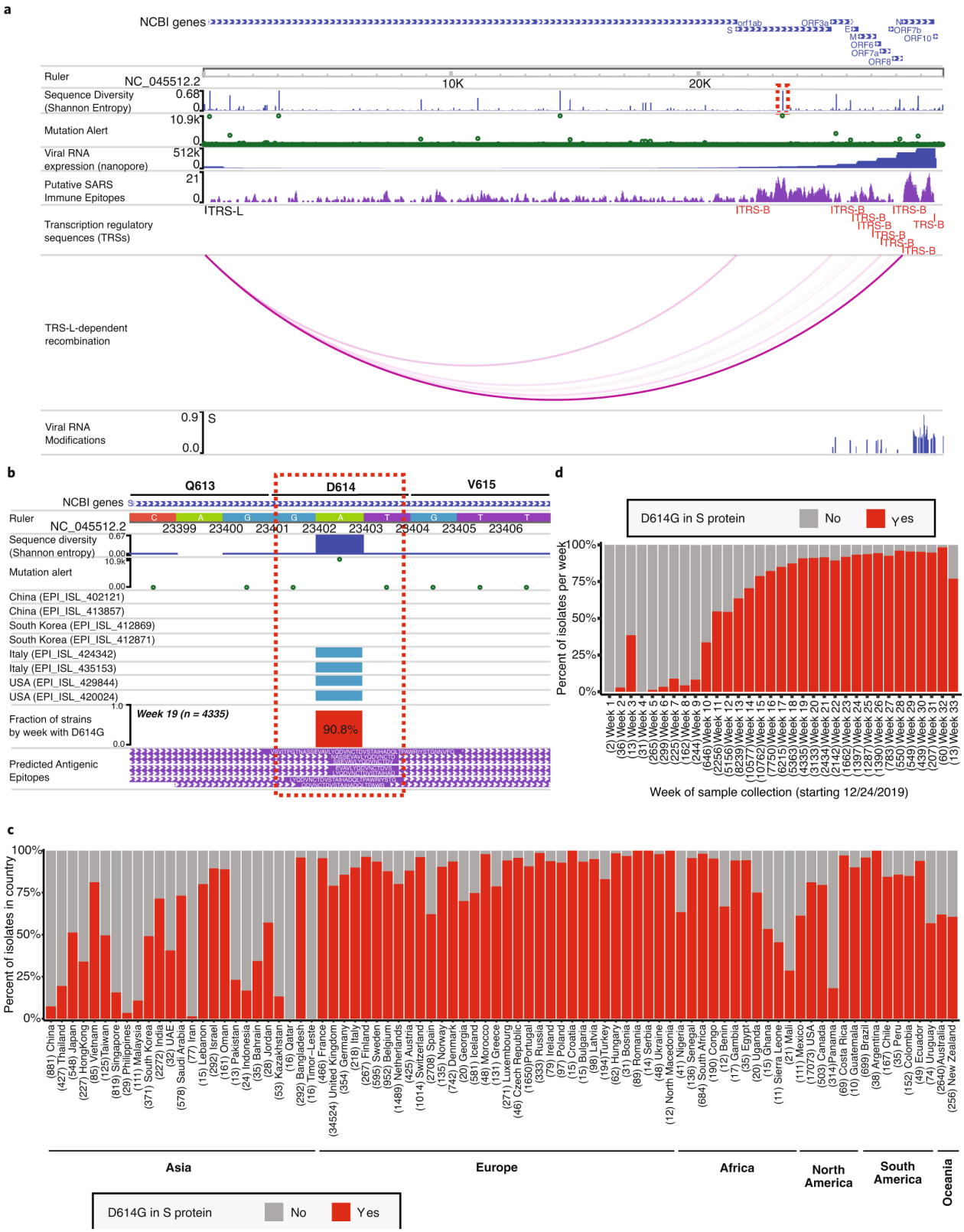


Figure 6.3 Using the WashU Virus Genome Browser to study the prevalence of the S-protein p.Asp614Gly alteration.

a. Default browser view of the SARS-CoV-2 genome. The SARS-CoV-2 genome browser includes various genomic datasets in track format (from top to bottom: gene annotations, sequence diversity, mutation alerts, gene expression, predicted immune epitopes, recombination events and RNA modifications). The p.Asp614Gly alteration within the S protein is circled in red within the sequence-diversity track. b. A zoomed-in view of the p.Asp614Gly alteration circled in a. The ruler track displays a color-coded nucleotide sequence. The sequence-diversity and mutation-alert tracks reveal a high degree of variation across SARS-CoV-2 strains at this location. Sequence variation within individual strains is displayed below in the form of SNV tracks, which report variations relative to the reference genome in a color-coded format. Notably, mutations at this position appear to be enriched in non-Asian accessions. Beneath these tracks is a dynamic track showing the weekly percentage of strains with the p.Asp614Gly alteration (beginning on 24 December 2019). Finally, the bottom track displays predicted antigenic linear epitopes that were experimentally identified in SARS-CoV by using T-cell, B-cell and major-histocompatibility-complex ligand assays[13]. c. Plot showing the percentage of SARS-CoV-2 isolates in each country that contain the p.Asp614Gly alteration (data retrieved from GISAID, accessed on 19 August 2020)[8]. Countries are ordered by continent and then by date of virus introduction. Countries with fewer than ten cases are excluded. Sample sizes are displayed in parentheses beneath each country. d. Plot showing the percentage of SARS-CoV-2 isolates per week (beginning on 24 December 2019) that contain the p.Asp614Gly alteration. Sample sizes for each week are displayed in parentheses.

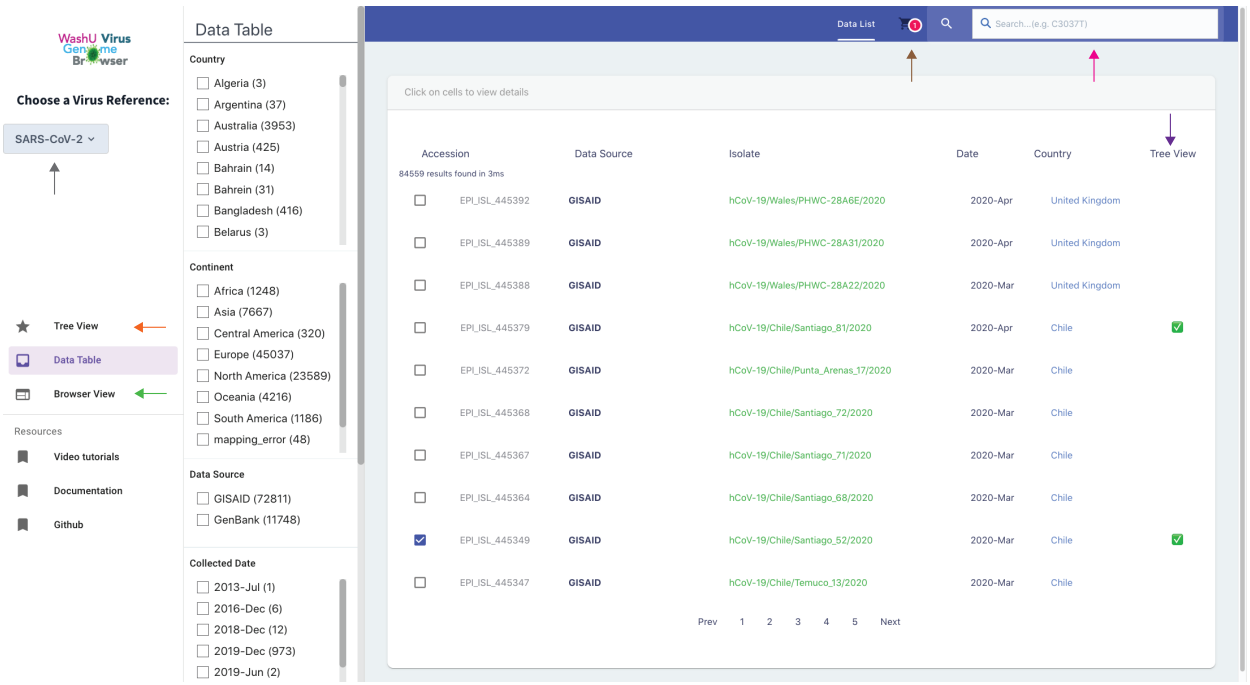


Figure 6.4: The WashU Virus Genome Browser data table.
The "Data Table" view displays all of the strains that are available in the browser for a given virus species. From this page, the user can select which reference genome to use (gray arrow), and which individual strains to view based on several sortable metadata features, including country, continent, data source, collection date, tree view availability, and clade membership. The user can also search for specific accession IDs, isolate names, and strains with a particular SNV (pink arrow). Selected strains of interest will be added to the user's cart (brown arrow) and can then be viewed in the traditional track-based genome browser view (green arrow) or the phylogenetic tree view (orange arrow). In the genome browser view, strains will show up as pairwise alignments to the reference genome (i.e. SNV tracks). In the tree view, strains will be highlighted within the phylogenetic tree. Please note that for SARS-CoV-2, only strains present in the Nextstrain database[9] are available for viewing in the tree view (as indicated by the metadata column, "Tree View" (purple arrow)).

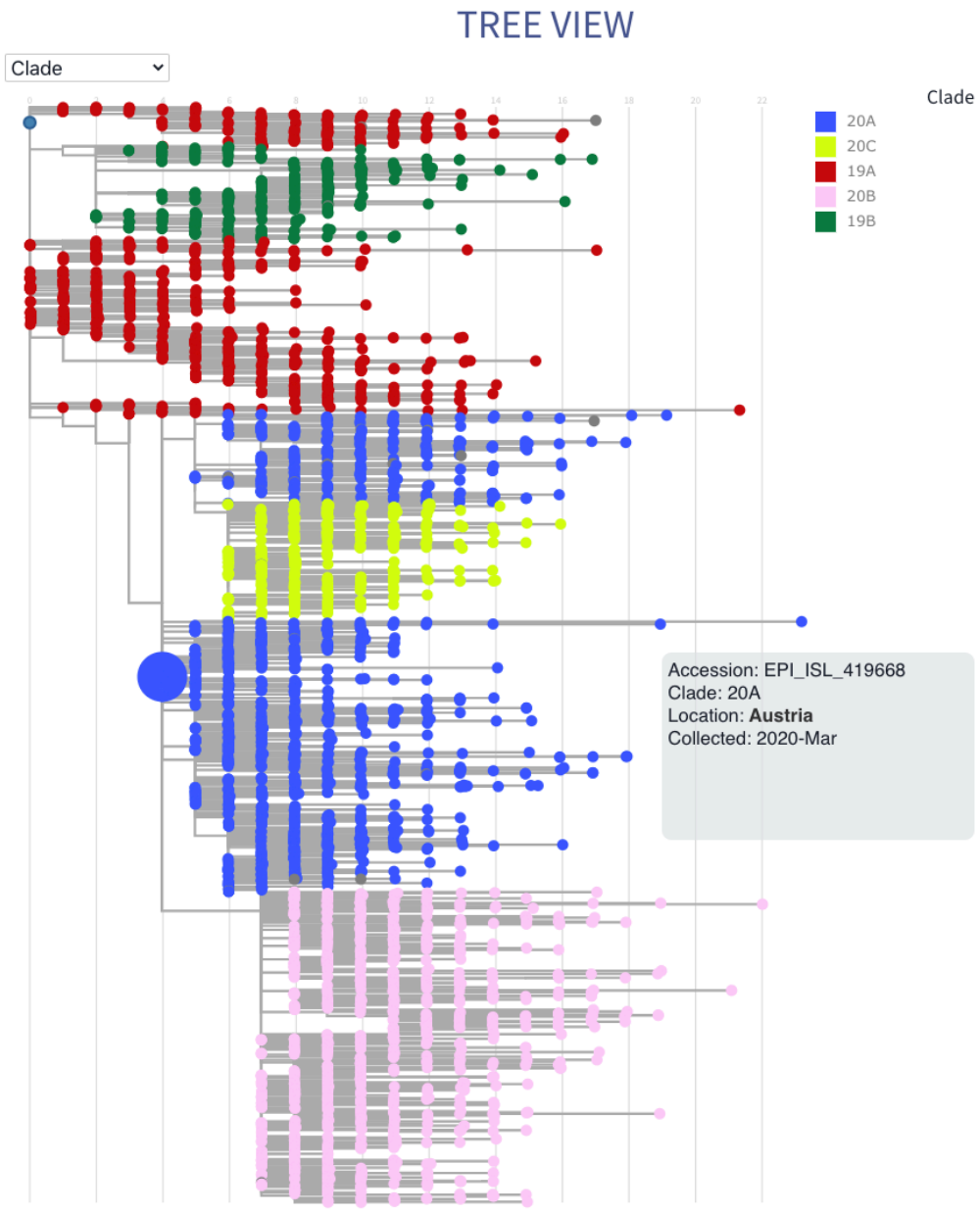


Figure 6.5: The WashU Virus Genome Browser tree view.
Phylogenetic tree view of all SARS-CoV-2 sequences available through Nextstrain. The default view colors individual strains by clade status, but can be changed by the user to reflect collection date or location (Appendix D.1). Strains in the user's cart will be highlighted in this view as an enlarged dot. Note that only strains from the Nextstrain database can be viewed in the tree view. Metadata information (e.g. accession ID, clade designation, isolate name, location, and collection date) about each strain within the tree can be viewed by clicking on the strain (see pop-up display box). The phylogeny is provided by Nextstrain (http://data.nextstrain.org/ncov_global.json)[9].

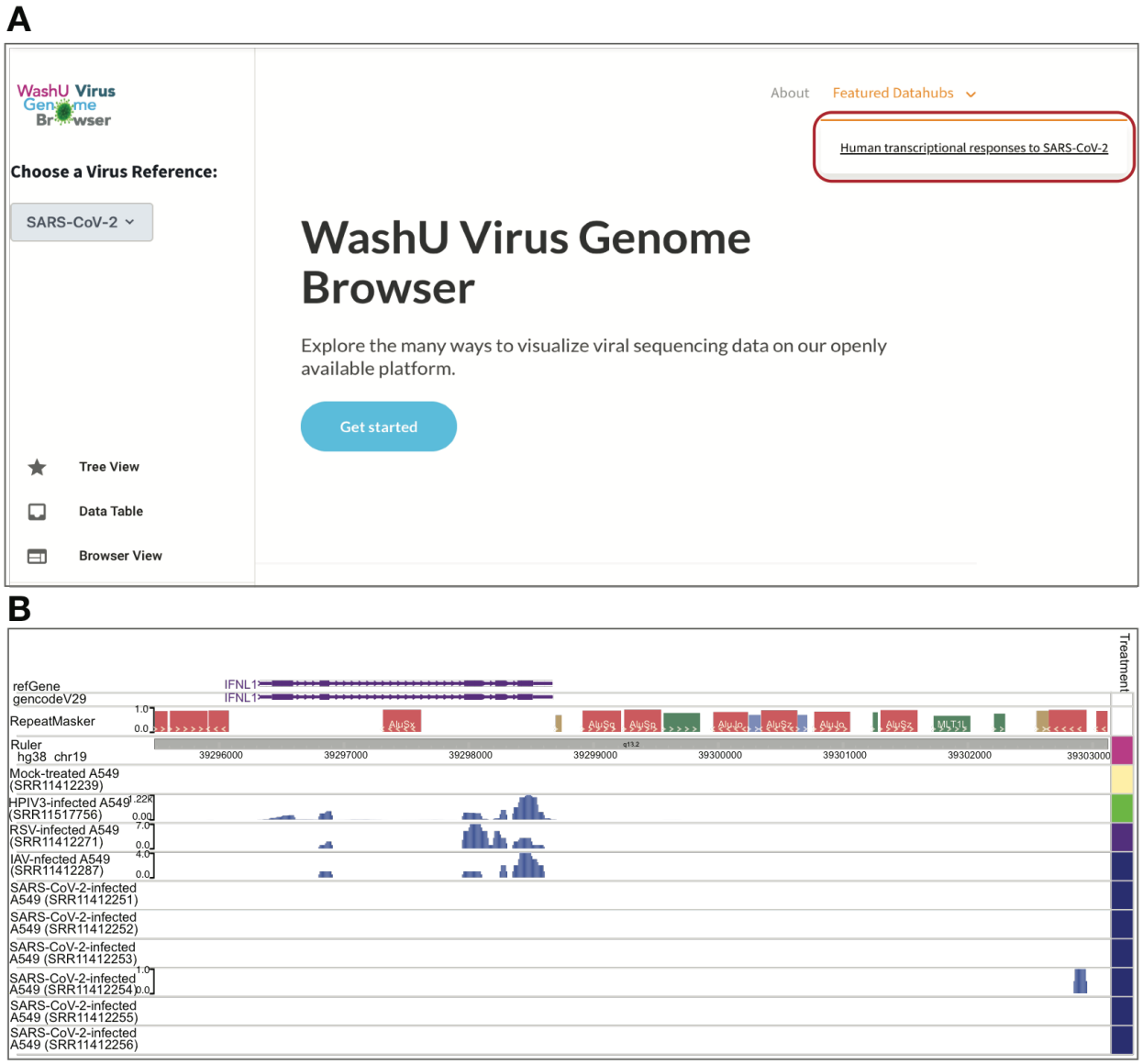


Figure 6.6: Using the WashU Virus Genome Browser to explore host transcriptional responses to SARS-CoV-2 infection.

(A) When navigating to the WashU Virus Genome Browser landing page, host transcriptional responses to SARS-CoV-2 infection can be viewed on the WashU Epigenome Browser[10–13] by selecting the drop-down menu "Featured Datahubs" and then "Human transcriptional response to SARS-CoV-2" (circled in red). (B): Once selected, the user is redirected to the hg38 genome where 12 RNA-seq tracks are pre-loaded into view. The complete data hub contains a total of 195 RNA-seq tracks which can be added to view (Appendix D.3). In this view, RNA-seq tracks are displayed for A549 cells that were infected with either HPIV3, RSV, IAV, SARS-CoV-2, or no virus (mock)[5]. Expression is shown over the *IFNL1* gene, revealing high expression in HPIV3-infected cells, low expression in RSV-infected and IAV-infected cells, and no expression

in mock-infected and SARS-CoV-2-infected cells. This result suggests that interferon pathways are suppressed during SARS-CoV-2 infection.

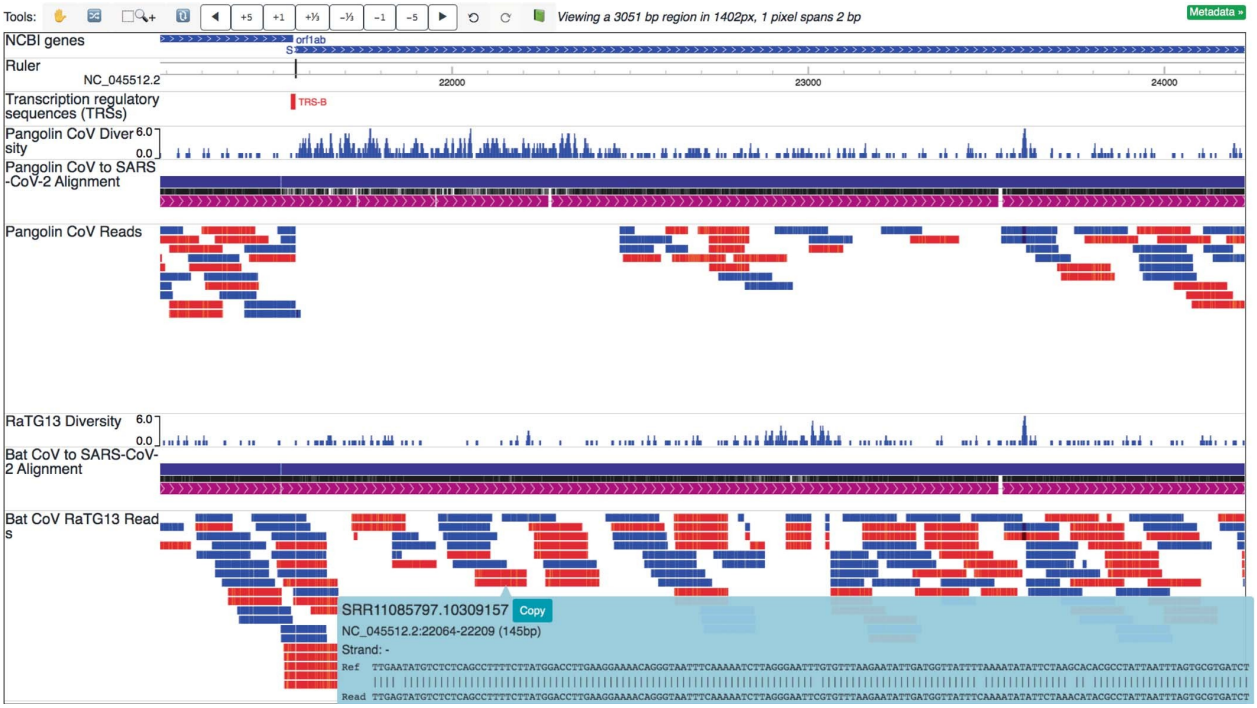


Figure 6.7: Using the WashU Virus Genome Browser to study sequence conservation across viral species.

Genome-browser view of the 5′ end of the SARS-CoV-2 S gene. Tracks loaded in the view are as follows from top to bottom: gene annotations, ruler, transcription regulatory sequence (TRS) sites, pangolin coronavirus (EPI_ISL_410721) comparison tracks (SNV track, genome-alignment track and read-alignment track) and bat coronavirus (EPI_ISL_402131) comparison tracks (SNV track, genome-alignment track and read-alignment track). In the SNV track (displayed in density mode), the *y* axis represents the density of sequence variation in bat or pangolin coronaviruses compared with the SARS-CoV-2 reference. In the genome-alignment track, the SARS-CoV-2 reference is represented by the solid blue band, and the query sequence (in this case, the bat or pangolin coronavirus genome) is represented by the purple band. Black bars between the reference and the query represent matches, whereas the absence of a bar represents a variant (mismatch or gap). A gap in the reference or query is represented by a gap in the blue or purple band, respectively. In the read-alignment track, data are loaded in bam format. Each read is displayed as a colored bar (blue, plus strand; red, minus strand), and gaps within a bar represent mismatches to the reference. The alignment of a read to the reference can be displayed by selecting a particular read (pop-up display window). This genome-browser view shows that the 5′ end of the S gene is highly variable between SARS-CoV-2 and the pangolin coronavirus, but not between SARS-CoV-2 and the bat coronavirus.

6.7 References

1. De Wit, E., Van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: Recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **14**, 523–534 (2016).
2. Cui, J., Li, F. & Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).
3. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
4. Kim, D. *et al.* The Architecture of SARS-CoV-2 Transcriptome. *Cell* **181**, 914-921.e10 (2020).
5. Blanco-Melo, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**, 1036-1045.e9 (2020).
6. Bojkova, D. *et al.* Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **583**, 469–472 (2020).
7. Agarwala, R. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, D8–D13 (2018).
8. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* **22**, 2–4 (2017).
9. Hadfield, J. *et al.* NextStrain: Real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
10. Li, D., Hsu, S., Purushotham, D., Sears, R. L. & Wang, T. WashU Epigenome Browser update 2019. *Nucleic Acids Res.* **47**, W158–W165 (2019).
11. Zhou, X. *et al.* Epigenomic annotation of genetic variants using the Roadmap Epigenome Browser. *Nat. Biotechnol.* **33**, 345–346 (2015).
12. Zhou, X. *et al.* Exploring long-range genome interactions using the WashU Epigenome Browser. *Nat. Methods* **10**, 375–376 (2013).
13. Zhou, X. *et al.* The human epigenome browser at Washington University. *Nat. Methods* **8**, 989–990 (2011).
14. Van Der Made, C. I. *et al.* Presence of Genetic Variants among Young Men with Severe COVID-19. *JAMA - J. Am. Med. Assoc.* **324**, 663–673 (2020).
15. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
16. Vita, R. *et al.* The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2019).
17. Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812–827 (2020).
18. Liu, P., Chen, W. & Chen, J. P. Viral metagenomics revealed sendai virus and coronavirus infection of malayan pangolins (manis javanica). *Viruses* **11**, (2019).
19. Fernandes, J. D. *et al.* The UCSC SARS-CoV-2 Genome Browser. *Nat. Genet.* **52**, 986–998 (2020).

Chapter 7: Conclusions and Future Directions

The overarching goal of my dissertation was to better understand how epigenomic changes contribute to tumorigenesis and cancer progression. My thesis started by comparing DNA methylation changes over enhancers across distinct cancer types in order to identify commonalities and cancer-specific enhancer methylation signatures. The main result of this study was the observation that both endometrial adenocarcinoma (EAC) and glioblastoma multiforme (GBM) tumors exhibited recurrent increases in methylation over enhancers that were likely related to maintaining original cell type transcriptional programs[1]. Recurrent EAC enhancer hyperDMRs contained groupings of enriched transcription factor binding motifs. Enriched disease ontology pertaining to the majority of differentially methylated enhancer groups largely encompassed uterine-specific terms, suggesting that enhancers gaining methylation in EAC might contribute to the regulation of normal uterine function (Figure 2.9). In a similar analysis, GBM hyperDMRs were found to encompass significantly more enhancers active in adult brain as compared to fetal brain (Figure 2.10). Further examination of brain enhancer hyperDMRs revealed increased H3K27ac and H3K4me1 signal in adult brain compared to fetal brain, suggesting that the majority of enhancers being methylated were specific to the adult brain.

While these findings are consistent with a mechanism whereby cancer cells silence their original cell type-specific programming by increasing methylation over enhancers, the causal relationship remains to be established. In fact, an unanticipated finding of our analysis was that transcription factors that contained enriched binding motifs in EAC hyperDMRs showed a loss of expression in EAC samples relative to normal endometrium. Likewise, transcription factors

containing enriched binding motifs in EAC hypoDMRs were more highly expressed in EAC relative to normal. These findings suggest that it is in fact loss of transcription factor occupancy that predicates the increased methylation observed over these enhancers. Future studies aimed at understanding the loss of cell identity in primary cancers should work to tease apart the mechanism underpinning original cell type enhancer silencing. For example, technology has now enabled targeted genome editing *in vivo*[23]. A potential follow-up experiment that would begin to address whether or not the presence of DNA methylation over the enhancer was contributing to the transcriptional state of the target gene could be to performed targeted removal of DNA methylation over a candidate enhancer region and assess the impact on the target gene's transcription. An additional outstanding question remains: regardless of the mechanism of silencing, how does the loss of original cell type-specific transcriptional programming contribute to cancer (or is this just a passenger event)? Current hypotheses speculate that the loss of original cell type features may promote transition back to a more undifferentiated, stem-like state[4]. Further studies are needed to assess if and how this might be achieved.

Knowing that enhancers contribute heavily to the maintenance of cell type-specific gene expression, that enhancers functions are often altered in cancer, and that methylation changes frequently accompany enhancer abnormalities, we next sought to determine if epigenetic changes, particularly over enhancers, might play a role in metastatic organotropism. We were specifically interested in gaining a better understanding of how non-small cell lung cancer (NSCLC) routinely, successfully colonizes the brain. While previous studies have compared genomic mutations in primary NSCLC to those of brain metastases, few (if any) metastasis-specific mutations have been identified[5–9]. Since genomic, transcriptomic, and methylomic alterations may each provide functionally equivalent outcomes, we expanded upon previous

analyses by performing whole exome-seq, RNA-seq, and MeDIP-seq plus MRE-seq on subsets of paired primary and brain metastasis samples from a cohort of 45 patients. While only four genes were significantly mutated in metastasis samples but not primaries (*POLDIP2*, *KRAS*, *KEAP1*, and *DCDC1*), the paired nature of our dataset allowed for the identification of metastasis enriched variants (variants with a variant allele frequency (VAF) of at least 30% in the metastasis and less than 10% in the primary tumor that had at least 30x coverage), resulting in 75 genes with recurrent MEVs. Pathway analysis of recurrent MEVs identified enriched processes, the two most significantly enriched being focal adhesion and extracellular matrix receptor interactions (Appendix B.2). One MEV recurrent in 3 patients fell within *TET1*, an active DNA demethylase (Figure 3.3), prompting us to look at VAF changes across epigenetic regulators as a whole, regardless of their individual recurrence. While many gene VAFs were private to individual patients, the signal stemming from examining this set of genes as a whole demonstrated a large increase in VAFs in metastases (Figure 3.7), suggesting that epigenetic misregulation may be selected for in metastasis. Consistent with this idea, methylation profiling across patients revealed a substantial degree of variation in the number of differentially methylated regions (DMRs) identified (Figure 3.8). However, recurrent hyperDMRs were found overlapping brain-specific active enhancers, correlating with increased nearby gene expression. Transcriptome analyses corroborated these finding, revealing an overall increase of expression of genes belonging to brain development pathways and a loss of immune-related transcriptional programming. Together the results of this study suggest that a contributing factor of NSCLC's successful colonization of brain is related to activating developmental brain transcriptional programming.

The findings reported in our study are consistent with an overall epigenetic reprogramming of NSCLC towards a more "brain-like" state. There are, however, many additional follow-up experiments necessary to fully understand how this process may be occurring. One of the largest limitations with this project is the relatively small sample size. Given the nature of brain metastatic disease, resection is not always advised, limiting the availability of samples. While we have been fortunate to have a cohort of 45 patients from which samples were available, statistical claims can be challenging, especially when looking across the entire genome (which increases the number of hypotheses needing correction), necessitating repeated analyses once larger cohorts become available. Another confounding factor present in all bulk cancer sample analyses is heterogeneity within tumors stemming from the presence of different tumor clones as well as infiltrating non-diseased cells. One of the major challenges of the present study was determining whether or not the "brain-like" signatures that were appearing in our metastasis samples were truly from transformed tumor cells or from contaminating brain tissue. To address this question, we asked whether or not somatic mutations identified in exome-seq data were also present in transcripts coming from over-expressed brain genes. While we could not recover somatic mutations in all transcripts (often due to low coverage), many different mutations across several different patients were recovered in the RNA-seq data (Table 8.1), arguing that the metastasis samples were in fact exhibiting an increased brain-expression signature. Like proposed follow-up experiments from Chapter 2, a better understanding of methylation's role within brain-specific enhancers in regulating nearby gene expression is needed. In this case, a conceivable experiment would be to perform targeted methylation of an enhancer that frequently undergoes hypomethylation in metastasis and determine if there is subsequent altered expression. In addition, mutational status and expression patterns of genes

with enriched motifs within these enhancers might also lend clues as to the order of causal events.

The largest recurrent methylation changes from primaries to metastases resided within DNA methylation valleys (DMVs) (Figure 3.8). Upon further characterization of these regions, we observed that the subset of DMVs gaining methylation as a function of disease progression were enriched for the histone mark H3K9me3 as well as the bivalent marks H3K27me3 and H3K4me1 in normal lung (Figure 4.6). The genes encompassed within DMVs are largely developmental, and the enrichment of H3K27me3 suggested that they are repressed by polycomb repressive complex 2 (PRC2) in normal lung. Publicly available ChIP-seq data[10] probing the genomic binding locations of EZH2, the catalytic subunit of PRC2, confirmed that these regions were indeed occupied by PRC2 in a normal lung context. This observation lead to a series of new questions, beginning with asking whether or not PRC2 binding over DMVs was lost in metastasis since DNA methylation and H3K27me3 have demonstrated mutual exclusivity in CpG islands[11]. To begin addressing this question, we shifted our focus to the non-small cell lung cancer lymph node-derived cell line, H1299, and profiled the binding of EZH2. Our results indicated that the increase in DNA methylation over DMVs was accompanied by a loss of EZH2 binding (Figure 4.8). Finally, developmental genes within DMVs that exhibited increased methylation (and likely PRC2 binding loss) displayed what might be a general, low-level increase in expression (Figure 3.8), although significance is lacking, likely due to the variable quality of RNA-seq data. These results lead me to the following working hypothesis, outlined in Figure 7.1. As normal cells transition to cancer and further to metastatic disease, there is an increase in polycomb group protein mutations (both novel and in the form of increased VAFs) (Figures 3.7 and 4.6). As a result, polycomb complexes dissociate from their positions in normal

lung over developmental genes. DNA methlytransferases are then able to bind and increase DNA methylation over the regions. Finally, loss of polycomb repression leads to the reactivation of development genes which is not fully remediated by the increased DNA methylation.

As stated above, this is a working hypothesis, with much needed follow-up attention. The first question that needs to be more definitively addressed is whether or not developmental genes loosing polycomb repression regain their expression despite increased DNA methylation. Since many clinical samples come from formalin-fixed paraffin-embedded tissues, RNA integrity is often compromised, resulting in inconclusive data. Therefore, repeated experiments would greatly benefit from the use of fresh-frozen tissue samples. If our preliminary findings of increased expression over DMV genes can be confirmed, the next question to address would be whether DNA methylation displaces polycomb binding or polycomb dissociation allows for DNA methylation to accumulate. Increased mutations within polycomb genes would suggest that perhaps it is the dissociation of polycomb that comes first; however, variants of both PRC1 and PRC2 contain subunits that target the complexes to unmethylated DNA sequences, suggesting that increased DNA methylation might prevent their binding[12]. To begin addressing this question, a first experiment could be to perform targeted DNA demethylation within a DMV in H1299 and determine if there is a subsequent re-gain of polycomb binding and loss of gene expression. Another experiment would be to perform a forced recruitment of PRC2 in H1299 to a region that had gained methylation and previously lost polycomb binding. The success of the forced recruitment could shed some light on DNA methylation's ability to exclude polycomb binding. In a similar manner, targeted displacement of polycomb in a non-malignant lung cell line could be performed and the effects on DNA methylation accumulation and expression changes measured. In addition, the success of targeted recruitment of DNMTs to DMVs

repressed by polycomb in a non-malignant lung cell line may offer some indication of DNA methylation's capacity to displace polycomb occupancy. In addition to assaying the mechanism responsible for the epigenetic switch, it is conceivable to also test the effects DNA methylation increase, polycomb binding decrease, and resulting expression changes on phenotype in the form of growth rate comparisons, invasiveness (scratch assays) and brain matrix survival.

The work presented in this thesis pertaining to the loss of PRC2 binding over DMVs in metastases may help inform future approaches to epigenetic therapy. As described in Chapter 1, EZH2 is frequently over expressed and exhibits gain-of-function mutations in a variety of cancers, leading to the repression of several tumor suppressor genes. This phenomenon has sparked interest in designing drugs to inhibit EZH2, with the first EZH2 inhibitor recently gaining FDA approval[13,14]. The work put forth in this thesis suggests that tumorigenic properties of polycomb misregulation may not be strictly in the form of increased binding over tumor suppressors, but rather in a redistribution of binding. Therefore, while EZH2 inhibitors might relieve repression of some TSGs, it also seems possible that they might additionally perpetuate the loss of PRC2 binding, and therefore interfere with maintenance, of developmental genes. As described above, additional work will be required to confirm increased expression of DMV genes as well as determine if PRC2 dissociates from developmental genes prior to methylation increase.

In conclusion, widespread epigenetic changes occur in both primary tumorigenesis and metastatic disease. These methylation changes are non-randomly dispersed throughout the genome and often target enhancers responsible for tissue-specific gene regulation. In primary cancer, we have shown that enhancers related to the original cell type exhibit increased methylation, whereas in NSCLC brain metastases, brain-specific enhancers are demethylated and

correlate with increased nearby brain gene expression. Additional methylation changes in metastases are found over DMVs where an apparent epigenetic switch occurs: there is a gain of methylation and a concordant loss of polycomb repression. While the implications of this epigenetic switch need further follow-up, preliminary data suggest that developmental genes encompassed within these DMVs may increase in expression, possibly contributing to a more stem-like cancer cell phenotype.

7.2 Figures

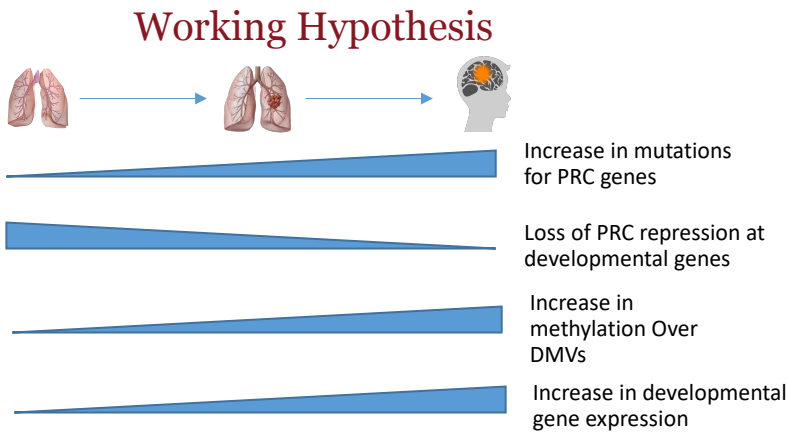


Figure 7.1 Schematic of current working hypothesis of epigenetic misregulation of DNA methylation valleys in cancer.
As normal lung progresses to non-small cell lung cancer and further to brain metastasis, there is an increase in polycomb group proteins mutations, an accompanying loss of polycomb binding replaced by acquired DNA methylation, potentially resulting in an increase of target developmental gene expression.

7.3 Tables

Table 7.1 Somatic mutations recovered in brain-related gene transcripts

Gene Name	Chromosome	Start Position	End Position	Variant Classi	Variant Type	Reference All	Variant Allele	Sample ID	Exome Sequencing Depth	Exome-Seq VAF	Total RNA-seq Reads	RNA-Seq VAF
FBXL16	16	745558	745558	Silent	SNP	G	T	707076_17814_BM	76	21.05263	1	1
DDX17	22	38882342	38882342	Silent	SNP	C	G	707076_17814_BM	83	77.10843	1	1
ATP1B1	1	169076067	169076067	De_novo_Star	SNP	C	A	707076_702524_BM	22	43.86839	31	0.38709677
TAGLN3	3	111732414	111732414	3'UTR	SNP	C	T	707076_702524_BM	47	50.26738	46	0.47826087
KCNMB2	3	178560613	178560613	Nonsense_Mu	SNP	C	A	707076_8062_BM	168	39.88095	1	1
LENG8	19	54965830	54965830	Missense_Mut	SNP	G	C	707076_8062_BM	206	13.59223	85	0.25882353
BEX2	X	102565355	102565355	5'UTR	SNP	C	A	707076_8062_BM	26	103.1746	1	1
CPE	4	166405675	166405675	Missense_Mut	SNP	C	A	AugMoffitt003_BM	104	24.03846	1	1
DNM1	9	130985144	130985144	Intron	SNP	G	T	AugMoffitt004_BM	7	111.1111	1	1
PRRT2	16	29825059	29825059	Missense_Mut	SNP	G	C	AugMoffitt004_BM	171	11.69591	10	0.3
RHOB	2	20647816	20647816	Silent	SNP	G	A	um2b	27	11.11111	42	0.28571429
ZXDC	3	126180480	126180480	Missense_Mut	SNP	C	A	um2b	80	18.75	20	0.7
APLP2	11	130003551	130003551	Missense_Mut	SNP	G	C	um2b	108	5.555556	4	0.25
MBP	18	74729065	74729065	Missense_Mut	SNP	G	A	um4b	35	51.42857	24	0.58333333
DVL3	3	183881451	183881451	Silent	SNP	G	T	WU102B	37	62.16216	5	0.2
CALM2	2	47399532	47399532	Intron	SNP	C	A	WU104B	57	8.77193	2	0.5
SCG2	2	224462715	224462715	Missense_Mut	SNP	T	C	WU104B	139	18.70504	35	0.4
CDR1	X	139866382	139866382	Silent	SNP	A	T	WU104B	115	28.69565	12	0.16666667
RAB6B	3	133547648	133547648	Missense_Mut	SNP	C	A	WU105B	55	40	2	0.5
LENG8	19	54964739	54964739	Missense_Mut	SNP	C	A	WU105B	13	46.15385	5	0.4
APP	21	27347466	27347466	Missense_Mut	SNP	C	A	WU105B	33	66.66667	12	1
APP	21	27347467	27347467	Missense_Mut	SNP	C	A	WU105B	33	66.66667	12	1
CDR1	X	139865895	139865895	Nonsense_Mu	SNP	C	A	WU105B	283	93.63958	83	0.87951807
APLP2	11	130005503	130005503	Missense_Mut	SNP	C	T	WU107_BM	88	40.90909	13	0.38461538
APP	21	27354695	27354695	Missense_Mut	SNP	C	G	WU107_BM	115	47.82609	35	0.37142857
UCHL1	4	41270006	41270006	Missense_Mut	SNP	C	A	5925_BM	127	21.25984	4	1

Exome-seq mutation was recovered in RNA-seq (<10 reads)

Exome-seq mutation was recovered in RNA-seq, with high coverage (>= 10 reads)

7.4 References

1. Karlow, J. A., Miao, B., Xing, X., Wang, T. & Zhang, B. Common DNA methylation dynamics in endometriod adenocarcinoma and glioblastoma suggest universal epigenomic alterations in tumorigenesis. *Commun. Biol.* **4**, (2021).
2. Lei, Y. *et al.* Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. *Nat. Commun.* **8**, 1–10 (2017).
3. Morita, S. *et al.* Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat. Biotechnol.* **34**, 1060–1065 (2016).
4. Kron, K. J., Bailey, S. D. & Lupien, M. Enhancer alterations in cancer: A source for a cell identity crisis. *Genome Med.* **6**, 1–12 (2014).
5. Shih, D. J. H. *et al.* Genomic characterization of human brain metastases identifies drivers of metastatic lung adenocarcinoma. *Nat. Genet.* **52**, 371–377 (2020).
6. Aljohani, H. M. *et al.* Genetic mutations associated with lung cancer metastasis to the brain. *Mutagenesis* **33**, 137–145 (2018).
7. Laughney, A. M. *et al. Regenerative lineages and immune-mediated pruning in lung cancer metastasis*. *Nature Medicine* **26**, (Springer US, 2020).
8. Brastianos, P. K. *et al.* Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov.* **5**, 1164–1177 (2015).
9. Song, Z. *et al.* Genomic profiles and tumor immune microenvironment of primary lung carcinoma and brain oligo-metastasis. *Cell Death Dis.* **12**, (2021).
10. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
11. Atlasi, Y. & Stunnenberg, H. G. The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* **18**, 643–658 (2017).
12. Blackledge, N. P. & Klose, R. J. The molecular principles of gene regulation by Polycomb repressive complexes. *Nat. Rev. Mol. Cell Biol.* **0123456789**, (2021).
13. Bhat, K. P., Ümit Kaniskan, H., Jin, J. & Gozani, O. Epigenetics and beyond: targeting writers of protein lysine methylation to treat disease. *Nat. Rev. Drug Discov.* **20**, 265–286 (2021).
14. Rothbart, S. B. & Baylin, S. B. Epigenetic Therapy for Epithelioid Sarcoma. *Cell* **181**, 211 (2020).

Appendix

Appendix A: Chapter 2 extended data

A.1 Tumor suppressor gene DMR overlap analysis

	All GBM HyperDMRs	All GBM HypoDMRs	All EAC HyperDMRs	All EAC HypoDMRs	Shared HyperDMRs	Shared HypoDMRs
Population size (Number of total genes)	25409	25409	26446	26446	25409	25409
Number of successes in the population (number of genes with a DMR within their core promoter)	1745	165	1944	203	515	5
Sample size (number of tumor suppressor genes)	1178	1178	1216	1216	1178	1178
Number of successes in the sample (number of tumor suppressor genes with a DMR within their core promoter)	116	11	142	14	39	0
Hypergeometric P-value	5.27E-05	1.45E-01	2.22E-08	8.61E-02	1.94E-03	2.11E-01

* Note: that only genes with core promoters overlapping 500 base pair regions with: at least 1 CpG, no blacklisted CpGs, MeDIP and/or MRE signal for at least one sample, and excluding the ends of chromosomes were considered. The Y chromosome was not considered for the EAC analysis, and neither the X nor Y chromosome was considered for the GBM analysis. These restrictions were imposed since only 500 base pair regions meeting these criteria have the potential to be called a DMR.

** Blacklisted CpGs were defined as DAC Blacklisted Regions or Duke Excluded Regions generated by the ENCODE project (http://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeMapability)

*** Gene annotations were defined according to RefGene, downloaded from the UCSC Gene Annotation Database (Haeussler M, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* 2019;47(D1):D853-D858. doi: 10.1093/nar/gky1095)

**** Tumor suppressor genes were defined according to TSGene (1216/1217 TSGs had annotations in RefGene)) (Zhao M, Sun J, Zhao Z. TSGene: A web resource for tumor suppressor genes. *Nucleic Acids Res.* 2013;41(D1):D970-976. doi:10.1093/nar/gks937; Zhao M, Kim P, Mitra R, Zhao J, Zhao Z. TSGene 2.0: An updated literature-based knowledgebase for Tumor Suppressor Genes. *Nucleic Acids Res.* 2016;44(D1):D1023-D1031. doi:10.1093/nar/gkv1268)

A.2 Tumor suppressor genes with core promoter hyperDMRs in both EAC and GBM

AKAP12
ALDH1A2
BCL6B
BIN1
CCDC67
CNTNAP2
CXCL12
CXXC5
DKK1
DLK1
DSC3
EGR2
EPB41L3
FOXA2
FOXD3
GATA4
GREM1
GRIN2A
IRF8
IRX1
KL
MIR124-2
MIR1247
MIR23B
MTSS1
NUAK1
PAX6
PCDH8
PRKCB
PRKCDBP
RASSF1
RUNX3
SALL4
SFRP1
SLC39A4
SLC5A8
SOX15
TBX5
THBD
TWIST2
UHRF2
WT1
ZIC1

A.3 Tumor suppressor gene DMR overlap analysis

A.3.1 TSG EAC hyperDMR overlaps

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
ABCG2		0	0
ABI2		0	0
ACHE		0	0
ACVR1C		0	0
ACY1		0	0
ADAMTS18		0	0
ADAMTS8		0	1
ADAMTS9		0	0
ADAMTS9-AS2		1	0
ADARB1		0	0
ADPRH		0	0
AFAP1L2		0	0
AGTR1		1	0
AHCYL1		0	0
AH0K		0	1
AHR		0	0
AHRR		0	1
AIF1		0	0
AIM2		0	0
AIMP1		0	0
AIMP2		0	0
AIP		0	0
AJAP1		1	1
AKAP12		1	0
AKR1B1		0	0
ALDH1A2		1	1
ALOX15		0	0
ALOX15B		1	0
ALPL		0	0
AMER1		0	0
AMH		0	0
A0PC1		0	0
ANGPTL4		0	0
ANP32A		0	0
ANXA1		0	0
ANXA7		0	0
APAF1		0	1
APC		0	0
APITD1		0	1
ARF1		0	1
ARG1		0	1
ARHGAP29		0	0
ARHGEF12		0	0
ARID1A		0	0
ARID2		0	0
ARL11		0	0
ARL6IP5		0	0
ARMC10		0	0
ARMC5		0	0
ARNTL		0	0
ASCL1		1	1
ASS1		0	1
ASXL1		0	0
ATF3		0	0
ATM		0	0
ATMIN		0	0
ATR		0	0
AXIN1		0	0
AXIN2		0	1
AZGP1		0	0
BANP		0	1
BAP1		0	1
BARD1		0	0
BASP1		0	1
BATF2		0	0
BAX		0	0
BBC3		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
BCL10		0	0
BCL2L11		1	1
BCL6B		1	0
BCORL1		0	0
BCR		0	1
BECN1		0	0
BEX2		0	0
BHLHE41		0	0
BIK		0	0
BIN1		1	1
BLCAP		0	0
BLID		0	0
BLM		0	0
BLNK		0	0
BMF		0	0
BMP10		0	0
BMP2		0	0
BMP4		1	0
BMPR1A		0	1
BMPR2		0	0
BNIP3L		0	1
BRCA1		0	0
BRCA2		0	0
BRD7		0	0
BRF1		0	1
BRINP1	DBC1	1	1
BRMS1		0	0
BRSK1		0	0
BTG2		0	1
BTG3		0	0
BTG4		1	0
BTK		0	0
C2orf40		0	0
CABLES1		0	0
CAC02D3		0	0
CADM1		0	0
CADM2		1	0
CADM3		1	0
CADM4		0	1
CAMK2N1		0	1
CAMTA1		0	1
CAPG		0	0
CASC1		0	0
CASC2		0	1
CASP2		0	0
CASP5		0	0
CASP8		0	0
CAT		0	0
CAV1		0	0
CBFA2T3		0	0
CBL		0	0
CBX5		0	0
CCAR1		0	0
CCAR2	DBC1	0	1
CCDC136		0	0
CCDC154		0	0
CCDC67		1	0
CCNC		0	0
CCNDBP1		0	0
CD4		0	0
CD44		0	1
CD82		0	1
CDC73		0	0
CDCP1		0	1
CDH1		0	0
CDH11		0	1
CDH13		1	1

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
CDH17		0	1
CDH4		1	1
CDH5		0	1
CDK2		0	0
CDK2AP1		0	0
CDK6		0	0
CDKN1A		0	0
CDKN1B		0	0
CDKN1C		0	1
CDKN2A		1	1
CDKN2B		0	0
CDKN2C		0	0
CDO1		1	0
CDX2		1	1
CEACAM1		0	0
CEBPA		0	0
CEBPD		0	0
CFTR		1	0
CHD1		0	0
CHD5		0	0
CHEK1		0	0
CHEK2		0	0
CHFR		0	1
CHST10		0	0
CHUK		0	1
CIC		0	0
CITED2		0	0
CIZ1		0	0
CLDN1		0	0
CLDN23		0	1
CLU		0	0
CMTM3		1	0
CMTM5		0	0
CNDP2		0	1
CNN1		0	0
CNOT3		0	0
CNT0P2		1	1
COPS2		0	0
CPNE7		0	0
CREB3L1		0	1
LOC401317	CREB5	0	1
CREBBP		0	0
CREM		0	0
CRNN		0	0
CSF2		0	0
CSMD1		1	0
CSNK1A1		0	0
CSRNP1		0	0
CST5		0	0
CST6		0	0
CTCF		0	0
CTCFL		0	0
CTDSPL		0	0
CTGF		0	0
CTN02		0	1
CTN03		0	0
CTNNBIP1		0	0
CTNND1		0	0
CUL1		0	0
CUL2		0	0
CUL5		0	0
CUX1		0	1
CXCL10		0	0
CXCL12		1	1
CXCL14		1	0
CXCR2		0	0
CXXC4		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
CXXC5		1	0
CYB561D2		0	0
CYB5A		0	0
CYB5R2		0	0
CYGB		0	0
CYLD		0	1
DAB2		0	0
DAB2IP		1	1
DACH1		0	1
DACT1		0	0
DAPK1		0	0
DAPK2		1	0
DAPK3		0	0
DCC		1	1
DCDC2		0	0
DCLRE1A		0	0
DCN		0	0
DCUN1D3		0	0
DDB2		0	0
DDR2		1	1
DDX3X		0	0
DDX58		0	0
DEC1		0	0
DEFB1		0	0
DENND2D		0	0
DFFA		0	0
DF05		0	0
DIABLO		0	0
DICER1		0	1
DIDO1		0	1
DIRAS1		0	0
DIRAS3		0	0
DKK1		1	0
DKK3		0	1
DLC1		0	1
DLEC1		0	0
DLEU1		0	0
DLEU2		0	1
DLG1		0	1
DLK1		1	0
DMBT1		0	0
DMD		0	1
DMTF1		0	0
D0JA3		0	0
D0JB1		0	0
D0JB4		0	0
D0JC11		0	0
DND1		1	0
DNMT1		0	0
DNMT3A		0	1
DNMT3B		0	0
DOK1		1	0
DOK2		0	0
DOK3		0	0
DPH1		0	0
DPP4		0	1
DSC3		1	1
DSP		0	0
DUSP1		0	1
DUSP22		0	0
DUSP26		0	0
DUSP5		0	0
DUSP6		0	0
DUSP9		0	0
E2F1		0	0
E2F2		0	0
E2F3		0	0
EAF1		0	0
EAF2		0	0
EBF3		0	1
ECT2		0	0
EDA2R		0	1
EDNRB		1	0
EED		0	1
EEF1A1		0	0
EEF1E1		0	0
EF05		0	0
EGLN1		0	0
EGLN3		0	0
EGR1		0	1
EGR2		1	0
EHD3		0	0
EHF		0	0
EI24		0	0
EIF3F		0	0
EMP1		0	0
EMP2		0	0
EPAS1		0	1
EPB41		0	0
EPB41L3		1	1
EPHA1		0	0
EPHA2		0	1
EPHA3		0	1
EPHB2		0	0
EPHB3		0	1
EPHB4		0	1
EPHB6		0	0
ERBB4		0	0
ERF		0	0
ERRFI1		0	1
ESR1		0	0
ESR2		0	0
ESRP1		0	0
ESRRB		0	0
ETS2		0	1
ETV6		0	0
EXT1		0	1
EXT2		0	0
EXTL1		0	0
EXTL2		0	0
EXTL3		0	1
EYA4		1	0
EZH1		0	0
EZH2		0	0
FABP3		0	0
FADD		0	0
FAM172A		0	0
FAM188A		0	0
FANCG		0	0
FAS		0	0
FAT1		0	1
FAT4		1	1
FBLN1		0	1
FBP1		0	0
FBXL13		0	0
FBXO25		0	0
FBXO31		0	1
FBXO32		0	0
FBXW7		0	0
FER1L4		0	1
FH		0	0
FHIT		0	0
FHL1		0	0
FLCN		0	0
FL0		0	0
FLT3		0	0
FOXA1		0	0
FOXA2		1	0
FOXC1		0	1
FOXD3		1	0
FOXL2		0	0
FOXO1		0	1
FOXO3		0	1
FOXO4		0	0
FOXO6		0	0
FOXP1		1	1
FOXP3		0	0
FRK		0	0
FRS3		0	0
FUS		0	0
FXN		0	0
FZR1		0	1
G0S2		0	0
GABARAP		0	0
GADD45A		0	1
GADD45B		0	1
GADD45G		0	0
GADD45GIP1		0	0
GALR1		1	0
GA0B		0	0
GAS1		0	1
GAS5		0	0
GATA4		1	1
GATA5		0	0
GBP1		0	0
GGNBP2		0	0
GJA1		0	0
GJB2		0	0
GKN1		0	0
GKN2		0	0
GLI1		0	0
GLIPR1		0	0
GLS2		0	0
GLTSCR1		0	1
GLTSCR2		0	0
G0T1		0	0
GNB2L1		0	0
GNMT		0	0
GORAB		0	0
GPC3		0	0
GPC5		1	0
GPRC5A		0	0
GPX3		0	0
GREM1		1	1
GRIN2A		1	0
GSK3B		0	0
GSN		0	1
GSTP1		0	0
GSTT1		0	0
GTPBP4		0	1
GUCY2C		0	0
H19		0	1
H2AFX		0	0
PTPLAD2		0	0
HACE1		0	0
HBP1		0	0
HCAR2	GPR109A	0	1
HDAC1		0	0
HDAC3		0	0
HECA		0	0
HEPACAM		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
HIC1		1	0
HIF1A		0	0
HINT1		0	0
HIPK2		0	1
HIRA		0	0
HIVEP1		0	0
HIVEP3		0	1
HLTF		0	0
HNF4A		0	0
HOMER2		0	0
HOPX		1	0
HOTS		0	0
HOXB13		0	0
HPGD		0	0
HRASLS2		0	0
HRG		0	0
HSP90B1		0	0
HSPB7		1	0
HSPD1		0	0
HTATIP2		0	1
HTRA1		0	1
HTRA2		0	0
HTRA3		0	1
ID4		0	0
IDH1		0	0
IER3		0	0
IFI16		0	0
IFT88		0	0
IGF1		0	1
IGF2R		0	0
IGFALS		0	0
IGFBP3		0	0
IGFBP4		0	1
IGFBP5		0	1
IGFBP7		0	1
IGFBPL1		0	0
IKZF1		0	0
IKZF2		0	0
IKZF3		0	0
IL17A		0	0
IL17RD		0	0
IL24		0	0
ILK		0	0
ING1		0	0
ING2		0	1
ING3		0	0
ING4		0	0
ING5		0	0
INPP4B		0	0
INTS6		0	0
IQGAP2		0	0
IRF1		0	0
IRF3		0	0
IRF4		1	0
IRF5		0	0
IRF8		1	1
IRX1		1	1
ISG15		0	0
ITGA5		0	1
ITGA7		0	0
ITGAV		0	0
ITGB1		0	0
ITGB3		0	0
JDP2		1	1
JUP		0	0
KANK1		0	1
KAT5		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
KCNRG		0	0
KDM3A		0	0
KDM3B		0	0
KDM5A		0	0
KDM6A		0	0
KDM8	JMJD5	0	1
KIF1B		0	0
KIF7		0	0
KISS1		0	0
KL		1	1
KLF10		0	1
KLF4		0	1
KLF5		0	0
KLF6		0	1
KLK10		0	0
KLK6		0	0
KMT2C		0	0
KRIT1		0	0
KRT19		0	0
L3MBTL4		0	0
LAT2		0	0
LATS1		0	0
LATS2		0	0
LEFTY1		0	0
LEFTY2		1	0
LGALS7		0	0
LHX4		0	0
LHX6		1	0
LIFR		0	0
LIMA1		0	0
LIMD1		0	0
LIN9		0	0
LINC-PINT		0	1
LITAF		0	1
LLGL1		0	1
LMNTD1		0	0
LOX		0	0
LRIG1		0	1
LRIG3		0	0
LRMP		0	0
LRP1B		0	0
LRRC3B		0	0
LRRC4		0	1
LSAMP		1	0
LTF		0	0
LXN		0	0
LZTS1		0	0
MAD1L1		1	1
MADD		0	0
MAL		0	1
MAP2K4		0	0
MAP3K4		0	1
MAP3K8		0	0
MAP4K1		0	0
MAPK10		1	0
MAPK9		0	0
MARCKS		0	0
MARVELD1		0	0
MAT2A		0	0
MAX		0	0
MBD4		0	0
MCC		0	1
MCM9		0	0
MCPH1		0	1
MDC1		0	0
MEG3		1	0
MEN1		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
MFSD2A		0	0
MIA		0	0
MIA2		0	0
MIR1-1		0	0
MIR1-2		0	0
MIR100		0	0
MIR101-1		0	0
MIR101-2		0	0
MIR106A		0	0
MIR107		0	0
MIR10A		0	0
MIR122		0	1
MIR1226		0	0
MIR124-1		1	1
MIR124-2		1	0
MIR124-3		1	1
MIR1247		1	0
MIR125A		1	0
MIR125B1		0	1
MIR125B2		0	0
MIR126		0	0
MIR127		0	0
MIR129-1		0	0
MIR129-2		1	0
MIR1291		0	0
MIR1297		0	0
MIR130A		0	1
MIR132		0	0
MIR133A1		0	0
MIR133A2		0	0
MIR134		0	0
MIR135A1		0	0
MIR135A2		0	0
MIR136		0	0
MIR137		1	0
MIR138-1		0	1
MIR138-2		0	0
MIR140		0	0
MIR141		0	0
MIR142		0	0
MIR143		0	0
MIR145		1	0
MIR146A		0	0
MIR147A		0	0
MIR148A		0	1
MIR148B		0	0
MIR149		0	0
MIR150		0	0
MIR152		0	0
MIR155		0	0
MIR15A		0	0
MIR16-1		0	0
MIR16-2		0	0
MIR17		0	0
MIR181A1		0	0
MIR181A2		0	0
MIR181B1		0	0
MIR181B2		0	0
MIR181C		0	1
MIR182		0	0
MIR183		0	0
MIR185		0	0
MIR186		0	0
MIR187		0	0
MIR18A		0	0
MIR18B		0	0
MIR192		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
MIR193A		0	0
MIR193B		0	0
MIR194-1		0	0
MIR194-2		0	0
MIR195		1	0
MIR196A2		0	0
MIR196B		0	0
MIR198		0	0
MIR199A1		1	0
MIR200A		0	0
MIR200B		0	0
MIR200C		0	0
MIR203A		0	0
MIR204		0	0
MIR205		0	0
MIR206		0	0
MIR20A		0	0
MIR210		0	0
MIR211		0	1
MIR214		0	0
MIR215		0	0
MIR217		0	0
MIR218-1		0	0
MIR218-2		0	0
MIR219A1		0	0
MIR22		0	0
MIR222		0	0
MIR223		0	0
MIR23A		0	1
MIR23B		1	0
MIR24-1		0	0
MIR25		1	0
MIR26A1		0	0
MIR26A2		0	0
MIR26B		1	0
MIR27A		0	0
MIR27B		0	0
MIR28		0	0
MIR296		0	0
MIR29A		0	0
MIR29B1		0	0
MIR29C		0	0
MIR302B		0	0
MIR30A		0	0
MIR30C1		0	0
MIR31		0	0
MIR320A		0	0
MIR326		0	0
MIR329-1		0	0
MIR335		0	0
MIR338		0	0
MIR33A		0	0
MIR340		0	0
MIR34A		0	0
MIR34B		1	0
MIR34C		1	0
MIR367		0	0
MIR370		0	0
MIR375		0	0
MIR378A		1	0
MIR383		0	0
MIR409		0	0
MIR410		0	0
MIR422A		0	1
MIR424		0	0
MIR449A		0	0
MIR449B		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
MIR451A		0	0
MIR483		0	1
MIR486-1		0	0
MIR487B		0	0
MIR490		0	0
MIR493		0	0
MIR494		0	0
MIR495		0	0
MIR497		1	0
MIR502		0	0
MIR503		0	0
MIR504		0	0
MIR505		0	0
MIR508		0	0
MIR509-3		0	0
MIR511		0	0
MIR517A		0	0
MIR519D		0	0
MIR520B		0	0
MIR520C		0	0
MIR551A		0	1
MIR574		0	1
MIR615		0	0
MIR636		0	0
MIR7-1		0	0
MIR7-2		0	0
MIR7-3		0	0
MIR708		0	0
MIR874		0	0
MIR888		0	0
MIR9-1		1	0
MIR9-2		0	0
MIR9-3		1	0
MIR941-1		0	0
MIR98		0	0
MIR99A		0	1
MIRLET7A1		0	0
MIRLET7A2		0	0
MIRLET7A3		1	0
MIRLET7B		1	0
MIRLET7C		0	1
MIRLET7D		0	0
MIRLET7E		1	0
MIRLET7F1		0	0
MIRLET7F2		0	0
MIRLET7G		0	0
MIRLET7I		0	1
MLH1		1	0
MLH3		0	0
MME		1	0
MNT		0	1
MOB1A		0	0
MOB1B		0	0
MRVI1		0	0
MSH2		0	0
MSMB		0	0
MST1		0	0
MST1R		0	0
MT1DP		0	0
MT1F		0	1
MT1G		0	0
MT1M		0	0
MT2A		0	0
MTAP		0	1
MTSS1		1	1
MTUS1		0	0
MUS81		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
MXI1		0	0
MYBBP1A		0	1
MYH9		0	1
MYO18B		0	1
MYO1A		0	0
MZB1		0	0
0PEPLD		0	0
NBN		0	0
NCOA4		0	0
NCOA5		0	0
NDN		0	0
NDRG1		0	1
NDRG2		0	0
NDRG4		0	0
NDST4		0	0
NEDD4		0	0
NEDD4L		0	0
NEURL1	NEURL	0	1
NF1		0	0
NF2		0	0
NFATC2		0	1
NFKB1		0	1
NGFR		0	0
NINJ1		0	1
NIT2		0	0
NKX2-8		0	0
NKX3-1		0	0
NME1		0	0
N0T		1	0
NOL7		0	0
NOTCH1		0	0
NOTCH2		0	0
NOTCH3		0	0
NOV		0	0
NPAS2		0	0
NPM1		0	0
NPRL2		0	0
NR0B2		0	0
NR1I2		0	0
NR2C2		0	0
NR4A1		0	0
NR4A3		0	0
NRBP1		0	0
NRCAM		0	0
NRF1		0	0
NRSN2		0	0
NTRK3		1	0
NUAK1		1	1
NUMB		0	0
NUP98		0	0
NUPR1		0	0
OLFM4		0	0
ONECUT1		1	1
OPCML		1	1
OSCP1		0	0
OSGIN1		0	0
PACRG		0	0
PAEP		0	0
PAFAH1B1		0	0
PAIP2		0	0
PALB2		0	0
PANO1		0	0
PANX2		0	0
PARK2		0	0
PARK7		0	0
PARP1		0	1
PAWR		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
PAX4		0	0
PAX5		0	1
PAX6		1	1
PBRM1		0	0
PCDH10		1	0
PCDH17		1	1
PCDH8		1	0
PCDH9		1	0
PCDHGC3		0	0
PCGF2		0	1
PDCD4		0	0
PDCD5		0	0
PDGFRL		0	1
PDLIM4		0	0
PDS5B		0	0
PDSS2		0	0
PDX1		0	0
PEA15		0	0
PEBP1		0	1
PEG3		1	1
PER2		0	0
PF4		0	0
PFN1		0	0
PGR		0	0
PGRMC2		0	0
PHACTR4		0	0
PHB		0	0
PHC3		0	0
PHF6		0	0
PHLDA2		0	0
PHLDA3		1	1
PHLPP1		0	0
PHLPP2		0	0
PHOX2A		1	1
PIAS1		0	0
PIN1		0	0
PINX1		0	0
PIWIL2		0	0
PKD1		0	0
PKNOX1		0	0
PLA2G16		0	0
PLA2G2A		0	1
PLA2G7		0	0
PLA2R1		0	0
PLAGL1		0	0
PLCB3		0	1
PLCD1		1	1
PLCE1		0	0
PLD1		0	0
PLEKHO1		1	0
PLK1		0	0
PLK2		0	0
PLK5	PLK5P	0	1
PLXNC1		0	0
PML		0	0
PNN		0	0
POU2F3		0	1
POU6F2		0	0
PPARA		0	0
PPARG		0	0
PPM1A		0	1
PPM1L		0	0
PPP1CA		0	0
PPP1R1B		1	0
PPP2CA		0	0
PPP2CB		0	0
PPP2R1B		0	1
PPP2R2C		0	1
PPP2R4		0	1
PPP2R5C		0	0
PPP3CC		0	0
PRDM1		0	1
PRDM11		0	1
PRDM2		0	1
PRDM4		0	0
PRDM5		1	0
PRICKLE1		0	1
PRKAA1		0	0
PRKAA2		0	0
PRKAR1A		1	0
PRKCB		1	1
PRKCD		0	0
PRKCDBP		1	0
PRKCE		0	1
PRODH		0	0
PROX1		0	0
PRR5		0	1
PTCH1		1	1
PTCH2		0	0
PTCSC3		0	0
PTEN		0	1
PTENP1		0	0
PTGDR		1	0
PTPN1		0	1
PTPN11		0	0
PTPN12		0	0
PTPN13		0	0
PTPN2		0	1
PTPN23		0	0
PTPN6		0	0
PTPRC		0	0
PTPRD		0	0
PTPRJ		0	0
PTPRK		0	0
PTPRT		1	0
PWAR4		0	0
PYCARD		0	0
PYHIN1		0	0
RAB25		0	0
RAB7A		0	0
RAD23B		0	1
RAD51C		0	0
RANBP9		0	0
RAP1A		0	0
RAP1GAP		0	1
RARB		0	0
RARRES3		0	0
RASAL1		0	1
RASAL2		0	0
RASL10A		0	1
RASL10B		0	0
RASL11A		0	0
RASSF1		1	0
RASSF10		0	1
RASSF2		0	0
RASSF3		0	0
RASSF4		0	0
RASSF5		0	0
RASSF8		0	0
RB1		0	0
RB1CC1		0	0
RBBP7		0	0
RBBP8		0	0
RBL1		0	0
RBL2		0	0
RBM14		0	0
RBM38		0	1
RBM4		0	0
RBM5		0	0
RBM6		0	0
RBMS3		0	0
RBMX		0	0
RBP1		0	0
RCHY1		0	0
RECK		0	0
RFWD2		0	0
RHOA		0	0
RHOB		0	1
RHOBTB2		0	1
RINT1		0	0
RITA1	C12orf52	0	1
R0SEL		0	0
R0SET2		0	1
RND3		0	0
RNF111		0	0
RNF144A		0	0
RNF180		0	0
RNF8		0	0
RNH1		0	1
ROBO1		1	1
ROR2		0	1
RPA1		0	0
RPL10		0	0
RPL11		0	0
RPL5		0	0
RPRM		1	0
RPS6KA2		0	1
RPS6KA6		1	0
RTN4		0	0
RTN4IP1		0	0
RUNX1		0	1
RUNX2		0	1
RUNX3		1	0
S100A11		0	0
S100A2		0	0
SAA1		0	0
SAFB		0	0
SAFB2		0	0
SALL2		0	0
SALL4		1	0
SAMD9L		0	0
SASH1		0	1
SCGB3A1		0	0
SCRIB		0	0
SCUBE2		0	0
SCYL1		0	0
SDHA		0	0
SDHB		0	0
SDHD		0	0
SEC14L2		0	1
SELENBP1		0	0
SEMA3B		0	0
SEMA3F		0	0
SEPT4		0	0
SERPINB5		0	1
SERPINI2		0	0
SETD2		0	0
SFN		0	0
SFRP1		1	1
SFRP2		1	1
SFRP4		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
SFRP5		1	0
SGMS1		0	0
SH2B3		0	0
SH3GLB1		0	0
SHISA3		0	0
SHPRH		0	0
SHQ1		0	1
SIAH1		0	0
SIK1		0	1
SIRT1		0	0
SIRT2		0	0
SIRT3		0	0
SIRT4		0	0
SIRT6		0	0
SKIL		0	0
SKP2		0	0
SLC39A1		0	0
SLC39A4		1	0
SLC5A8		1	0
SLC9A3R1		0	0
SLIT2		0	0
SLX4		0	0
SMAD2		0	0
SMAD4		0	0
SMARCA2		0	1
SMARCA4		0	1
SMARCB1		0	1
SMARCC1		0	0
SMCHD1		0	0
SMYD4		0	0
SNORD50A		0	0
SOCS1		0	1
SOCS3		0	1
SOD2		0	0
SOX1		1	1
SOX11		1	1
SOX15		1	0
SOX7		1	1
SP100		0	0
SPARC		1	1
SPARCL1		0	1
SPI1		0	0
SPINK7		0	0
SPINT2		0	0
SPOP		0	0
SPRY2		0	0
SPRY4		0	0
SPTBN1		0	1
SRGAP3		0	0
SRPX		0	0
SSBP2		0	0
ST13		0	0
ST20		0	0
ST5		0	1
ST7		0	0
STARD13		0	1
STAT1		0	0
STAT3		0	0
STAT5A		0	1
STK10		0	0
STK11		0	1
STRADA		0	0
STUB1		0	0
SUFU		0	0
SUZ12		0	0
SYK		0	0
SYNM		0	0
SYNPO2		0	0
SYT13		0	0
TAGLN		1	0
TANK		0	0
TAT		0	0
TBL2		0	0
TBRG1		0	0
TBX5		1	1
TCEAL7		0	0
TCEB3		0	0
TCF3		0	0
TCF4		1	0
TCF7L2		0	0
TCHP		0	0
TDGF1		0	0
TDRG1		0	0
TES		0	0
TET2		0	0
TFAP2A		0	1
TFPI2		0	0
TGFB1		0	0
TGFBI		0	0
TGFBR2		0	1
TGFBR3		0	1
TGM3		0	1
THBD		1	0
THBS1		0	0
THRA		0	0
THRB		0	0
THSD1		0	0
THY1		1	1
TIMP3		0	0
TMEFF1		0	0
TMEFF2		1	0
TMEM127		0	0
TMPRSS11A		0	0
TMPRSS6		0	0
TNFAIP3		0	0
TNFAIP8L2		0	0
TNFRSF10A		0	0
TNFRSF10B		0	0
TNFRSF12A		0	0
TNFRSF18		0	0
TNFSF12		0	0
TNFSF9		0	1
TNK1		0	1
TOM1L2		0	0
TOPORS		0	0
TP53		0	0
TP53BP1		0	0
TP53BP2		0	1
TP53COR1		0	0
TP53INP1		0	0
TP63		0	0
TP73		1	1
TPTE2		0	0
TREX2		0	0
TRIM13		0	0
TRIM15		0	0
TRIM24		0	0
TRIM3		0	0
TRIM31		0	0
TRIM32		0	0
TRIM35		0	0
TRIM62		0	1
TRIT1		0	0
TSC1		0	0
TSC2		0	0
TSC22D1		0	1
TSG101		0	0
TSLP		1	0
TSPAN13		0	0
TSPAN32		0	0
TSSC4		0	1
TTC4		0	0
TTF1		0	0
TUSC1		0	0
TUSC2		0	0
TUSC7		0	0
TWIST2		1	1
TXNIP		0	0
UBE2QL1		0	0
UBE4B		0	0
UBIAD1		0	1
UCHL1		0	0
UFL1		0	0
UHRF2		1	1
UIMC1		0	0
UNC5A		0	0
UNC5B		0	0
UNC5C		0	0
UNC5D		1	0
USP12		0	0
USP33		0	0
UVRAG		0	0
VDR		0	0
VEGFA		0	1
VEZT		0	0
VHL		0	0
VIL1		0	0
VIM		0	0
VPS53		0	0
VSNL1		0	0
VTR02-1		0	0
VWA5A		0	0
WDR11		0	0
WDR48		0	0
WFDC1		0	1
WHSC1L1		0	0
WIF1		0	0
WISP3		0	1
WNK2		0	1
WNT11		0	1
WNT5A		0	1
WNT7A		0	1
WT1		1	1
WWOX		0	1
XAF1		0	0
XIST		0	0
XPO5		0	0
XRCC5		0	0
YAP1		0	0
YPEL3		0	0
ZBTB16		1	0
ZBTB18	ZNF238	0	1
ZBTB4		0	1
ZBTB48		0	0
ZBTB7C		0	1
ZC3H10		0	0
ZDHHC2		0	0
ZFAS1		0	0
ZFHX3		0	1
ZFP36		0	0
ZFP36L2		0	1

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
ZFP82		0	0
ZHX2		0	1
ZIC1		1	1
ZMYND10		0	0
ZMYND11		0	0
ZNF185		0	0
ZNF292		0	1
ZNF366		0	0
ZNF382		0	0
ZNF668		0	0
ZYX		0	0
Total		142	262

List of TSGs with EAC hyperDMR in either Promoter and/or Active Enhancer (Aliases Substituted when Needed)

ADAMTS8	CDKN1C	EGR1	HTRA3	MIR181C	PAX6	RBM38	TGM3
ADAMTS9-AS	CDKN2A	EGR2	IGF1	MIR195	PCDH10	RHOB	THBD
AGTR1	CDO1	EPAS1	IGFBP4	MIR199A1	PCDH17	RHOBTB2	THY1
AH0K	CDX2	EPB41L3	IGFBP5	MIR211	PCDH8	RNH1	TMEFF2
AHRR	CFTR	EPHA2	IGFBP7	MIR23A	PCDH9	ROBO1	TNFSF9
AJAP1	CHFR	EPHA3	ING2	MIR23B	PCGF2	ROR2	TNK1
AKAP12	CHUK	EPHB3	IRF4	MIR25	PDGFRL	RPRM	TP53BP2
ALDH1A2	CLDN23	EPHB4	IRF8	MIR26B	PEBP1	RPS6KA2	TP73
ALOX15B	CMTM3	ERRFI1	IRX1	MIR34B	PEG3	RPS6KA6	TRIM62
APAF1	CNDP2	ETS2	ITGA5	MIR34C	PHLDA3	RUNX1	TSC22D1
APITD1	CNT0P2	EXT1	JDP2	MIR378A	PHOX2A	RUNX2	TSLP
ARF1	CREB3L1	EXTL3	JMJD5	MIR422A	PLA2G2A	RUNX3	TSSC4
ARG1	CREB5	EYA4	KANK1	MIR483	PLCB3	SALL4	TWIST2
ASCL1	CSMD1	FAT1	KL	MIR497	PLCD1	SASH1	UBIAD1
ASS1	CTN02	FAT4	KLF10	MIR551A	PLEKHO1	SEC14L2	UHRF2
AXIN2	CUX1	FBLN1	KLF4	MIR574	PLK5P	SERPINB5	UNC5D
BANP	CXCL12	FBXO31	KLF6	MIR9-1	POU2F3	SFRP1	VEGFA
BAP1	CXCL14	FER1L4	LEFTY2	MIR9-3	PPM1A	SFRP2	WFDC1
BASP1	CXXC5	FOXA2	LHX6	MIR99A	PPP1R1B	SFRP5	WISP3
BCL2L11	CYLD	FOXC1	LINC-PINT	MIRLET7A3	PPP2R1B	SHQ1	WNK2
BCL6B	DAB2IP	FOXD3	LITAF	MIRLET7B	PPP2R2C	SIK1	WNT11
BCR	DACH1	FOXO1	LLGL1	MIRLET7C	PPP2R4	SLC39A4	WNT5A
BIN1	DAPK2	FOXO3	LRIG1	MIRLET7E	PRDM1	SLC5A8	WNT7A
BMP4	DBC1	FOXP1	LRRC4	MIRLET7I	PRDM11	SMARCA2	WT1
BMPR1A	DCC	FZR1	LSAMP	MLH1	PRDM2	SMARCA4	WWOX
BNIP3L	DDR2	GADD45A	MAD1L1	MME	PRDM5	SMARCB1	ZBTB16
BRF1	DICER1	GADD45B	MAL	MNT	PRICKLE1	SOCS1	ZBTB4
BTG2	DIDO1	GALR1	MAP3K4	MT1F	PRKAR1A	SOCS3	ZBTB7C
BTG4	DKK1	GAS1	MAPK10	MTAP	PRKCB	SOX1	ZFHX3
C12orf52	DKK3	GATA4	MCC	MTSS1	PRKCDBP	SOX11	ZFP36L2
CADM2	DLC1	GLTSCR1	MCPH1	MYBBP1A	PRKCE	SOX15	ZHX2
CADM3	DLEU2	GPC5	MEG3	MYH9	PRR5	SOX7	ZIC1
CADM4	DLG1	GPR109A	MIR122	MYO18B	PTCH1	SPARC	ZNF238
CAMK2N1	DLK1	GREM1	MIR124-1	N0T	PTEN	SPARCL1	ZNF292F2:F350
CAMTA1	DMD	GRIN2A	MIR124-2	NDRG1	PTGDR	SPTBN1	
CASC2	DND1	GSN	MIR124-3	NEURL	PTPN1	ST5	
CCDC67	DNMT3A	GTPBP4	MIR1247	NFATC2	PTPN2	STARD13	
CD44	DOK1	H19	MIR125A	NFKB1	PTPRT	STAT5A	
CD82	DPP4	HIC1	MIR125B1	NINJ1	R0SET2	STK11	
CDCP1	DSC3	HIPK2	MIR129-2	NTRK3	RAD23B	TAGLN	
CDH11	DUSP1	HIVEP3	MIR130A	NUAK1	RAP1GAP	TBX5	
CDH13	EBF3	HOPX	MIR137	ONECUT1	RASAL1	TCF4	
CDH17	EDA2R	HSPB7	MIR138-1	OPCML	RASL10A	TFAP2A	
CDH4	EDNRB	HTATIP2	MIR145	PARP1	RASSF1	TGFBR2	
CDH5	EED	HTRA1	MIR148A	PAX5	RASSF10	TGFBR3	

Total: 349 (Note: this is 349 as opposed to 350 here because BRINP1 and CCAR2 were replaced with their alias DBC1)

TSGs with expression data (not highlighted in orange above): 310

A.3.2 TSG GBM hyperDMR overlaps

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
ABCG2		0	0
ABI2		0	0
ACHE		0	0
ACVR1C		0	0
ACY1		0	0
ADAMTS18		0	0
ADAMTS8		0	0
ADAMTS9		0	0
ADAMTS9-AS2		0	0
ADARB1		0	1
ADPRH		0	0
AFAP1L2		0	0
AGTR1		0	0
AHCYL1		0	0
AHNAK		0	0
AHR		0	0
AHRR		0	1
AIF1		0	0
AIM2		0	0
AIMP1		0	1
AIMP2		0	0
AIP		0	0
AJAP1		0	1
AKAP12		1	0
AKR1B1		0	0
ALDH1A2		1	0
ALOX15		0	0
ALOX15B		0	0
ALPL		0	1
AMER1		0	0
AMH		0	0
ANAPC1		0	0
ANGPTL4		0	0
ANP32A		0	0
ANXA1		0	0
ANXA7		0	0
APAF1		0	0
APC		0	0
APITD1		0	0
ARF1		0	1
ARG1		0	0
ARHGAP29		0	0
ARHGEF12		0	0
ARID1A		0	0
ARID2		0	0
ARL11		0	1
ARL6IP5		0	0
ARMC10		0	0
ARMC5		0	0
ARNTL		0	0
ASCL1		0	0
ASS1		0	0
ASXL1		0	1
ATF3		0	1
ATM		0	0
ATMIN		0	0
ATR		0	0
AXIN1		0	0
AXIN2		0	1
AZGP1		0	0
BANP		0	0
BAP1		0	0
BARD1		0	0
BASP1		1	1
BATF2		0	0
BAX		0	0
BBC3		0	1

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
BCL10		0	0
BCL2L11		0	1
BCL6B		1	0
BCORL1		0	0
BCR		1	0
BECN1		0	0
BEX2		0	0
BHLHE41		0	1
BIK		0	0
BIN1		1	0
BLCAP		0	0
BLID		0	0
BLM		0	0
BLNK		0	0
BMF		0	0
BMP10		0	0
BMP2		0	0
BMP4		0	0
BMPR1A		0	0
BMPR2		0	0
BNIP3L		0	1
BRCA1		0	0
BRCA2		0	0
BRD7		0	0
BRF1		1	1
BRINP1		0	1
BRMS1		0	0
BRSK1		0	0
BTG2		0	0
BTG3		0	0
BTG4		0	0
BTK		0	0
C2orf40		1	0
CABLES1		0	1
CACNA2D3		0	0
CADM1		0	0
CADM2		0	0
CADM3		0	0
CADM4		0	0
CAMK2N1		0	1
CAMTA1		0	0
CAPG		0	0
CASC1		0	0
CASC2		0	0
CASP2		1	0
CASP5		0	0
CASP8		0	0
CAT		0	0
CAV1		1	0
CBFA2T3		0	0
CBL		0	0
CBX5		0	0
CCAR1		0	0
CCAR2		0	0
CCDC136		0	0
CCDC154		0	0
CCDC67		1	0
CCNC		0	0
CCNDBP1		0	0
CD4		0	0
CD44		0	0
CD82		0	0
CDC73		0	0
CDCP1		1	0
CDH1		0	0
CDH11		0	0
CDH13		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
CDH17		0	1
CDH4		0	0
CDH5		0	0
CDK2		0	0
CDK2AP1		0	0
CDK6		0	0
CDKN1A		0	0
CDKN1B		0	0
CDKN1C		0	1
CDKN2A		0	0
CDKN2B		0	0
CDKN2C		0	0
CDO1		0	0
CDX2		0	0
CEACAM1		0	0
CEBPA		0	0
CEBPD		0	0
CFTR		0	0
CHD1		0	0
CHD5		0	0
CHEK1		0	0
CHEK2		0	0
CHFR		0	0
CHST10		0	1
CHUK		1	0
CIC		0	0
CITED2		0	0
CIZ1		1	0
CLDN1		0	0
CLDN23		1	1
CLU		0	0
CMTM3		0	0
CMTM5		0	0
CNDP2		0	1
CNN1		0	0
CNOT3		0	0
CNTNAP2		1	0
COPS2		0	0
CPNE7		0	0
CREB3L1		0	0
CREB5		0	1
CREBBP		0	1
CREM		1	0
CRNN		0	0
CSF2		0	0
CSMD1		0	1
CSNK1A1		0	0
CSRNP1		0	0
CST5		0	0
CST6		0	0
CTCF		0	0
CTCFL		0	0
CTDSPL		0	0
CTGF		0	1
CTNNA2		0	1
CTNNA3		0	0
CTNNBIP1		0	1
CTNND1		0	0
CUL1		0	1
CUL2		0	0
CUL5		0	0
CUX1		0	1
CXCL10		0	0
CXCL12		1	0
CXCL14		0	0
CXCR2		0	0
CXXC4		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
CXXC5		1	1
CYB561D2		0	0
CYB5A		0	1
CYB5R2		1	0
CYGB		0	0
CYLD		0	0
DAB2		0	0
DAB2IP		0	1
DACH1		0	0
DACT1		0	0
DAPK1		0	0
DAPK2		0	0
DAPK3		0	0
DCC		0	0
DCDC2		1	0
DCLRE1A		0	0
DCN		0	0
DCUN1D3		0	0
DDB2		0	0
DDR2		0	0
DDX3X		0	0
DDX58		0	0
DEC1		0	0
DEFB1		0	0
DENND2D		0	0
DFFA		0	0
DFNA5		1	1
DIABLO		0	1
DICER1		0	1
DIDO1		0	1
DIRAS1		0	1
DIRAS3		0	0
DKK1		1	0
DKK3		0	0
DLC1		1	0
DLEC1		1	0
DLEU1		0	0
DLEU2		0	1
DLG1		0	0
DLK1		1	0
DMBT1		0	0
DMD		0	0
DMTF1		0	0
DNAJA3		0	0
DNAJB1		0	0
DNAJB4		0	0
DNAJC11		0	0
DND1		0	0
DNMT1		0	0
DNMT3A		0	1
DNMT3B		0	0
DOK1		0	0
DOK2		0	0
DOK3		0	0
DPH1		0	0
DPP4		0	0
DSC3		1	1
DSP		1	0
DUSP1		0	1
DUSP22		1	0
DUSP26		0	0
DUSP5		0	0
DUSP6		0	0
DUSP9		0	0
E2F1		1	0
E2F2		0	0
E2F3		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
EAF1		0	0
EAF2		0	0
EBF3		1	1
ECT2		0	0
EDA2R		0	0
EDNRB		0	0
EED		0	0
EEF1A1		0	0
EEF1E1		0	0
EFNA5		0	0
EGLN1		0	0
EGLN3		0	1
EGR1		0	0
EGR2		1	0
EHD3		0	1
EHF		0	0
EI24		0	0
EIF3F		0	0
EMP1		0	0
EMP2		0	0
EPAS1		0	0
EPB41		0	0
EPB41L3		1	0
EPHA1		0	0
EPHA2		0	0
EPHA3		0	1
EPHB2		0	0
EPHB3		0	0
EPHB4		0	0
EPHB6		0	0
ERBB4		0	0
ERF		0	0
ERRFI1		0	0
ESR1		0	0
ESR2		0	0
ESRP1		0	0
ESRRB		0	0
ETS2		0	0
ETV6		0	0
EXT1		0	0
EXT2		0	0
EXTL1		0	0
EXTL2		0	0
EXTL3		0	0
EYA4		0	0
EZH1		0	0
EZH2		0	0
FABP3		0	0
FADD		0	0
FAM172A		0	0
FAM188A		0	0
FANCG		0	0
FAS		0	0
FAT1		0	0
FAT4		0	0
FBLN1		0	0
FBP1		0	0
FBXL13		0	0
FBXO25		0	0
FBXO31		0	0
FBXO32		0	0
FBXW7		0	0
FER1L4		0	0
FH		0	0
FHIT		0	1
FHL1		0	0
FLCN		0	1

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
FLNA		0	0
FLT3		1	0
FOXA1		1	1
FOXA2		1	0
FOXC1		1	1
FOXD3		1	0
FOXL2		0	0
FOXO1		0	0
FOXO3		0	1
FOXO4		0	0
FOXO6		0	1
FOXP1		0	0
FOXP3		0	0
FRK		0	0
FRS3		0	0
FUS		0	0
FXN		0	0
FZR1		0	1
G0S2		0	0
GABARAP		0	0
GADD45A		0	0
GADD45B		0	1
GADD45G		0	0
GADD45GIP1		0	0
GALR1		0	1
GANAB		0	0
GAS1		0	0
GAS5		0	0
GATA4		1	1
GATA5		0	0
GBP1		0	0
GGNBP2		0	0
GJA1		0	0
GJB2		0	0
GKN1		0	0
GKN2		0	0
GLI1		0	0
GLIPR1		0	0
GLS2		1	0
GLTSCR1		0	1
GLTSCR2		0	0
GNAT1		0	0
GNB2L1		0	0
GNMT		0	0
GORAB		0	0
GPC3		0	0
GPC5		0	0
GPRC5A		0	0
GPX3		0	0
GREM1		1	1
GRIN2A		1	0
GSK3B		0	0
GSN		1	0
GSTP1		0	0
GSTT1		0	0
GTPBP4		0	0
GUCY2C		0	0
H19		0	0
H2AFX		0	0
HACD4		0	0
HACE1		0	0
HBP1		0	0
HCAR2		0	1
HDAC1		0	0
HDAC3		0	0
HECA		0	0
HEPACAM		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
HIC1		0	0
HIF1A		0	0
HINT1		0	0
HIPK2		0	0
HIRA		0	0
HIVEP1		0	0
HIVEP3		0	0
HLTF		1	0
HNF4A		0	0
HOMER2		1	1
HOPX		0	0
HOTS		0	0
HOXB13		0	0
HPGD		0	0
HRASLS2		0	0
HRG		0	0
HSP90B1		0	0
HSPB7		0	0
HSPD1		0	0
HTATIP2		1	0
HTRA1		0	0
HTRA2		0	0
HTRA3		0	0
ID4		0	0
IDH1		0	0
IER3		1	0
IFI16		0	0
IFT88		0	0
IGF1		0	0
IGF2R		0	0
IGFALS		0	0
IGFBP3		1	1
IGFBP4		0	0
IGFBP5		0	0
IGFBP7		1	0
IGFBPL1		1	0
IKZF1		1	0
IKZF2		0	0
IKZF3		0	0
IL17A		0	0
IL17RD		0	0
IL24		0	0
ILK		0	0
ING1		0	0
ING2		0	0
ING3		0	0
ING4		0	0
ING5		0	1
INPP4B		0	0
INTS6		0	0
IQGAP2		0	0
IRF1		0	0
IRF3		0	0
IRF4		0	0
IRF5		0	0
IRF8		1	1
IRX1		1	1
ISG15		0	0
ITGA5		0	0
ITGA7		0	0
ITGAV		0	0
ITGB1		0	0
ITGB3		0	0
JDP2		0	1
JUP		0	0
KANK1		0	1
KAT5		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
KCNRG		0	0
KDM3A		0	0
KDM3B		0	0
KDM5A		0	0
KDM6A		0	0
KDM8		0	0
KIF1B		0	1
KIF7		0	0
KISS1		0	0
KL		1	1
KLF10		0	0
KLF4		0	1
KLF5		0	0
KLF6		0	1
KLK10		0	0
KLK6		0	0
KMT2C		0	0
KRIT1		0	0
KRT19		0	0
L3MBTL4		0	0
LAT2		0	0
LATS1		0	0
LATS2		0	1
LEFTY1		0	0
LEFTY2		0	0
LGALS7		0	0
LHX4		0	0
LHX6		0	0
LIFR		0	0
LIMA1		0	0
LIMD1		0	0
LIN9		0	0
LINC-PINT		0	1
LITAF		0	0
LLGL1		0	0
LMNTD1		0	0
LOX		0	0
LRIG1		0	0
LRIG3		0	0
LRMP		0	0
LRP1B		0	0
LRRC3B		0	0
LRRC4		0	1
LSAMP		0	0
LTF		0	0
LXN		0	0
LZTS1		0	0
MAD1L1		0	1
MADD		0	0
MAL		0	1
MAP2K4		0	0
MAP3K4		0	0
MAP3K8		0	0
MAP4K1		0	0
MAPK10		0	1
MAPK9		0	1
MARCKS		0	0
MARVELD1		0	0
MAT2A		0	0
MAX		0	0
MBD4		0	0
MCC		0	0
MCM9		0	0
MCPH1		0	0
MDC1		0	0
MEG3		0	0
MEN1		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
MFSD2A		0	0
MIA		0	0
MIA2		0	0
MIR1-1		0	0
MIR1-2		0	0
MIR100		0	0
MIR101-1		0	0
MIR101-2		0	0
MIR106A		0	0
MIR107		0	0
MIR10A		1	0
MIR122		0	0
MIR1226		0	0
MIR124-1		0	1
MIR124-2		1	0
MIR124-3		0	0
MIR1247		1	0
MIR125A		0	0
MIR125B1		0	0
MIR125B2		0	0
MIR126		0	0
MIR127		0	0
MIR129-1		0	0
MIR129-2		0	0
MIR1291		0	0
MIR1297		0	0
MIR130A		0	0
MIR132		0	0
MIR133A1		0	0
MIR133A2		0	0
MIR134		0	0
MIR135A1		0	0
MIR135A2		0	0
MIR136		0	0
MIR137		0	0
MIR138-1		0	0
MIR138-2		0	0
MIR140		0	0
MIR141		0	0
MIR142		0	0
MIR143		0	0
MIR145		0	0
MIR146A		0	0
MIR147A		0	0
MIR148A		1	1
MIR148B		0	0
MIR149		0	0
MIR150		0	0
MIR152		0	0
MIR155		0	0
MIR15A		0	0
MIR16-1		0	0
MIR16-2		0	0
MIR17		0	0
MIR181A1		0	0
MIR181A2		0	0
MIR181B1		0	0
MIR181B2		0	0
MIR181C		0	0
MIR182		0	0
MIR183		0	0
MIR185		0	0
MIR186		0	0
MIR187		0	0
MIR18A		0	0
MIR18B		0	0
MIR192		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
MIR193A		1	0
MIR193B		0	0
MIR194-1		0	0
MIR194-2		0	0
MIR195		0	0
MIR196A2		0	1
MIR196B		1	1
MIR198		0	0
MIR199A1		0	1
MIR200A		0	0
MIR200B		0	0
MIR200C		0	0
MIR203A		1	0
MIR204		0	0
MIR205		0	0
MIR206		0	0
MIR20A		0	0
MIR210		1	0
MIR211		0	0
MIR214		0	0
MIR215		0	0
MIR217		0	0
MIR218-1		0	0
MIR218-2		0	0
MIR219A1		0	0
MIR22		0	0
MIR222		0	0
MIR223		0	0
MIR23A		0	0
MIR23B		1	0
MIR24-1		1	0
MIR25		0	0
MIR26A1		0	0
MIR26A2		0	0
MIR26B		0	0
MIR27A		0	0
MIR27B		1	0
MIR28		0	0
MIR296		0	0
MIR29A		0	0
MIR29B1		0	0
MIR29C		0	1
MIR302B		0	0
MIR30A		0	0
MIR30C1		0	0
MIR31		0	0
MIR320A		0	0
MIR326		0	0
MIR329-1		0	0
MIR335		0	0
MIR338		1	1
MIR33A		0	0
MIR340		0	1
MIR34A		0	0
MIR34B		0	0
MIR34C		0	0
MIR367		0	0
MIR370		0	0
MIR375		1	0
MIR378A		0	0
MIR383		0	0
MIR409		0	0
MIR410		0	0
MIR422A		0	0
MIR424		0	0
MIR449A		0	0
MIR449B		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
MIR451A		0	0
MIR483		0	0
MIR486-1		0	0
MIR487B		0	0
MIR490		0	0
MIR493		0	0
MIR494		0	0
MIR495		0	0
MIR497		0	0
MIR502		0	0
MIR503		0	0
MIR504		0	0
MIR505		0	0
MIR508		0	0
MIR509-3		0	0
MIR511		0	0
MIR517A		0	0
MIR519D		0	0
MIR520B		0	0
MIR520C		0	0
MIR551A		0	0
MIR574		0	0
MIR615		0	0
MIR636		0	0
MIR7-1		0	0
MIR7-2		0	0
MIR7-3		0	0
MIR708		0	0
MIR874		0	0
MIR888		0	0
MIR9-1		0	0
MIR9-2		0	0
MIR9-3		0	0
MIR941-1		0	1
MIR98		0	0
MIR99A		0	0
MIRLET7A1		0	0
MIRLET7A2		0	0
MIRLET7A3		0	0
MIRLET7B		0	0
MIRLET7C		0	0
MIRLET7D		0	0
MIRLET7E		0	0
MIRLET7F1		0	0
MIRLET7F2		0	0
MIRLET7G		0	0
MIRLET7I		0	0
MLH1		0	1
MLH3		0	0
MME		0	0
MNT		0	1
MOB1A		0	0
MOB1B		0	0
MRVI1		0	0
MSH2		0	0
MSMB		0	0
MST1		0	0
MST1R		0	0
MT1DP		0	0
MT1F		0	0
MT1G		0	0
MT1M		0	0
MT2A		0	0
MTAP		0	0
MTSS1		1	0
MTUS1		0	0
MUS81		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
MXI1		0	0
MYBBP1A		0	1
MYH9		0	1
MYO18B		0	0
MYO1A		0	0
MZB1		0	0
NAPEPLD		0	0
NBN		0	0
NCOA4		0	0
NCOA5		0	1
NDN		0	0
NDRG1		0	1
NDRG2		0	0
NDRG4		0	0
NDST4		0	0
NEDD4		0	0
NEDD4L		1	1
NEURL1		0	0
NF1		0	0
NF2		0	0
NFATC2		0	0
NFKB1		0	0
NGFR		0	0
NINJ1		0	0
NIT2		1	0
NKX2-8		0	0
NKX3-1		0	0
NME1		0	0
NNAT		0	0
NOL7		0	0
NOTCH1		0	0
NOTCH2		0	0
NOTCH3		0	0
NOV		0	0
NPAS2		0	0
NPM1		0	0
NPRL2		0	0
NR0B2		0	0
NR1I2		0	0
NR2C2		0	0
NR4A1		0	0
NR4A3		0	0
NRBP1		0	0
NRCAM		0	0
NRF1		0	0
NRSN2		0	0
NTRK3		0	1
NUAK1		1	1
NUMB		0	0
NUP98		0	0
NUPR1		0	0
OLFM4		0	0
ONECUT1		0	1
OPCML		0	1
OSCP1		0	0
OSGIN1		0	1
PACRG		0	1
PAEP		0	0
PAFAH1B1		0	0
PAIP2		0	0
PALB2		0	0
PANO1		0	0
PANX2		0	0
PARK2		0	0
PARK7		0	0
PARP1		0	0
PAWR		0	1

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
PAX4		0	0
PAX5		0	0
PAX6		1	0
PBRM1		0	0
PCDH10		0	0
PCDH17		0	0
PCDH8		1	0
PCDH9		0	1
PCDHGC3		0	0
PCGF2		0	0
PDCD4		0	0
PDCD5		0	0
PDGFRL		0	0
PDLIM4		0	0
PDS5B		0	0
PDSS2		0	0
PDX1		0	0
PEA15		0	0
PEBP1		0	1
PEG3		0	0
PER2		0	0
PF4		0	0
PFN1		0	0
PGR		1	0
PGRMC2		0	0
PHACTR4		0	0
PHB		0	0
PHC3		0	0
PHF6		0	0
PHLDA2		0	0
PHLDA3		0	1
PHLPP1		0	0
PHLPP2		0	0
PHOX2A		0	0
PIAS1		0	0
PIN1		0	0
PINX1		0	0
PIWIL2		0	0
PKD1		0	0
PKNOX1		0	0
PLA2G16		0	0
PLA2G2A		0	0
PLA2G7		0	0
PLA2R1		0	0
PLAGL1		0	0
PLCB3		0	0
PLCD1		0	0
PLCE1		0	0
PLD1		0	0
PLEKHO1		0	0
PLK1		0	0
PLK2		0	0
PLK5		0	0
PLXNC1		0	1
PML		0	0
PNN		0	0
POU2F3		0	0
POU6F2		0	0
PPARA		0	0
PPARG		1	0
PPM1A		0	0
PPM1L		0	0
PPP1CA		0	0
PPP1R1B		0	0
PPP2CA		0	0
PPP2CB		0	1
PPP2R1B		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
PPP2R2C		0	1
PPP2R4		0	0
PPP2R5C		0	1
PPP3CC		0	0
PRDM1		0	1
PRDM11		0	0
PRDM2		1	1
PRDM4		0	0
PRDM5		0	0
PRICKLE1		1	1
PRKAA1		0	0
PRKAA2		1	1
PRKAR1A		0	0
PRKCB		1	0
PRKCD		0	0
PRKCDBP		1	0
PRKCE		0	1
PRODH		0	0
PROX1		0	0
PRR5		0	0
PTCH1		0	1
PTCH2		1	0
PTCSC3		0	0
PTEN		0	0
PTENP1		0	0
PTGDR		0	0
PTPN1		0	0
PTPN11		0	0
PTPN12		0	0
PTPN13		0	0
PTPN2		0	1
PTPN23		0	0
PTPN6		0	0
PTPRC		0	0
PTPRD		0	0
PTPRJ		0	0
PTPRK		0	1
PTPRT		0	0
PWAR4		0	0
PYCARD		0	0
PYHIN1		0	0
RAB25		0	0
RAB7A		0	0
RAD23B		0	0
RAD51C		0	0
RANBP9		0	0
RAP1A		0	0
RAP1GAP		1	1
RARB		0	0
RARRES3		0	0
RASAL1		0	0
RASAL2		0	0
RASL10A		0	1
RASL10B		0	0
RASL11A		0	0
RASSF1		1	0
RASSF10		1	0
RASSF2		0	0
RASSF3		0	0
RASSF4		0	0
RASSF5		1	1
RASSF8		0	0
RB1		0	1
RB1CC1		0	0
RBBP7		0	0
RBBP8		0	0
RBL1		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
RBL2		0	0
RBM14		0	0
RBM38		1	0
RBM4		0	0
RBM5		0	0
RBM6		0	0
RBMS3		0	0
RBMX		0	0
RBP1		0	0
RCHY1		0	0
RECK		0	1
RFWD2		0	0
RHOA		0	0
RHOB		0	0
RHOBTB2		0	0
RINT1		0	0
RITA1		0	0
RNASEL		0	0
RNASET2		0	0
RND3		0	0
RNF111		0	0
RNF144A		0	0
RNF180		0	0
RNF8		0	0
RNH1		0	1
ROBO1		0	0
ROR2		1	0
RPA1		0	1
RPL10		0	0
RPL11		0	0
RPL5		0	0
RPRM		0	0
RPS6KA2		0	1
RPS6KA6		0	0
RTN4		0	1
RTN4IP1		0	0
RUNX1		0	1
RUNX2		0	1
RUNX3		1	1
S100A11		0	0
S100A2		0	0
SAA1		0	0
SAFB		0	1
SAFB2		0	0
SALL2		0	1
SALL4		1	0
SAMD9L		0	0
SASH1		0	1
SCGB3A1		0	0
SCRIB		0	0
SCUBE2		0	0
SCYL1		0	0
SDHA		0	0
SDHB		0	0
SDHD		0	0
SEC14L2		0	0
SELENBP1		0	0
SEMA3B		0	0
SEMA3F		0	0
SEPT4		0	0
SERPINB5		0	0
SERPINI2		0	0
SETD2		0	0
SFN		0	0
SFRP1		1	0
SFRP2		0	0
SFRP4		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
SFRP5		0	0
SGMS1		0	1
SH2B3		0	0
SH3GLB1		0	0
SHISA3		0	0
SHPRH		0	0
SHQ1		0	0
SIAH1		0	1
SIK1		0	0
SIRT1		0	0
SIRT2		0	0
SIRT3		0	0
SIRT4		0	0
SIRT6		0	0
SKIL		0	0
SKP2		0	0
SLC39A1		0	0
SLC39A4		1	0
SLC5A8		1	0
SLC9A3R1		1	0
SLIT2		0	0
SLX4		0	0
SMAD2		0	0
SMAD4		0	0
SMARCA2		0	1
SMARCA4		0	1
SMARCB1		0	0
SMARCC1		0	1
SMCHD1		0	0
SMYD4		0	0
SNORD50A		0	0
SOCS1		0	0
SOCS3		0	1
SOD2		0	0
SOX1		0	1
SOX11		0	0
SOX15		1	0
SOX7		0	1
SP100		0	0
SPARC		0	0
SPARCL1		0	0
SPI1		0	0
SPINK7		0	0
SPINT2		1	0
SPOP		0	0
SPRY2		0	0
SPRY4		0	0
SPTBN1		1	1
SRGAP3		0	0
SRPX		0	0
SSBP2		0	0
ST13		0	0
ST20		0	0
ST5		0	0
ST7		0	0
STARD13		0	0
STAT1		0	0
STAT3		0	0
STAT5A		0	0
STK10		0	0
STK11		0	1
STRADA		0	0
STUB1		0	0
SUFU		0	0
SUZ12		0	0
SYK		1	1
SYNM		1	0
SYNPO2		0	0
SYT13		0	0
TAGLN		0	0
TANK		0	0
TAT		0	0
TBL2		1	1
TBRG1		0	0
TBX5		1	1
TCEAL7		0	0
TCEB3		0	0
TCF3		0	0
TCF4		0	0
TCF7L2		0	0
TCHP		0	0
TDGF1		0	0
TDRG1		0	0
TES		1	0
TET2		0	0
TFAP2A		0	1
TFPI2		0	0
TGFB1		0	0
TGFBI		0	0
TGFBR2		0	0
TGFBR3		0	0
TGM3		0	0
THBD		1	0
THBS1		0	0
THRA		0	0
THRB		0	0
THSD1		0	0
THY1		0	0
TIMP3		0	1
TMEFF1		0	1
TMEFF2		0	0
TMEM127		0	0
TMPRSS11A		0	0
TMPRSS6		0	0
TNFAIP3		0	0
TNFAIP8L2		0	0
TNFRSF10A		1	1
TNFRSF10B		0	0
TNFRSF12A		0	0
TNFRSF18		0	0
TNFSF12		1	0
TNFSF9		1	1
TNK1		0	0
TOM1L2		0	0
TOPORS		0	0
TP53		0	0
TP53BP1		0	0
TP53BP2		0	1
TP53COR1		0	0
TP53INP1		0	0
TP63		0	0
TP73		0	0
TPTE2		0	0
TREX2		0	0
TRIM13		0	0
TRIM15		0	0
TRIM24		0	0
TRIM3		0	0
TRIM31		0	1
TRIM32		0	1
TRIM35		0	0
TRIM62		0	0
TRIT1		0	0
TSC1		1	0
TSC2		0	1
TSC22D1		0	1
TSG101		0	0
TSLP		0	0
TSPAN13		0	0
TSPAN32		0	0
TSSC4		0	0
TTC4		0	0
TTF1		0	0
TUSC1		0	0
TUSC2		0	0
TUSC7		0	0
TWIST2		1	1
TXNIP		0	0
UBE2QL1		0	0
UBE4B		0	0
UBIAD1		1	0
UCHL1		0	0
UFL1		0	0
UHRF2		1	1
UIMC1		1	0
UNC5A		0	1
UNC5B		0	0
UNC5C		0	0
UNC5D		0	0
USP12		0	0
USP33		0	0
UVRAG		0	0
VDR		0	0
VEGFA		0	0
VEZT		0	0
VHL		1	0
VIL1		0	0
VIM		0	0
VPS53		0	0
VSNL1		0	0
VTRNA2-1		0	0
VWA5A		0	0
WDR11		0	0
WDR48		0	0
WFDC1		1	0
WHSC1L1		0	0
WIF1		0	0
WISP3		0	0
WNK2		1	0
WNT11		0	0
WNT5A		0	0
WNT7A		0	1
WT1		1	1
WWOX		0	0
XAF1		0	0
XIST		0	0
XPO5		0	0
XRCC5		0	0
YAP1		0	0
YPEL3		0	0
ZBTB16		0	1
ZBTB18		0	1
ZBTB4		0	1
ZBTB48		1	0
ZBTB7C		0	0
ZC3H10		0	0
ZDHHC2		0	0
ZFAS1		0	0
ZFHX3		0	1
ZFP36		0	0
ZFP36L2		0	0

TSGs Full List	Substituted Aliases	TSGs with EAC hyperDMR in Promoter	TSGs with EAC hyperDMR in Active Enhancer
ZFP82		0	0
ZHX2		0	0
ZIC1		1	0
ZMYND10		0	0
ZMYND11		0	1
ZNF185		0	0
ZNF292		0	0
ZNF366		0	0
ZNF382		0	0
ZNF668		0	0
ZYX		0	0
Total		116	165

List of TSGs with GBM hyperDMR in either Promoter and/or Active Enhancer (Aliases Substituted when Needed)						
ADARB1	CNDP2	EGR2	KANK1	MYH9	RASL10A	SYK
AHRR	CNTNAP2	EHD3	KIF1B	NCOA5	RASSF1	SYNM
AIMP1	CREB5	EPB41L3	KL	NDRG1	RASSF10	TBL2
AJAP1	CREBBP	EPHA3	KLF4	NEDD4L	RASSF5	TBX5
AKAP12	CREM	FHIT	KLF6	NIT2	RB1	TES
ALDH1A2	CSMD1	FLCN	LATS2	NTRK3	RBM38	TFAP2A
ALPL	CTGF	FLT3	LINC-PINT	NUAK1	RECK	THBD
ARF1	CTNNA2	FOXA1	LRRC4	ONECUT1	RNH1	TIMP3
ARL11	CTNNBIP1	FOXA2	MAD1L1	OPCML	ROR2	TMEFF1
ASXL1	CUL1	FOXC1	MAL	OSGIN1	RPA1	TNFRSF10A
ATF3	CUX1	FOXD3	MAPK10	PACRG	RPS6KA2	TNFSF12
AXIN2	CXCL12	FOXO3	MAPK9	PAWR	RTN4	TNFSF9
BASP1	CXXC5	FOXO6	MIR10A	PAX6	RUNX1	TP53BP2
BBC3	CYB5A	FZR1	MIR124-1	PCDH8	RUNX2	TRIM31
BCL2L11	CYB5R2	GADD45B	MIR124-2	PCDH9	RUNX3	TRIM32
BCL6B	DAB2IP	GALR1	MIR1247	PEBP1	SAFB	TSC1
BCR	DCDC2	GATA4	MIR148A	PGR	SALL2	TSC2
BHLHE41	DFNA5	GLS2	MIR193A	PHLDA3	SALL4	TSC22D1
BIN1	DIABLO	GLTSCR1	MIR196A2	PLXNC1	SASH1	TWIST2
BNIP3L	DICER1	GREM1	MIR196B	PPARG	SFRP1	UBIAD1
BRF1	DIDO1	GRIN2A	MIR199A1	PPP2CB	SGMS1	UHRF2
BRINP1	DIRAS1	GSN	MIR203A	PPP2R2C	SIAH1	UIMC1
C2orf40	DKK1	HCAR2	MIR210	PPP2R5C	SLC39A4	UNC5A
CABLES1	DLC1	HLTF	MIR23B	PRDM1	SLC5A8	VHL
CAMK2N1	DLEC1	HOMER2	MIR24-1	PRDM2	SLC9A3R1	WFDC1
CASP2	DLEU2	HTATIP2	MIR27B	PRICKLE1	SMARCA2	WNK2
CAV1	DLK1	IER3	MIR29C	PRKAA2	SMARCA4	WNT7A
CCDC67	DNMT3A	IGFBP3	MIR338	PRKCB	SMARCC1	WT1
CDCP1	DSC3	IGFBP7	MIR340	PRKCDBP	SOCS3	ZBTB16
CDH17	DSP	IGFBPL1	MIR375	PRKCE	SOX1	ZBTB18
CDKN1C	DUSP1	IKZF1	MIR941-1	PTCH1	SOX15	ZBTB4
CHST10	DUSP22	ING5	MLH1	PTCH2	SOX7	ZBTB48
CHUK	E2F1	IRF8	MNT	PTPN2	SPINT2	ZFHX3
CIZ1	EBF3	IRX1	MTSS1	PTPRK	SPTBN1	ZIC1
CLDN23	EGLN3	JDP2	MYBBP1A	RAP1GAP	STK11	ZMYND11

Total: 245

TSGs with expression data (not highlighted in orange above): 220

A.4 Gene ontology enrichment

A.4.1 EAC hyperDMRs overlapping active enhancers and not promoters GREAT GO Biological Processes significant terms (terms significant in both the binomial and hypergenometric tests (BinomFDRQ<=0.05 and HyperFDRQ<=0.05, were within the top 500 ranked binomial test terms, and had a region fold change of 2 or greater)).

# Ontology	ID	Desc	Binom Rank	Binom P	Binom Bonf P	Binom Fdr Q	Region Fold Enrich	Exp Regions	Obs Regions	Genome Frac	Set Cov	Hyper Rank	Hyper P	Hyper Bonf P	Hyper Fdr Q	Gene Fold Enrich	Exp Genes	Obs Genes	Total Genes	Gene Set Cov	Term Cov	Genes
GO Biological Process	GO:0007179	transforming growth factor beta receptor signaling pathway	163	6.31698E-33	6.59493E-29	4.04597E-31	2.377787	102.1958	243	0.01358626	0.03230524	726	0.000357897		1	0.005146621	1.462058	40.35408	59	126	0.01021115	0.468254	ACVRL1,ARHGEF18,BAMBI,BMPR1A,CCL2,CDK8,CDKN2B,CGN,CITED2,ENG,F11R,FERMT2,FMOD,FUT8,GDF5,HIPK2,KLF10,LEFTY1,LTBP2,LTBP4,MAP3K1,MSTN,MYC,NCOR1,NCOR2,NEDD4L,NODAL,PARD3,PARP1,PDGFA,PDGFB,PDGFRB,PMEPA1,PML,PPM1A,PPP1CA,PPP1CB,PPP1CC,PRKCZ,SKI,SMAD3,SMAD5,SMAD6,SMAD7,STUB1,TAB1,TFDP1,TGFB1,TGFB3,TGFBR1,TGFBR2,TGFBR3,TGFBRAP1,TGIF1,TGIF2,UBE2D1,USP9X,WFIKKN2,WWTR1
GO Biological Process	GO:0045668	negative regulation of osteoblast differentiation	164	7.15242E-33	7.46712E-29	4.55312E-31	3.495145	37.48056	131	0.004982792	0.01741558	794	0.000670615		1	0.008817657	1.89234	10.56893	20	33	0.003461405	0.6060606	AXIN2,CDK6,HAND2,HDAC4,HDAC5,HDAC7,ID2,IGFBP5,LIMD1,LRP5,NOTCH1,PTCH1,SFRP1,SKI,SMAD3,SOX2,SUFU,TNF,TOB1,TWIST2
GO Biological Process	GO:0048844	artery morphogenesis	178	1.503E-30	1.56913E-26	8.81534E-29	2.756929	62.38826	172	0.008294105	0.02286626	289	3.92117E-08	0.0040937	1.41651E-06	2.211672	15.37298	34	48	0.005884389	0.7083333	APOB,CHD7,CITED2,DLL4,ENG,EYA1,FOXC1,FOXC2,FOXF1,FOXS1,HAND2,HES1,HEY1,HOXA1,JAG1,LRP1,MYLK,NOTCH1,PDGFRB,PROX1,PRRX1,PRRX2,RBPJ,SMAD7,SOX4,SRF,STRA6,TBX1,TBX2,TFAP2B,TGFBR1,VEGFA,WNT11,ZMIZ1
GO Biological Process	GO:0060840	artery development	189	7.74626E-30	8.0871E-26	4.27889E-28	2.596285	71.64082	186	0.009524172	0.02472747	240	3.71531E-09	3.87879E-05	1.61616E-07	2.214038	17.61488	39	55	0.00674974	0.7090909	ACVR2B,ACVRL1,APOB,CHD7,CITED2,DLL4,ENG,EYA1,FOXC1,FOXC2,FOXF1,FOXS1,GLI3,HAND2,HES1,HEY1,HOXA1,JAG1,LRP1,MYLK,NKX3-1,NOTCH1,PDGFRB,PROX1,PRRX1,PRRX2,RBPJ,SHH,SMAD7,SOX4,SRF,STRA6,TBX1,TBX2,TFAP2B,TGFBR1,VEGFA,WNT11,ZMIZ1
GO Biological Process	GO:0006909	phagocytosis	193	1.90095E-29	1.9846E-25	1.02829E-27	2.315894	98.45011	228	0.01308829	0.03031109	566	4.93019E-05	0.5147117	0.000909385	1.505023	44.5176	67	139	0.01159571	0.4820144	ABCA1,ABCA7,ABL1,ACTB,ACTG1,ACTR2,ADORA1,ADORA2A,ALOX15,ANXA11,ARPC5,BAIAP2,CD247,CD93,CDC42,COLEC12,CORO1C,DOCK1,FCER1G,FCGR2A,FCGR3A,FCN2,FYN,GAS6,GATA2,GRB2,HSP90AA1,HSP90AB1,IGLL5,IRF8,ITPR1,ITPR2,ITPR3,LBP,LIMK1,LRP1,LYN,MAPK1,MAPK3,MARCO,MEGF10,MYH2,MYO10,MYO7A,NCKIPSD,PEAR1,PIK3R1,PIP5K1C,PLCG1,PLCG2,PPAPDC1A,PRKCE,PTK2,RAB20,RAB7A,RARA,SCARB1,SRC,SYK,TGM2,TUB,TULP1,UNC13D,VAV2,VAV3,WASF2,WIPF3
GO Biological Process	GO:2001243	negative regulation of intrinsic apoptotic signaling pathway	201	4.80588E-29	5.01734E-25	2.49619E-27	3.669966	29.42806	108	0.003912266	0.01435788	890	0.00141072		1	0.01654822	1.696935	14.73244	25	46	0.004326757	0.5434783	ACKR3,AKT1,ARHGEF2,BCL2,CD44,CLU,CXCL12,DDX3X,EPO,ING2,KDM1A,KIAA1967,MAPK7,NDUFA13,NOC2L,PPIF,SFRP2,SNAI1,SNAI2,SOD2,SRC,TMEM161A,TRIM32,URI1,WFS1
GO Biological Process	GO:0071560	cellular response to transforming growth factor beta stimulus	206	9.54188E-29	9.96173E-25	4.83579E-27	2.017216	150.207	303	0.01996903	0.04028184	469	9.87097E-06	0.1030529	0.000219729	1.52115	49.9622	76	156	0.01315334	0.4871795	ACVRL1,ARG1,ARHGEF18,BAMBI,BMPR1A,CCL2,CDK8,CDKN2B,CGN,CITED2,CLEC3B,COL1A1,COL4A2,CYP11A1,ENG,F11R,FERMT2,FMOD,FUT8,GDF5,HIPK2,KLF10,LEFTY1,LTBP2,LTBP4,MAP3K1,MAPK7,MEF2C,MSTN,MYC,NCOR1,NCOR2,NEDD4L,NODAL,PARD3,PARP1,PDE2A,PDGFA,PDGFB,PDGFRB,PMEPA1,PML,PPM1A,PPP1CA,PPP1CB,PPP1CC,PRKCZ,SFRP1,SKI,SMAD3,SMAD5,SMAD6,SMAD7,SOX6,SOX9,STUB1,TAB1,TFDP1,TGFB1,TGFB3,TGFBR1,TGFBR2,TGFBR3,TGFBRAP1,TGIF1,TGIF2,UBE2D1,USP9X,WFIKKN2,WNT4,WNTSA,WNT7A,WWOX,WWTR1,XCL1,ZEB1

# Ontology	ID	Desc	Binom Rank	Binom P	Binom Bonf P	Binom Fdr Q	Region Fold Enrich	Exp Regions	Obs Regions	Genome Frac	Set Cov	Hyper Rank	Hyper P	Hyper Bonf P	Hyper Fdr Q	Gene Fold Enrich	Exp Genes	Obs Genes	Total Genes	Gene Set Cov	Term Cov	Genes
GO Biological Process	GO:0043467	regulation of generation of precursor metabolites and energy	209	1.25462E-28	1.30982E-24	6.26708E-27	2.992348	47.12019	141	0.006264316	0.01874501	1004	0.002822299	1	0.02934741	1.56118	19.21623	30	60	0.005192108	0.5	ADRA1B,AKT1,AKT2,C1QTNF2,CISD1,DDIT4,GCK,GFPT1,GRB10,HDAC4,IER3,IGF1,IGF2,INSR,IRS1,IRS2,MNF1,NCOR1,NOA1,NOS2,PPARA,PPIF,PPP1CA,PPP1CB,PPP1CC,PPP1R3B,PRDM16,PRKAG2,SORBS1,TRAP1
GO Biological Process	GO:0043647	inositol phosphate metabolic process	219	9.67101E-28	1.00965E-23	4.61029E-26	3.173913	39.38357	125	0.005235784	0.01661792	988	0.002754063	1	0.02910164	1.589565	17.61488	28	55	0.004845967	0.5090909	CALM1,CALM2,CALM3,INPP4A,INPP5A,INPP5J,INPPL1,IP6K2,IP6K3,IPPK,ISYNA1,ITPK1,ITPKA,ITPKB,MTMR2,MTMR7,NUDT10,NUDT11,NUDT4,PLCB2,PLCB3,PLCD1,PLCD3,PLCG1,PLCG2,PLCH2,PLD4,PTH2
GO Biological Process	GO:0007219	Notch signaling pathway	221	1.86134E-27	1.94324E-23	8.79294E-26	2.167464	112.574	244	0.01496596	0.0323818	548	0.000042427	0.4429379	0.000808281	1.548278	38.75273	60	121	0.01038422	0.4958678	AGO2,APH1B,APP,ARRB1,ASCL1,CDK6,CDK8,CHAC1,CREB1,CREBBP,DLK1,DLL1,DLL4,DTX1,EP300,EPN1,EPN2,FBXW7,FOXC1,FOXC2,FOXS1,GMDS,HDAC5,HDAC7,HES1,HEY1,HEYL,HOXD3,IL2RA,ITGB1BP1,JAG1,LFNG,MAML2,MAML3,MESP2,MIB1,MYC,NCOR1,NCOR2,NEURL,NEURL1B,NOTCH1,NOTCH3,NOTCH4,PSEN1,PSEN2,RBPJ,RBX1,SNAI1,SNAI2,SPEN,TBL1X,TBL1XR1,TLE1,TLE2,TLE3,TMEM100,TNRC6C,TP63,ZNF423
GO Biological Process	GO:0001570	vasculogenesis	235	4.37598E-26	4.56852E-22	1.94405E-24	2.308483	87.50334	202	0.01163299	0.02685456	357	4.29591E-07	0.004484926	1.25628E-05	1.928517	21.77839	42	68	0.007268951	0.6176471	ACKR3,AGGF1,CITED2,CTNNB1,EGFL7,ENG,EPHA2,FBXW7,FOXF1,FOXM1,GJC1,HDAC7,HEG1,HEY1,HHEX,MYO18B,MYO1E,NKX2-5,NOTCH1,PAXIP1,PDGFRB,PITX2,PTK2,PTPRJ,QKI,SHH,SMARCA4,SMO,SOX17,T,TBX5,TDGF1,TEAD2,TGFBR2,TMEM100,TNNI3,VEGFA,WNT7A,WNT7B,WT1,ZFP36L1,ZMIZ1
GO Biological Process	GO:0043627	response to estrogen stimulus	244	2.31323E-25	2.41501E-21	9.89757E-24	2.04492	125.6773	257	0.01670796	0.03416645	316	8.97449E-08	0.000936937	2.96499E-06	1.688388	43.23652	73	135	0.01263413	0.5407407	ADCYAP1R1,ALDH1A2,ARNT2,ARSB,ASH2L,BDH1,BGLAP,BID,CCND1,CITED2,CITED4,CRIPAK,CTGF,CTNNA1,CTNNB1,CYP11A1,DTYMK,DUSP1,EGFR,EP300,EPO,ETS1,F7,FOXA1,GAL,GATA3,GATA4,GATA6,GHRHR,GHRL,GPER,HMOX1,HSP90AA1,HTR5A,IGF2,IGFBP2,IL10,IL4R,MAPK1,MMP14,MSTN,MSX2,MYOD1,MYOG,NCOR2,NR2F2,PDGFA,PDGFB,PDGFRB,PPARG,PRL,PTK2,PTPRN,RAMP3,RARA,RGS9,SERPINA1,SERPINB9,SFRP1,SLC6A1,SMAD6,SOCS2,SSTR1,STAT3,TBL1X,TGFB1,TGFB3,TGFBR2,TNFRSF11B,TNFRSF1A,TXNIP,WFDC1,WNT7A
GO Biological Process	GO:0007009	plasma membrane organization	248	5.49366E-25	5.73538E-21	2.31265E-23	2.383215	75.94784	181	0.01009676	0.02406275	973	0.002483053	1	0.02664242	1.416627	34.58921	49	108	0.008480443	0.4537037	A4GALT,AKT1,AKT2,ANK1,ANK2,ANK3,ANO6,BBS1,BBS2,CDH1,COL5A1,CRB3,DLG1,DYSF,EHD2,FA2H,FNBP1L,GAS6,GOLPH3,KRT18,LAMA5,MICALL1,MYH10,MYH2,MYOF,MYRF,NDRG1,NFASC,NKD2,PACS2,PACSIN1,PACSIN2,PI2K1,PLEKHF1,PLSCR1,PLSCR3,PRKCZ,PTCH1,RAB31,RAB8A,RAMP1,RAMP3,SPTB,SPTBN1,SYT7,TESC,VAMP3,WNT4,ZDHHC22
GO Biological Process	GO:0043433	negative regulation of sequence-specific DNA binding transcription factor activity	252	1.00859E-24	1.05297E-20	4.17844E-23	2.1188	108.552	230	0.01443127	0.03057697	760	0.000494108	1	0.006787486	1.467776	37.47165	55	117	0.009518865	0.4700855	ARRB1,BHLHE40,BMP7,BTRC,CACTIN,CDKN2A,CEBPG,COMMD7,CTNNBIP1,CYLD,DAB2IP,DAP,FAM105B,FOXA2,FOXJ1,FOXS1,GAS6,GLIS2,HAND1,HAND2,HDAC4,HEYL,HMOX1,JDP2,KDM1A,KLF4,MSX2,NFKBIA,NLRC5,NLRP3,NOD2,PBX1,PEX14,PIAS4,PIM1,PRMT2,PRNP,PROX1,PTCH1,PTGIS,RBCK1,SIGIRR,SIK1,SIVA1,SMAD7,SUFU,SUMO1,TAX1BP1,TCF7L2,TNFAIP3,TRIB1,USP7,WFS1,WWP2,XCL1
GO Biological Process	GO:2001234	negative regulation of apoptotic signaling pathway	262	3.82851E-24	3.99697E-20	1.52556E-22	2.372973	74.16855	176	0.009860217	0.02339803	703	0.000293924	1	0.004364954	1.544747	30.4257	47	95	0.008134303	0.4947368	ACKR3,AKT1,AR,ARHGEF2,BCL2,CD44,CLU,CSF2,CXCL12,DDX3X,EPO,FYN,GDNF,HMOX1,HTT,IGF1,IL1B,ING2,ITGA6,KDM1A,KIAA1967,MAPK7,MCL1,NDUFA13,NOC2L,NOS3,NRG1,PARK7,PPIF,RNF34,SCG2,SFRP2,SH3RF1,SNAI1,SNAI2,SOD2,SRC,TCF7L2,TGFBR1,TMEM161A,TNF,TNFAIP3,TRIM32,UNC5B,URI1,WFS1,ZMYND11
GO Biological Process	GO:0071230	cellular response to amino acid stimulus	263	3.97601E-24	4.15096E-20	1.57831E-22	3.001423	38.98151	117	0.005182333	0.01555437	728	0.000363751	1	0.005216425	1.815326	13.77163	25	43	0.004326757	0.5813953	CAPN2,CDH1,CEBPB,COL16A1,COL1A1,COL4A1,COL6A1,CPEB3,DNMT1,DNMT3A,IPO5,LAMTOR1,LAMTOR3,LAMTOR4,MMP3,NEURL,NSMF,PDGFRA,RPTOR,RRAGC,SH3BP4,SOCS1,TNF,UBR2,ZEB1
GO Biological Process	GO:0035791	platelet-derived growth factor receptor-beta signaling pathway	292	1.78582E-21	1.8644E-17	6.38493E-20	8.457342	4.256656	36	0.000565894	0.004785961	854	0.001077296	1	0.01316975	3.122361	1.921623	6	6	0.001038422	1	CSF1R,PDGFA,PDGFB,PDGFRB,PDGFRL,PTPN1
GO Biological Process	GO:0071229	cellular response to acid	294	1.81829E-21	1.8983E-17	6.45679E-20	2.758758	42.77288	118	0.00568637	0.01568732	919	0.001809707	1	0.02055859	1.656763	15.69325	26	49	0.004499827	0.5306122	CAPN2,CDH1,CEBPB,COL16A1,COL1A1,COL4A1,COL6A1,CPEB3,DNMT1,DNMT3A,IPO5,LAMTOR1,LAMTOR3,LAMTOR4,MMP3,NEURL,NSMF,PDGFRA,RPTOR,RRAGC,S100A10,SH3BP4,SOCS1,TNF,UBR2,ZEB1
GO Biological Process	GO:0032355	response to estradiol stimulus	317	1.49674E-20	1.56259E-16	4.92931E-19	2.208377	77.88526	172	0.01035433	0.02286626	862	0.001140161	1	0.01380891	1.540905	24.66083	38	77	0.00657667	0.4935065	ADCYAP1R1,ALDH1A2,ARNT2,BDH1,BID,CTGF,CTNNB1,DUSP1,EGFR,ETS1,FOXA1,GPER,HTR5A,IGF2,IGFBP2,IL10,MSX2,MYOD1,MYOG,NCOR2,NR2F2,PDGFA,PDGFB,PDGFRB,PRL,PTK2,RAMP3,RARA,SERPINA1,SFRP1,SLC6A1,SOCS2,SSTR1,STAT3,TGFB1,TNFRSF1A,TXNIP,WFDC1
GO Biological Process	GO:0060323	head morphogenesis	318	1.69104E-20	1.76544E-16	5.55171E-19	2.777436	39.96492	111	0.005313072	0.01475671	863	0.001140746	1	0.01379999	1.836683	10.8892	20	34	0.003461405	0.5882353	ANKRD11,ARID5B,COL1A1,DKK1,LEF1,LRP6,MSX1,PAX9,PDGFRA,PLEKHA1,SKI,SSBP3,STRA6,TBX1,TCF7L2,TFAP2A,TGFB1,TGFB3,TWIST2,WNT3
GO Biological Process	GO:0033235	positive regulation of protein sumoylation	319	1.93971E-20	2.02506E-16	6.34815E-19	6.248453	6.881703	43	0.000914877	0.005716565	974	0.002547331	1	0.02730404	2.497889	3.202705	8	10	0.001384562	0.8	AHRR,CDKN2A,HDAC4,PIAS1,PIAS3,PIAS4,RASD2,TOLLIP
GO Biological Process	GO:0043393	regulation of protein binding	342	5.3631E-19	5.59907E-15	1.63716E-17	2.110466	82.92008	175	0.01102368	0.02326509	490	1.46064E-05	0.1524911	0.000311206	1.619002	34.58921	56	108	0.009691935	0.5185185	ADD1,AES,APP,ARRB1,AURKB,BAMBI,BCL2,CAPRIN2,CARM1,CDC42,CDK5,CDON,CREBBP,CSF3,CTNNBIP1,DVL1,EP300,GTPBP4,HDAC4,HDAC5,HIPK2,HSP90AB1,SL1,ITGB1BP1,KDM1A,LDLRAP1,LEF1,LFNG,LRPAP1,MAPK3,MAPK8,MEF2A,MEF2C,NES,NKX3-1,NRG1,NRP1,PARK7,PEX14,PKD1,PPARA,PRKACA,PRMT7,PRMT8,PSEN1,PSEN2,RAPGEF2,RFNG,SENP2,SPAG9,SRC,TCF7L2,TGFBR1,TGFBR3,VTN,WNT3A

# Ontology	ID	Desc	Binom Rank	Binom P	Binom Bonf P	Binom Fdr Q	Region Fold Enrich	Exp Regions	Obs Regions	Genome Frac	Set Cov	Hyper Rank	Hyper P	Hyper Bonf P	Hyper Fdr Q	Gene Fold Enrich	Exp Genes	Obs Genes	Total Genes	Gene Set Cov	Term Cov	Genes
GO Biological Process	GO:0010171	body morphogenesis	344	6.95186E-19	7.25774E-15	2.10981E-17	2.490628	49.38513	123	0.006565425	0.01635203	853	0.001063692	1	0.01301869	1.742713	13.77163	24	43	0.004153686	0.5581395	ANKRD11,ARID5B,CDON,COL1A1,DKK1,FUZ,GATAD2A,LEF1,LRP6,MSX1,PAX9,PDGFRA,PLEKHA1,SKI,SSBP3,STRA6,TBX1,TCF7L2,TFAP2A,TGFB1,TGFB3,TWIST2,WNT3,ZNF281
GO Biological Process	GO:0045834	positive regulation of lipid metabolic process	345	7.83894E-19	8.18385E-15	2.37213E-17	2.091828	84.61499	177	0.011249	0.02353098	600	9.91026E-05	1	0.001724385	1.57695	31.70678	50	99	0.008653513	0.5050505	ABCG1,ABHD5,AKT1,AKT2,BMP6,C1QTNF2,CCL19,CCL21,CCR7,CD81,CDC42,CNEP1R1,CPT1A,CREB1,CYR61,EPHA8,FGF1,FGFR3,FLT1,GHSR,HTR2A,HTR2B,IGF1R,IGF2,IL1B,IRS1,IRS2,KIT,LDLR,LDLRAP1,LYN,MAPK9,MID1IP1,NOD2,PDGFB,PDGFRA,PDGFRB,POR,PPARA,PPARG,PRKCE,PTK2,SCARB1,SORBS1,SREBF1,TGFB1,TNF,VAV2,VAV3,WNT4
GO Biological Process	GO:0002431	Fc receptor mediated stimulatory signaling pathway	376	1.15912E-17	1.21013E-13	3.21842E-16	2.340987	55.10496	129	0.007325839	0.01714969	587	8.17355E-05	0.8533181	0.001453694	1.691279	23.05948	39	72	0.00674974	0.5416667	ABL1,ACTB,ACTG1,ACTR2,ARPC5,BAIAP2,CD247,CDC42,DOCK1,FCER1G,FCGR2A,FCGR3A,FYN,GRB2,HSP90AA1,HSP90AB1,IGLL5,ITPR1,ITPR2,ITPR3,LIMK1,LYN,MAPK1,MAPK3,MYH2,MYO10,NCKIPSD,PIK3R1,PLCG1,PLCG2,PPAPDC1A,PRKCE,PTK2,SRC,SYK,VAV2,VAV3,WASF2,WIPF3
GO Biological Process	GO:0046890	regulation of lipid biosynthetic process	382	1.4713E-17	1.53603E-13	4.02103E-16	2.036241	85.94267	175	0.01142551	0.02326509	956	0.002260794	1	0.02468901	1.427365	33.6284	48	105	0.008307373	0.4571429	ABCG1,ACADVL,AKR1C3,AKT1,APOB,ATP1A1,BMP6,CNEP1R1,CREB1,CYR61,DHCR7,DKK3,FBXW7,FGF1,FGFR4,GPER,HTR2A,HTR2B,IGF1R,IGF2,IL1B,INSIG1,INSIG2,LDLR,LPCAT1,MAPK9,MID1IP1,NFKB1,NR3C1,NR5A1,PDGFA,PDGFB,POR,PRKAG2,PROX1,SCARB1,SEC14L2,SIK1,SLC27A1,SNAI1,SNAI2,SORBS1,SREBF1,STK11,TNF,TSPO,VDR,WNT4
GO Biological Process	GO:0002076	osteoblast development	383	1.51671E-17	1.58344E-13	4.13431E-16	3.151885	24.42982	77	0.003247783	0.01023664	542	3.53765E-05	0.3693303	0.000681421	2.465022	6.085139	15	19	0.002596054	0.7894737	BGLAP,GLI2,HDAC4,HDAC5,HEY1,HOXA2,JUND,LIMD1,LRP5,MSX2,PTH1R,RUNX2,SATB2,SHH,SMAD3
GO Biological Process	GO:0048546	digestive tract morphogenesis	384	1.53393E-17	1.60142E-13	4.17036E-16	2.064639	81.85452	169	0.01088202	0.02246743	291	4.04282E-08	0.00042207	1.45041E-06	2.139395	17.29461	37	54	0.0064036	0.6851852	ACVR2B,BCL2,CTNNB1,EGFR,EPHB3,FGFR2,FGFR3,FOXF1,FOXP4,GATA4,GLI2,GLI3,HLX,HNF1B,JD2,NKX2-3,NODAL,NOTCH1,PDGFRA,PITX2,SFRP1,SFRP2,SHH,SHOX2,SMAD3,SMO,SOX10,SOX11,SOX17,STRA6,TCF21,TCF7L2,TP63,TP73,VANGL2,WNT11,WNT5A
GO Biological Process	GO:0038096	Fc-gamma receptor signaling pathway involved in phagocytosis	391	2.59514E-17	2.70933E-13	6.92922E-16	2.324719	55.06041	128	0.007319916	0.01701675	634	0.00014066	1	0.002316238	1.671123	22.73921	38	71	0.00657667	0.5352113	ABL1,ACTB,ACTG1,ACTR2,ARPC5,BAIAP2,CD247,CDC42,DOCK1,FCGR2A,FCGR3A,FYN,GRB2,HSP90AA1,HSP90AB1,IGLL5,ITPR1,ITPR2,ITPR3,LIMK1,LYN,MAPK1,MAPK3,MYH2,MYO10,NCKIPSD,PIK3R1,PLCG1,PLCG2,PPAPDC1A,PRKCE,PTK2,SRC,SYK,VAV2,VAV3,WASF2,WIPF3
GO Biological Process	GO:0035307	positive regulation of protein dephosphorylation	394	2.79561E-17	2.91862E-13	7.40767E-16	4.758466	9.66698	46	0.001285161	0.006115395	877	0.001239874	1	0.01475973	2.289731	4.804057	11	15	0.001903773	0.7333333	ADORA1,CALM1,CALM2,CALM3,CNEP1R1,DLC1,GBA,NSMF,PPP2R4,PPP2R5A,TGFB1
GO Biological Process	GO:0038094	Fc-gamma receptor signaling pathway	395	2.83766E-17	2.96252E-13	7.50004E-16	2.321921	55.12677	128	0.007328739	0.01701675	674	0.000206518	1	0.00319888	1.647913	23.05948	38	72	0.00657667	0.5277778	ABL1,ACTB,ACTG1,ACTR2,ARPC5,BAIAP2,CD247,CDC42,DOCK1,FCGR2A,FCGR3A,FYN,GRB2,HSP90AA1,HSP90AB1,IGLL5,ITPR1,ITPR2,ITPR3,LIMK1,LYN,MAPK1,MAPK3,MYH2,MYO10,NCKIPSD,PIK3R1,PLCG1,PLCG2,PPAPDC1A,PRKCE,PTK2,SRC,SYK,VAV2,VAV3,WASF2,WIPF3
GO Biological Process	GO:0060322	head development	401	4.94924E-17	5.16701E-13	1.28853E-15	2.209572	63.36068	140	0.008423382	0.01861207	714	0.000337882	1	0.004940462	1.741317	16.65407	29	52	0.005019038	0.5576923	ALDH1A2,ALDH1A3,ANKRD11,ARID5B,CHD7,COL1A1,DKK1,GLI2,LEF1,LHX1,LRP6,MSX1,PAX9,PDGFRA,PGAP1,PLEKHA1,SKI,SOX3,SSBP3,STRA6,TBX1,TCF7L2,TFAP2A,TGFB1,TGFB3,TWIST2,WNT3,WNT5A,ZFAND5
GO Biological Process	GO:0035306	positive regulation of dephosphorylation	414	8.11783E-17	8.47502E-13	2.04711E-15	4.621489	9.953501	46	0.001323252	0.006115395	1001	0.002818652	1	0.02939733	2.146623	5.124328	11	16	0.001903773	0.6875	ADORA1,CALM1,CALM2,CALM3,CNEP1R1,DLC1,GBA,NSMF,PPP2R4,PPP2R5A,TGFB1
GO Biological Process	GO:0030512	negative regulation of transforming growth factor beta receptor signaling pathway	416	8.79095E-17	9.17775E-13	2.20619E-15	2.350005	51.48925	121	0.006845154	0.01608615	1068	0.003850488	1	0.0376396	1.513872	21.13785	32	66	0.005538249	0.4848485	ADAMTSL2,BAMBI,CHST11,CIDEA,ENG,FBN1,HTRA1,HTRA3,HTRA4,NEDD4L,NKX2-1,ONECUT1,ONECUT2,PIN1,PMEPA1,PPM1A,PPP1CA,PPP1CB,PPP1CC,PRDM16,SKI,SKOR1,SMAD3,SMAD6,SMAD7,STUB1,TDGF1,TGFB1,TGFB3,TGFBR1,TGFBR2,TGFBR3
GO Biological Process	GO:0007566	embryo implantation	429	2.13298E-16	2.22683E-12	5.19074E-15	2.947819	26.79947	79	0.003562812	0.01050253	1017	0.00294423	1	0.03022395	1.734645	11.52974	20	36	0.003461405	0.5555556	BSG,CALCA,DDR1,EPO,FBLN1,FKBP4,HMX3,IL1B,LIF,MMP9,NLRP5,PCSK5,PPARD,RPL29,RXRA,SCGB1A1,STC2,TEAD4,TGFBR2,UBTF
GO Biological Process	GO:0060325	face morphogenesis	439	5.37021E-16	5.6065E-12	1.27711E-14	2.565738	37.80589	97	0.005026043	0.01289551	895	0.0014637	1	0.01707377	1.873416	9.608115	18	30	0.003115265	0.6	ANKRD11,ARID5B,COL1A1,DKK1,LEF1,LRP6,MSX1,PAX9,PDGFRA,PLEKHA1,SKI,STRA6,TBX1,TCF7L2,TFAP2A,TGFB1,TGFB3,TWIST2
GO Biological Process	GO:0035924	cellular response to vascular endothelial growth factor stimulus	444	7.10874E-16	7.42152E-12	1.67151E-14	2.904358	26.8562	78	0.003570353	0.01036958	612	0.000109401	1	0.001866252	2.211672	7.686492	17	24	0.002942195	0.7083333	DAB2IP,DLL4,EGR3,FLT1,FLT4,GAS1,ITGB1BP1,MAPKAPK2,MT1G,NOTCH1,NR4A1,NRP1,NRP2,PDGFRA,PDGFRB,PRKD2,VEGFA
GO Biological Process	GO:0031063	regulation of histone deacetylation	453	1.242E-15	1.29665E-11	2.86236E-14	3.439962	17.44205	60	0.002318805	0.007976602	1094	0.004615117	1	0.04404188	1.93289	6.72568	13	21	0.002249913	0.6190476	CAMK2D,CTBP1,JDP2,LPIN1,MAPK8,NCOR1,PML,PRKD2,SKI,SREBF1,TADA2A,TGFB1,VEGFA
GO Biological Process	GO:0048008	platelet-derived growth factor receptor signaling pathway	462	3.79466E-15	3.96162E-11	8.57494E-14	2.585909	34.804	90	0.004626961	0.0119649	517	2.21398E-05	0.2311392	0.000447078	2.185652	9.608115	21	30	0.003634476	0.7	ABL1,ARID5B,BCAR1,BCR,CSF1R,MYO1E,NRP1,PDGFA,PDGFB,PDGFRA,PDGFRB,PDGFRL,PLAT,PLEKHA1,PTPN1,PTPRJ,RAPGEF1,RGS14,TXNIP,VEGFA,ZFAND5
GO Biological Process	GO:0003229	ventricular cardiac muscle tissue development	467	4.55349E-15	4.75385E-11	1.01795E-13	2.23691	53.19839	119	0.007072373	0.01582026	570	5.84652E-05	0.6103768	0.001070836	1.915994	14.0919	27	44	0.004672897	0.6136364	DLL4,DSP,FGFR2,FOXC1,FOXC2,FOXS1,GAA,GATA4,HAND1,HEY1,ISL1,LY6E,MYL3,NKX2-5,NOTCH1,NRG1,PKP2,PROX1,RBPJ,RXRA,SMAD7,TBX5,TGFBR3,TNNI1,TNNI3,TNNT2,TPM1
GO Biological Process	GO:0051291	protein heterooligomerization	500	2.04811E-14	2.13823E-10	4.27645E-13	2.2675	48.51158	110	0.006449293	0.01462377	1034	0.003304023	1	0.03335977	1.515263	21.77839	33	68	0.005711319	0.4852941	ACVRL1,C1QTNF2,C1QTNF6,COL1A1,COL6A1,COL6A2,CTNNA1,CTNNB1,FADD,GRB2,IDE,IGF1R,ILK,INSR,ITPR3,KCNAB2,KCNH7,MED24,P2RX1,P2RX2,PDSS2,PPP5C,PRKAB1,PRKCZ,RRM2,SCUBE3,SEPT11,SEPT9,SQSTM1,STK11,STXBP3,TNFRSF1A,TSC2

A.4.2 EAC hypoDMRs overlapping active enhancers and not promoters GREAT GO Biological Processes significant terms (terms significant in both the binomial and hypergenometric tests (BinomFDRQ<=0.05 and HyperFDRQ<=0.05, were within the top 500 ranked binomial test terms, and had a region fold change of 2 or greater)).

# Ontology	ID	Desc	Binom Rank	Binom P	Binom Bonf P	Binom Fdr Q	Region Fold Enrich	Exp Regions	Obs Regions	Genome Frac	Set Cov	Hyper Rank	Hyper P	Hyper Bonf P	Hyper Fdr Q	Gene Fold Enrich	Exp Genes	Obs Genes	Total Genes	Gene Set Cov	Term Cov	Genes
GO Biological Process	GO:0050872	white fat cell differentiation	6	1.01115E-25	1.05564E-21	1.7594E-22	7.843925	5.864411	46	0.001767454	0.01386377	459	0.001390993	1	0.03163827	3.45613	2.025387	7	12	0.002298851	0.5833333	AACS,CEBPA,CTBP1,CTBP2,NCOR2,PPARG,PRDM16
GO Biological Process	GO:0014015	positive regulation of gliogenesis	29	1.8667E-16	1.94884E-12	6.72012E-14	3.407818	18.78034	64	0.00566014	0.01928873	129	1.6513E-06	0.01723952	0.00013364	3.136656	5.738595	18	34	0.00591133	0.5294118	BMP2,CLCF1,DICER1,ETV5,GFAP,HDAC2,HES1,ID2,JAG1,NKX2-2,NKX6-2,NOTCH1,NTF3,OLIG2,PPARG,SHH,SOX8,TENM4
GO Biological Process	GO:0060579	ventral spinal cord interneuron fate commitment	87	2.65871E-11	2.7757E-07	3.19045E-09	4.139888	7.97123	33	0.00240242	0.00994575	385	0.00067846	1	0.01839771	3.770324	1.856604	7	11	0.002298851	0.6363636	DBX1,GLI2,GLI3,LHX3,NKX2-2,NKX6-2,SOX1
GO Biological Process	GO:0060687	regulation of branching involved in prostate gland morphogenesis	93	4.61942E-11	4.82267E-07	5.18567E-09	4.55835	6.361951	29	0.001917405	0.008740205	506	0.002120933	1	0.04375996	4.231996	1.181476	5	7	0.001642036	0.7142857	BMP4,BMP7,FGFR2,RXRA,SFRP1
GO Biological Process	GO:0071248	cellular response to metal ion	104	8.17883E-11	8.5387E-07	8.21029E-09	2.583839	23.60828	61	0.007115213	0.01838457	413	0.000864948	1	0.02186456	1.85596	14.00892	26	83	0.008538588	0.313253	AKR1C3,ALOXSAP,AOC1,BNIP3,CACNA1H,CAMK2D,CDH1,CEBPA,CYP11A1,CYP11B2,DPEP1,FBP1,FOSB,FST,L3,GRIN1,ID2,JUN,MEF2A,PPARG,RYR3,SHH,SNCA,TFAP2A,TH,TXNRD1,WNT5A
GO Biological Process	GO:0010893	positive regulation of steroid biosynthetic process	114	1.39837E-10	1.4599E-06	1.28062E-08	5.238147	4.581773	24	0.001380884	0.007233273	459	0.001390993	1	0.03163827	3.45613	2.025387	7	12	0.002298851	0.5833333	ABCG1,BMP6,CYP17A1,FGF1,IGF1R,IGF2,WNT4
GO Biological Process	GO:0021521	ventral spinal cord interneuron specification	115	1.60894E-10	1.67974E-06	1.46064E-08	4.064987	7.626101	31	0.002298403	0.009342978	545	0.002599991	1	0.04980533	3.554877	1.687822	6	10	0.001970443	0.6	DBX1,GLI2,GLI3,LHX3,NKX6-2,SOX1
GO Biological Process	GO:0014013	regulation of gliogenesis	120	2.14744E-10	2.24193E-06	1.86828E-08	2.164143	38.35236	83	0.01155888	0.02501507	115	5.05164E-07	0.005273915	4.58601E-05	2.62245	10.29572	27	61	0.008866995	0.4426234	ASCL2,BMP2,CLCF1,DAB1,DICER1,EPHA4,ETV5,GFAP,HDAC2,HES1,HES5,ID2,L6ST,LIN28A,LIN28B,LIN41,NKX2-2,NKX6-2,NOTCH1,NR2E1,NTF3,NTRK3,OLIG2,PPARG,SHH,SKI,SOX11,SOX8,TENM4
GO Biological Process	GO:0061180	mammary gland epithelium development	123	3.17486E-10	3.31455E-06	2.69476E-08	2.123499	40.02827	85	0.01206398	0.02561784	135	1.97254E-06	0.02059328	0.00015254	2.525322	10.29572	26	61	0.008538588	0.4262295	BTRC,CCND1,CEBPB,CSF1,DDR1,ELF5,ETV4,FGFR2,FOXF1,GLI2,ID2,LRP5,LRP6,MGMT,NCOA3,NR3C1,NTN1,PTCH1,SLC12A2,TBX3,TFAP2C,TNFSF11,WNT4,WNT5A,WNT7B,ZNF703
GO Biological Process	GO:0071241	cellular response to inorganic substance	142	8.31852E-10	8.68453E-06	6.11587E-08	2.370086	27.00324	64	0.008138409	0.01928873	366	0.000516714	1	0.01473906	1.86398	15.02162	28	89	0.009195402	0.3146067	AKR1C3,ALOXSAP,AOC1,BLM,BNIP3,CACNA1H,CAMK2D,CDH1,CEBPA,CYP11A1,CYP11B2,DPEP1,FBP1,FOSB,FOXO1,FST3,GRIN1,ID2,JUN,MEF2A,PPARG,RYR3,SHH,SNCA,TFAP2A,TH,TXNRD1,WNT5A
GO Biological Process	GO:0021514	ventral spinal cord interneuron differentiation	146	9.91791E-10	1.03543E-05	7.09198E-08	3.49617	9.724929	34	0.002930961	0.01024714	348	0.000375335	1	0.01126006	3.646028	2.194169	8	13	0.002627258	0.6153846	DBX1,GLI2,GLI3,LHX3,LMO4,NKX2-2,NKX6-2,SOX1
GO Biological Process	GO:0045662	negative regulation of myoblast differentiation	155	1.86924E-09	1.95148E-05	1.25902E-07	3.267362	11.01806	36	0.003320694	0.01084991	302	0.000169507	1	0.00585979	3.949863	2.025387	8	12	0.002627258	0.6666667	BMP4,MKX,NMRK2,NOTCH1,SOX8,SOX9,TBX3,ZFHX3
GO Biological Process	GO:0051148	negative regulation of muscle cell differentiation	156	1.97496E-09	2.06186E-05	1.3217E-07	2.528802	21.35398	54	0.006435799	0.01627486	282	0.000120852	1	0.00447409	2.693089	5.569813	15	33	0.004926108	0.4545455	BHLHE41,BMP2,BMP4,DKK1,HDAC4,MKX,MYOCD,NKX2-5,NMRK2,NOTCH1,SHH,SOX8,SOX9,TBX3,ZFHX3
GO Biological Process	GO:0045687	positive regulation of glial cell differentiation	164	3.34861E-09	3.49595E-05	2.13168E-07	2.760896	16.29905	45	0.004912313	0.01356239	125	7.00975E-07	0.007318178	5.85454E-05	3.702997	4.050773	15	24	0.004926108	0.625	BMP2,CLCF1,DICER1,HDAC2,HES1,JD2,L6ST,NKX2-2,NKX6-2,NOTCH1,NTF3,OLIG2,PPARG,SHH,TENM4
GO Biological Process	GO:0043313	regulation of neutrophil degranulation	167	4.36856E-09	4.56077E-05	2.731E-07	8.159097	1.715876	14	0.000517142	0.004219409	405	0.000810205	1	0.02088529	5.924795	0.675129	4	4	0.001313629	1	ABR,BCR,PRAM1,SYK
GO Biological Process	GO:0048663	neuron fate commitment	188	1.27834E-08	0.000133458	7.09885E-07	2.011898	39.26641	79	0.01183436	0.02380952	212	3.40195E-05	0.3551638	0.001675301	2.293469	10.4645	24	62	0.007881773	0.3870968	BMP4,DBX1,FOXA3,GLI2,GLI3,HES1,HES5,ID2,LBX1,LHX3,NKX2-2,NKX6-2,NOTCH1,NR6S1,NTRK3,OLIG1,OLIG2,PAX3,PAX7,PTF1A,SHH,SOX1,TFAP2C,TFGBR1
GO Biological Process	GO:0007440	foregut morphogenesis	219	5.68959E-08	0.000593993	2.7123E-06	3.471008	7.77872	27	0.0023444	0.008137432	302	0.000169507	1	0.00585979	3.949863	2.025387	8	12	0.002627258	0.6666667	EPB41L5,FOXF1,GATA4,NOTCH1,SHH,SMAD2,SMAD3,SOX17
GO Biological Process	GO:0035019	somatic stem cell maintenance	255	1.60152E-07	0.001671983	6.5568E-06	2.186011	25.15999	55	0.007582877	0.01657625	495	0.001867134	1	0.03937955	2.241814	6.244942	14	37	0.004597701	0.3783784	ASCL2,BCL9,FZD7,HES1,NR2E1,POU5F1,PRDM16,SFRP1,SKI,SOX9,TFAP2C,WNT7A,ZFP36L2,ZHX2
GO Biological Process	GO:0055017	cardiac muscle tissue growth	265	2.05641E-07	0.002146891	8.10148E-06	3.020575	9.931882	30	0.002993334	0.009041591	448	0.001234306	1	0.02876374	2.962397	3.03808	9	18	0.002955665	0.5	FOXC1,GATA4,MYOCD,NKX2-5,NOTCH1,PRKAR1A,RXRA,TENM4,TGFBR3
GO Biological Process	GO:0060419	heart growth	275	2.46302E-07	0.002571392	9.35052E-06	2.994156	10.01952	30	0.003019746	0.009041591	501	0.001996071	1	0.04159477	2.806482	3.206862	9	19	0.002955665	0.4736842	FOXC1,GATA4,MYOCD,NKX2-5,NOTCH1,PRKAR1A,RXRA,TENM4,TGFBR3
GO Biological Process	GO:0060612	adipose tissue development	287	3.27965E-07	0.003423949	1.19301E-05	2.641159	13.63038	36	0.00410801	0.01084991	503	0.002004442	1	0.04160313	2.506644	4.388338	11	26	0.003612479	0.4230769	AACS,ARID5B,CREB5,CSF1,FTO,JD2,LRP5,PAXIP1,SH3PXD2B,SOX8,UBBP4
GO Biological Process	GO:0032148	activation of protein kinase B activity	288	3.28835E-07	0.003433041	1.19203E-05	3.090799	9.059147	28	0.002730304	0.008438819	275	0.000113349	1	0.00430313	3.485173	2.869298	10	17	0.003284072	0.5882353	ADRA2A,ADRA2C,GAS6,INS,INSR,NTF3,NTRK3,PDGFB,PRKCZ,WNT5A
GO Biological Process	GO:0010830	regulation of myotube differentiation	342	8.79489E-07	0.009181862	2.68476E-05	2.021577	28.69047	58	0.008646916	0.01748041	437	0.001150829	1	0.02749349	2.091104	8.607893	18	51	0.00591133	0.3529412	BHLHE41,BMP4,CYP26B1,EHD2,HDAC4,IGFBP3,MKX,MYOCD,MYOD1,NKX2-5,NMRK2,NOTCH1,RBM38,SIK1,SOX8,SOX9,TBX3,ZFHX3
GO Biological Process	GO:0072080	nephron tubule development	355	9.8578E-07	0.01029154	2.89903E-05	2.090088	25.35779	53	0.007642492	0.01597348	457	0.001371125	1	0.03132286	2.304087	6.07616	14	36	0.004597701	0.3888889	DLL1,HES1,HES5,JRK1,JRK2,JRK3,JAG1,MEF2A,NOTCH1,PKD1,POU3F3,SOX8,SOX9,WNT7B
GO Biological Process	GO:0048617	embryonic foregut morphogenesis	372	1.46061E-06	0.01524878	4.09913E-05	3.166128	7.580237	24	0.00228458	0.007233273	385	0.00067846	1	0.01839771	3.770324	1.856604	7	11	0.002298851	0.6363636	EPB41L5,FOXF1,GATA4,SHH,SMAD2,SMAD3,SOX17
GO Biological Process	GO:0045655	regulation of monocyte differentiation	379	1.77436E-06	0.01852435	4.88769E-05	4.606401	3.256338	15	0.000981416	0.004520796	545	0.002599991	1	0.04980533	3.554877	1.687822	6	10	0.001970443	0.6	ACIN1,CSF1,DCSTAMP,JUN,MYC,ZBTB46
GO Biological Process	GO:0061326	renal tubule development	381	1.85773E-06	0.01939472	5.09048E-05	2.028522	26.62037	54	0.008023016	0.01627486	402	0.000776943	1	0.02017731	2.338735	6.413724	15	38	0.004926108	0.3947368	DLL1,HES1,HES5,JRK1,JRK2,JRK3,JAG1,MEF2A,NOTCH1,PKD1,POU3F3,SOX8,SOX9,TTC8,WNT7B
GO Biological Process	GO:0006208	pyrimidine nucleobase catabolic process	410	3.1984E-06	0.03339126	8.14421E-05	3.919997	4.336738	17	0.001307034	0.005123568	506	0.002120933	1	0.04375996	4.231996	1.181476	5	7	0.001642036	0.7142857	CRMP1,DPYS,DPYSL3,DPYSL4,DPYSL5
GO Biological Process	GO:0090031	positive regulation of steroid hormone biosynthetic process	419	3.67406E-06	0.03835716	9.15445E-05	4.986282	2.607153	13	0.00078576	0.003918023	388	0.000704474	1	0.01895542	4.937329	1.012693	5	6	0.001642036	0.8333333	BMP6,CYP17A1,IGF1R,IGF2,WNT4
GO Biological Process	GO:0021532	neural tube patterning	429	5.08338E-06	0.05307047	0.000123707	2.273937	16.7111	38	0.005036499	0.01145268	486	0.001758496	1	0.0377751	2.33401	5.569813	13	33	0.004269294	0.3939394	BMP4,GBX2,GLI2,GLI3,GSC,HES1,HES3,LRP6,PAX7,PTCH1,RPGRIP1L,SHH,SOX17
GO Biological Process	GO:0072070	loop of Henle development	434	5.59661E-06	0.05842862	0.000134628	2.845065	8.787146	25	0.002648326	0.007534659	149	4.35952E-06	0.0455134	0.000305459	4.847559	1.856604	9	11	0.002955665	0.8181818	DLL1,HES5,JRK1,JRK2,JRK3,JAG1,PKD1,POU3F3,WNT7B
GO Biological Process	GO:0072273	metanephric nephron morphogenesis	435	5.88027E-06	0.06139	0.000141126	2.350933	14.88771	35	0.004486952	0.01054852	430	0.001068428	1	0.02594044	2.821331	3.544427	10	21	0.003284072	0.4761905	BMP4,HES1,HES5,KIF26B,LHX1,PKD1,SIX2,SOX8,SOX9,WNT4
GO Biological Process	GO:0021575	hindbrain morphogenesis	455	7.59054E-06	0.07924524	0.000174165	2.134372	19.67792	42	0.005930657	0.01265823	127	1.43545E-06	0.0149861	0.000118001	2.962397	6.751289	20	40	0.006568144	0.5	ATP2B2,CACNA1A,CBS,DAB1,DLC1,FAM2,GLI2,GNPAT,HES1,HES3,KNDC1,LHX1,LHX5,LMX1B,LRP6,PROX1,RORA,SKOR2,ULK1,WNT7A
GO Biological Process	GO:0045600	positive regulation of fat cell differentiation	494	1.36267E-05	0.1422623	0.00028798	2.395476	12.94106	31	0.00390026	0.009342978	363	0.000488222	1	0.01404143	2.513549	5.569813	14	33	0.004597701	0.4242424	BMP2,CEBPA,CMKLR1,FNDC3B,JD2,NS1RP5,MEDAG,PPARG,SFRP1,SHSPXD2B,WNT5B,ZBTB16,ZC3H12A

A.4.3 GBM hyperDMRs overlapping active enhancers and not promoters GREAT GO Biological Processes significant terms (terms significant in both the binomial and hypergenometric tests (BinomFDRQ<=0.05 and HyperFDRQ<=0.05, were within the top 500 ranked binomial test terms, and had a region fold change of 2 or greater)).

# Ontology	ID	Desc	Binom Rank	Binom P	Binom Bonf P	Binom Fdr Q	Region Fold Enrich	Exp Regions	Obs Regions	Genome Frac	Set Cov	Hyper Rank	Hyper P	Hyper Bonf P	Hyper Fdr Q	Gene Fold Enrich	Exp Genes	Obs Genes	Total Genes	Gene Set Cov	Term Cov	Genes
GO Biological Process	GO:0009952	anterior/posterior pattern specification	113	8.80978E-20	9.19741E-16	8.1393E-18	2.034663	96.33043	196	0.0267436	0.05441421	107	3.64972E-11	3.81031E-07	3.56104E-09	1.980802	41.39737	82	195	0.02140992	0.4205128	ACVR2A,ACVR2B,ARC,AXIN1,AXIN2,BARX1,BASP1,BMP4,BMPR1A,CDX2,CELSR2,COBL,CREBBP,CTNNBIP1,DLL1,FOXA2,FOXB1,FOXC1,FOXF1,FZD5,GATA4,GBX2,GLI3,GRSF1,HES1,HEY1,HIPK1,HOXA1,HOXA11,HOXA2,HOXA3,HOXA4,HOXA6,HOXA7,HOXB1,HOXB2,HOXB3,HOXB4,HOXB5,HOXB6,HOXB8,HOXB9,HOXC10,HOXC11,HOXC4,HOXC5,HOXC6,HOXC8,HOXC9,HOXD10,HOXD3,HOXD8,HTT,KDM2B,KMT2A,LFNG,LRP6,NOTCH1,NR2F2,OSR1,PAX1,PBX1,PBX3,PCSK6,PGAP1,PLD6,RARG,SF3B1,SHH,SKI,SMAD3,SOX17,SSBP3,T,TBX3,TCF7L1,TSHZ1,WNT1,WNT3A,WT1,ZBTB16,ZEB2
GO Biological Process	GO:0021546	rhombomere development	126	3.5096E-18	3.66403E-14	2.90796E-16	8.45646	3.547584	30	0.000984893	0.008328706	572	0.001712651	1	0.03125887	3.532833	1.698354	6	8	0.00156658	0.75	GBX2,HOXA1,HOXA2,HOXB1,HOXB2,HOXB3
GO Biological Process	GO:0042552	myelination	137	7.71817E-17	8.05777E-13	5.88159E-15	2.756765	32.64697	90	0.009063566	0.02498612	645	0.002862774	1	0.04633699	1.673447	16.13436	27	76	0.007049608	0.3552632	AFG3L2,AKT1,AKT2,ARHGEF10,CD9,CNTN2,CNTNAP1,DICER1,FA2H,FGFR3,JAM3,MAL,MBP,MYO5A,NAB2,NDRG1,NFASC,NKX6-2,PLLP,PMP22,POU3F1,QKI,SCN8A,SERINC5,SKI,TGFB1,UGT8
GO Biological Process	GO:0032012	regulation of ARF protein signal transduction	139	8.57764E-17	8.95506E-13	6.44249E-15	3.646314	16.45497	60	0.004568288	0.01665741	501	0.001091465	1	0.0227443	1.962685	10.19012	20	48	0.005221932	0.4166667	ACAP3,ADAP1,AGAP1,AGAP2,AGAP5,AGFG2,ARAP2,ARFGAP3,ASAP1,ASAP2,CYTH1,CYTH2,CYTH3,CYTH4,DAB2IP,GIT1,IQSEC1,IQSEC3,PSD2,PSD3
GO Biological Process	GO:0010721	negative regulation of cell development	149	4.92845E-16	5.1453E-12	3.45322E-14	2.139833	64.95833	139	0.01803396	0.03858967	221	3.89648E-06	0.04067926	0.000184069	1.853289	25.89989	48	122	0.01253264	0.3934426	AATK,ARHGEF2,ATF5,BCL11A,BMPR1A,CDK5,CDKSR1,CIT,DAB2IP,DICER1,DMTN,EPHB2,F2,FGFR3,FOXA1,FOXA2,HES1,HOOK3,ID2,ITGB1BP1,KANK1,LINGO1,MAG,MBP,NGFR,NKX2-1,NKX6-1,NKX6-2,NOTCH1,NTRK3,PTK2,RAPGEF2,RTN4,RTN4R,RUFY3,RYK,SKI,SLIT1,SMAD7,TBX3,TGFB1,THY1,TNR,TRIM11,TTC3,ULK1,WNT3A,WNT7A
GO Biological Process	GO:0072659	protein localization to plasma membrane	183	3.30603E-14	3.4515E-10	1.88607E-12	2.840475	24.99582	71	0.006939427	0.01971127	473	0.000798252	1	0.01761893	1.78233	15.70977	28	74	0.007310705	0.3783784	AKT2,ANK1,ANK2,ANK3,BBS2,DLG1,EHD2,EHD3,GAS6,GOLGA4,GOLPH3,KIF13A,MACF1,MYO5A,NFASC,NKD2,PALM,PLEKHF1,PRKCZ,PTCH1,RAB31,RAMP1,SPTBN1,SPTBN4,TESC,VAMP3,WNT4,ZDHHC22
GO Biological Process	GO:0045668	negative regulation of osteoblast differentiation	189	4.23755E-14	4.42401E-10	2.34074E-12	3.231555	17.94802	58	0.004982792	0.01610217	551	0.001581999	1	0.02997472	2.141111	7.005709	15	33	0.003916449	0.4545455	AXIN2,CDK6,HAND2,HDAC4,ID2,NOTCH1,PTCH1,SEMA4D,SKI,SMAD3,SOX2,TOB1,TWIST1,TWIST2,TWSG1
GO Biological Process	GO:0048704	embryonic skeletal system morphogenesis	224	6.56091E-13	6.84959E-09	3.05785E-11	2.16908	48.86864	106	0.01356708	0.0294281	100	7.17141E-12	7.48696E-08	7.48696E-10	2.569333	18.68189	48	88	0.01253264	0.5454545	AXIN1,BMP4,CHST11,DLG1,DSCAML1,FGFR2,FUZ,GLI3,GSC,HOXA1,HOXA11,HOXA2,HOXA3,HOXA4,HOXA6,HOXA7,HOXB1,HOXB2,HOXB3,HOXB4,HOXB5,HOXB6,HOXB8,HOXC11,HOXC9,HOXD10,HOXD3,HYAL1,IRX5,MDFI,MEF2A,MEF2C,NDST1,NIPBL,OSR1,OSR2,PDGFRA,PRRX1,RDH10,SATB2,SHOX2,SIX2,SMAD3,SMO,TBX15,TFAP2A,TWIST1,TWIST2
GO Biological Process	GO:0019228	regulation of action potential in neuron	227	8.59661E-13	8.97486E-09	3.95368E-11	2.206854	45.7665	101	0.01270586	0.02803998	441	0.00054021	1	0.01278864	1.699645	20.59254	35	97	0.009138381	0.3608247	AFG3L2,AKT1,AKT2,ANK3,ARHGEF10,CD9,CNTN2,CNTNAP1,DICER1,DPP6,FA2H,FGFR3,GPER,JAM3,KCNMA1,KCNMB4,MAL,MBP,MYO5A,NAB2,NDRG1,NFASC,NKX6-2,P2RX2,P2RX4,P2RX7,PLLP,PMP22,POU3F1,QKI,SCN8A,SERINC5,SKI,TGFB1,UGT8
GO Biological Process	GO:0043313	regulation of neutrophil degranulation	235	1.89298E-12	1.97627E-08	8.40967E-11	9.663161	1.862744	18	0.000517142	0.004997224	589	0.002028694	1	0.03595852	4.710444	0.849177	4	4	0.001044386	1	ABR,BCR,PRAM1,SYK
GO Biological Process	GO:0007009	plasma membrane organization	238	3.23334E-12	3.37561E-08	1.41832E-10	2.337185	36.36853	85	0.01009676	0.023598	444	0.000565427	1	0.01329516	1.657378	22.92778	38	108	0.009921671	0.3518519	AKT1,AKT2,ANK1,ANK2,ANK3,BBS2,COL5A1,DLG1,DYSF,EHD2,EHD3,FA2H,FNBP1L,GAS6,GOLGA4,GOLPH3,KIF13A,MACF1,MYH10,MYO5A,NDRG1,NFASC,NKD2,P2RX7,PACSIN2,PALM,PLEKHF1,PLSCR1,PRKCZ,PTCH1,RAB31,RAMP1,SPTBN1,SPTBN4,TESC,VAMP3,WNT4,ZDHHC22

# Ontology	ID	Desc	Binom Rank	Binom P	Binom Bonf P	Binom Fdr Q	Region Fold Enrich	Exp Regions	Obs Regions	Genome Frac	Set Cov	Hyper Rank	Hyper P	Hyper Bonf P	Hyper Fdr Q	Gene Fold Enrich	Exp Genes	Obs Genes	Total Genes	Gene Set Cov	Term Cov	Genes
GO Biological Process	GO:0043433	negative regulation of sequence-specific DNA binding transcription factor activity	256	1.06654E-11	1.11347E-07	4.3495E-10	2.058428	51.98142	107	0.01443127	0.02970572	327	8.25041E-05	0.8613426	0.002634075	1.731189	24.83842	43	117	0.01122715	0.3675214	ARRB2,CACTIN,CHP1,COMMD7,CTNNBIP1,CYLD,DAB2IP,DAP,ESR1,FAM105B,FOXA2,GAS6,GLIS2,HAND1,HAND2,HDAC4,JD2,IRAK3,KAT6A,KDM1A,KLF4,NLRC5,NLRP12,PBX1,PEX14,PIAS4,PRMT2,PRNP,PROX1,PTCH1,RB1,RPS3,SETD6,SIGIRR,SIK1,SIVA1,SMAD7,SUMO1,TRAF3,TWIST1,USP7,WFS1,WWP2
GO Biological Process	GO:0021903	rostrocaudal neural tube patterning anterior/posterior	276	3.81009E-11	3.97773E-07	1.44121E-09	4.935316	5.470774	27	0.001518816	0.007495836	640	0.002824613	1	0.0460765	2.997555	2.335236	7	11	0.001827676	0.6363636	GBX2,HES1,KDM2B,LRP6,SOX17,SSBP3,WNT1
GO Biological Process	GO:0008595	axis specification, embryo	290	7.60659E-11	7.94128E-07	2.73837E-09	3.87279	8.779199	34	0.002437312	0.0094392	368	0.000190738	1	0.005411165	3.140296	3.184413	10	15	0.002610966	0.6666667	BASP1,FZD5,HOXB6,HOXD8,PCSK6,PL D6,T,TBX3,TCF7L1,WT1
GO Biological Process	GO:0033762	response to glucagon stimulus	293	8.87259E-11	9.26298E-07	3.16143E-09	3.0239	15.54284	47	0.004315059	0.01304831	480	0.000857975	1	0.01866096	2.034055	9.340946	19	44	0.004960836	0.4318182	ADCY2,ADCY3,ADCY4,ADCY5,ADCY7,ADCY9,GCGR,GLP1R,GNB1,GNB2,GNB3,GNB4,GNG13,GNG7,GNGR,PRKACB,PRKAR1B,SREBF1,STK11
GO Biological Process	GO:2001242	regulation of intrinsic apoptotic signaling pathway	305	1.3022E-10	1.35949E-06	4.45736E-09	2.653222	21.48332	57	0.005964275	0.01582454	578	0.001841732	1	0.03326589	1.718675	15.70977	27	74	0.007049608	0.3648649	ACKR3,AKT1,ARHGEF2,BBC3,BID,CD44,CLU,CXCL12,CYLD,DFNA5,FIS1,GNB2L1,HSPB1,KDM1A,KIAA1967,LCK,NACC2,NFE2L2,PPIF,RIPK1,SIAH1,SNAI2,TPT1,TRIM32,URI1,WFS1,YBX3
GO Biological Process	GO:0072661	protein targeting to plasma membrane	309	1.57722E-10	1.64662E-06	5.32886E-09	4.320217	6.712625	29	0.001863583	0.008051083	350	0.000133089	1	0.003969858	3.047934	3.609002	11	17	0.002872063	0.6470588	ANK1,ANK2,ANK3,EHD3,GAS6,NFASC,NKD2,PALM,PTCH1,SPTBN1,TESC
GO Biological Process	GO:0071377	cellular response to glucagon stimulus	345	7.54862E-10	7.88076E-06	2.28428E-08	3.037344	13.82787	42	0.003838942	0.01166019	458	0.000674074	1	0.01536536	2.164258	7.854886	17	37	0.004438642	0.4594595	ADCY2,ADCY3,ADCY4,ADCY5,ADCY7,ADCY9,GCGR,GLP1R,GNB1,GNB2,GNB3,GNB4,GNG13,GNG7,GNG8,PRKACB,PRKAR1B
GO Biological Process	GO:0090002	establishment of protein localization to plasma membrane	347	8.23495E-10	8.59729E-06	2.47761E-08	3.07754	13.32233	41	0.003698593	0.01138257	366	0.000188599	1	0.005379701	2.190904	9.128651	20	43	0.005221932	0.4651163	ANK1,ANK2,ANK3,BBS2,EHD3,GAS6,GOLGA4,GOLPH3,KIF13A,MACF1,NFASC,NKD2,PALM,PTCH1,RAB31,SPTBN1,SPTBN4,TESC,VAMP3,WNT4
GO Biological Process	GO:0050885	neuromuscular process controlling balance	350	1.03023E-09	1.07556E-05	3.07304E-08	2.339661	27.7818	65	0.007712881	0.01804553	237	6.42459E-06	0.06707267	0.000283007	2.310784	11.25159	26	53	0.006788512	0.490566	ABR,ADCY5,APLP2,APP,ATP2B2,BCR,BLOC1S4,CACNA1A,CAMK2B,CDH23,CLIC5,CNTNAP1,GRIN2C,HERC1,HMX3,JPH3,JPH4,KCNMA1,MYH10,NKX6-2,PAFAH1B1,RBFOX1,RBFOX2,SHANK3,SOX2,TNR
GO Biological Process	GO:0051693	actin filament capping	370	1.69406E-09	1.7686E-05	4.77999E-08	4.272843	6.084941	26	0.001689323	0.007218212	631	0.002619572	1	0.04334125	2.261013	5.307355	12	25	0.003133159	0.48	ADD1,AVIL,CAPG,CAPZA2,CAPZB,DMTN,EPS8,SCIN,SPTAN1,SPTBN1,SPTBN4,VILL
GO Biological Process	GO:0060684	epithelial-mesenchymal cell signaling	371	1.7337E-09	1.80998E-05	4.87865E-08	4.600771	5.216517	24	0.001448228	0.006662965	436	0.000522729	1	0.01251673	4.037523	1.48606	6	7	0.00156658	0.8571429	BMP4,CDC42,FOXA1,PDGFA,SHH,SMO
GO Biological Process	GO:0033182	regulation of histone ubiquitination	375	1.89586E-09	1.97928E-05	5.27807E-08	11.12211	1.078932	12	0.000299537	0.003331483	589	0.002028694	1	0.03595852	4.710444	0.849177	4	4	0.001044386	1	OTUB1,TRIP12,UBE2N,UBR5
GO Biological Process	GO:0051101	regulation of DNA binding	384	2.71181E-09	2.83113E-05	7.37274E-08	2.077877	38.50084	80	0.01068874	0.02220988	471	0.000778885	1	0.01726445	1.827933	14.22371	26	67	0.006788512	0.3880597	ACD,BRD4,CTNNBIP1,DOT1L,HABP4,HAND2,HES1,HEY1,ID2,ISL1,JUN,KDM1A,LHX2,MDFI,PDX1,PITX2,PPARA,SKI,SMARCA4,SMO,SUMO1,TGFB1,TNKS,TWIST1,TWIST2,ZNF462
GO Biological Process	GO:0003166	bundle of His development	400	5.13917E-09	5.3653E-05	1.34132E-07	5.2037	3.843419	20	0.001067024	0.005552471	589	0.002028694	1	0.03595852	4.710444	0.849177	4	4	0.001044386	1	ID2,NKX2-5,TBX3,TBX5
GO Biological Process	GO:0031532	actin cytoskeleton reorganization	406	6.12323E-09	6.39266E-05	1.5745E-07	2.399237	23.34075	56	0.006479941	0.01554692	251	1.3051E-05	0.1362526	0.000542839	2.472983	8.491769	21	40	0.005483029	0.525	ANTXR1,BRSK2,CDC42BPA,CDC42BPB,DMTN,EPS8,FARP2,FER,KIT,LPIN1,MICALL2,MINK1,MYH9,PARVB,PHACTR1,PRKCZ,PTPN1,RAP2A,RICTOR,TNIK,TRPV4
GO Biological Process	GO:0030705	cytoskeleton-dependent intracellular transport	426	1.01498E-08	0.000105964	2.48742E-07	2.322451	24.97362	58	0.006933265	0.01610217	404	0.000340675	1	0.008803579	1.802762	17.19583	31	81	0.008093995	0.382716	AP3B1,AP3D1,APBA1,APP,ARHGAP21,BICD1,BLOC1S4,CDC42,CNTNAP1,DST,HTT,KIF13A,KIF1A,KIF1B,KLC1,LRPPRC,MAP1B,MAPT,MGARP,MYO10,NDE1,NDEL1,NEFM,OPA1,PAFAH1B1,PEX14,PRKCZ,RAB6A,RHOT1,SSNA1,UGT8
GO Biological Process	GO:0010970	microtubule-based transport	435	1.50155E-08	0.000156762	3.60373E-07	2.335344	23.97934	56	0.006657228	0.01554692	380	0.0002246	1	0.0061706	1.859386	16.13436	30	76	0.007832898	0.3947368	AP3B1,AP3D1,APBA1,APP,ARHGAP21,BICD1,BLOC1S4,CDC42,CNTNAP1,DST,HTT,KIF13A,KIF1A,KIF1B,KLC1,LRPPRC,MAP1B,MAPT,MGARP,NDE1,NDEL1,NEFM,OPA1,PAFAH1B1,PEX14,PRKCZ,RAB6A,RHOT1,SSNA1,UGT8
GO Biological Process	GO:0003207	cardiac chamber formation	462	2.9204E-08	0.000304889	6.59934E-07	3.038103	11.19119	34	0.003106939	0.0094392	333	0.000101551	1	0.003183749	3.532833	2.547531	9	12	0.002349869	0.75	HAND1,HAND2,MEF2A,MEF2C,NKX2-5,NOTCH1,SOX4,TBX20,TBX5
GO Biological Process	GO:0003211	cardiac ventricle formation	482	5.20351E-08	0.000543246	1.12707E-06	3.085358	10.37157	32	0.002879392	0.008883953	407	0.000362571	1	0.009300342	3.425777	2.335236	8	11	0.002088773	0.7272727	HAND1,HAND2,MEF2A,MEF2C,NKX2-5,NOTCH1,SOX4,TBX5

A.4.4 GBM hypoDMRs overlapping active enhancers and not promoters GREAT GO Biological Processes significant terms (terms significant in both the binomial and hypergenometric tests (BinomFDRQ<=0.05 and HyperFDRQ<=0.05, were within the top 500 ranked binomial test terms, and had a region fold change of 2 or greater)).

# Ontology	ID	Desc	Binom Rank	Binom P	Binom Bonf P	Binom Fdr Q	Region Fold Enrich	Exp Regions	Obs Regions	Genome Frac	Set Cov	Hyper Rank	Hyper P	Hyper Bonf P	Hyper Fdr Q	Gene Fold Enrich	Exp Genes	Obs Genes	Total Genes	Gene Set Cov	Term Cov	Genes
GO Biological Process	GO:0048534	hematopoietic or lymphoid organ development	23	2.48601E-19	2.59539E-15	1.12843E-16	2.004776	95.2725	191	0.05503899	0.1103408	113	1.78172E-07	0.001860114	1.64612E-05	1.779627	45.51516	81	447	0.04409363	0.1812081	ACE,ADD2,ANK1,BARX1,BCL2L11,BCL3,CARD11,CBFA2T3,CD1D,CD248,CEBPA,CEBPG,CHD7,CSF1,DLL1,DMTN,EPAS1,EPHB3,EPO,FADD,FAM20C,FGF3,FOXL1,FOXN1,GATA2,GATA3,GLI3,GLRX5,HDAC4,HES1,ID2,IKZF1,IL12B,INHBA,IRF4,IRF8,JAG1,JAM3,JUN,KLF4,LCK,LRP5,LTB,LY6D,LYN,MECOM,MEF2A,MEIS1,MELK,MMP9,NHEJ1,NKX2-5,PDE2A,PLCG2,POU2F2,RCOR1,RHOH,RUNX1,RUNX2,RYR1,SFRP2,SHH,SMAD5,SOX8,SYK,TACC3,TBX1,TGFBR2,TNF,TNRC6C,TPO,TRAF6,TTC7A,VEGFA,WNT4,ZBTB16,ZFAT,ZFP36L1,ZFP36L2,ZFPM1,ZNF683
GO Biological Process	GO:0055123	digestive system development	29	5.13836E-18	5.36445E-14	1.84981E-15	2.663447	37.92079	101	0.02190687	0.05834778	116	2.51243E-07	0.002622981	2.26119E-05	2.572141	12.82978	33	126	0.01796407	0.2619048	ALX4,BARX1,DACT1,EGFR,EPHB3,FGF9,FGFR3,FOXF1,FOXF2,FOXL1,FOXP4,GATA4,GATA5,GATA6,GLI2,GLI3,HNF1B,ID2,IGF1R,IGF2,IHH,NODAL,NOTCH1,PTK6,SFRP2,SHH,SMAD2,SMAD3,SOX11,SOX17,TGFBR2,TNF,WNT11
GO Biological Process	GO:0048562	embryonic organ morphogenesis	51	7.92247E-16	8.27106E-12	1.62178E-13	2.057426	70.96245	146	0.04099506	0.08434431	27	1.44354E-13	1.50705E-09	5.58167E-11	2.547528	27.08508	69	266	0.03756124	0.2593985	ALDH1A3,ALX3,ALX4,ATP6V1B1,CELSR1,CHD7,CHRNA9,CHST11,CLIC5,COL2A1,CRYAA,DLL1,DSCAML1,EPHB2,FGF3,FGF9,FGFR1,FOXF1,FOXF2,FOXL2,GATA2,GATA3,GATA4,GLI2,GLI3,HES1,HNF1B,HOXB1,HOXC4,HPN,HSPG2,ID2,IHH,INSIG1,INSIG2,LRP5,MAFB,MEF2A,MIB1,MYO15A,MYO7A,NKX2-5,NODAL,NOTCH1,NTN1,OSR1,OTOP1,ROR2,RPL38,RUNX2,RYR2,SETD2,SHH,SMAD2,SMAD3,SOX11,SOX17,SOX18,SOX9,TBX1,TBX15,TBX3,TEAD2,TGFBR2,TH,TMIE,TWIST2,WNT9A,ZEB1
GO Biological Process	GO:0042493	response to drug	61	6.05425E-15	6.32064E-11	1.03617E-12	2.016227	71.42052	144	0.04125969	0.08318891	101	6.05187E-08	0.000631815	6.2556E-06	1.920288	36.45286	70	358	0.03810561	0.1955307	ABCC1,ABCC6,ACSL1,ADCY1,ADORA2A,ALDH3A1,AOC1,CCL2,CCND1,CDKN1A,COMT,CPT1A,CRYAA,CYP2B6,DPEP1,DRD1,ENSG00000267545,FYN,GAD2,GATA3,GATA4,GATA6,GNA12,HDAC4,IGF2,IL1RN,INHBA,JUN,KCNH2,KCNK3,KCNQ1,LCK,LCN2,LPL,LYN,MAS1,MEF2A,MGMT,MYC,NDUFA10,NFATC2,NKX3-1,NNMT,NOS1,PDE2A,PDGFA,PGF,PRKCD,PTK2,PTPRM,RAB6C,RPH3AL,SEMA3C,SFRP2,SLC15A4,SLC18A1,SLC19A1,SLC22A2,SLC22A3,SLC6A11,SRP14,TERF1,TGFA,TGFBR2,TH,TNF,TOP1,UGT1A1,UQCRFS1,XPC
GO Biological Process	GO:0050870	positive regulation of T cell activation	63	6.73043E-15	7.02657E-11	1.11533E-12	2.667487	30.74054	82	0.01775884	0.04737146	234	1.7393E-05	0.1815824	0.000775993	2.095925	16.69907	35	164	0.0190528	0.2134146	ADAM8,CARD11,CCL2,CD1D,CD5,FADD,FYN,GATA3,GLI2,GLI3,GRAP2,HES1,HLA-DPA1,HLA-DRA,HLA-DRB5,ICOS,ICOSLG,IGF2,IHH,IKZF1,IL12B,IL1RL2,IL2RA,IL4R,LCK,LYN,PRKCQ,PRKCZ,PTPN6,SART1,SHH,SRC,SYK,TGFBR2,TRAF6
GO Biological Process	GO:0043588	skin development	76	6.98663E-14	7.29404E-10	9.59742E-12	2.115669	56.24698	119	0.03249392	0.06874639	196	5.74531E-06	0.05998106	0.000306026	1.86562	28.40879	53	279	0.02885139	0.1899642	ADAMTS2,ALX4,APCDD1,BARX2,CALML5,CELSR1,CLIC4,CLIC5,COL1A1,COL5A1,CST6,CUX1,DACT2,DBI,EDAR,EGFR,FGFR1,FOXN1,GJB5,GLI2,HDAC1,HES1,INHBA,IVL,JAG1,KLF4,KRT17,KRT31,KRT34,LCE1F,LOR,LTB,MSX2,MYO7A,NGFR,NOTCH1,NTF3,PDGFA,POU2F3,POU3F1,POU3F2,PPARA,RUNX1,RUNX3,RYR1,S100A7,SHH,SNAI1,SOX18,SOX9,SPRR4,TGM3,TNFRSF19
GO Biological Process	GO:0051251	positive regulation of lymphocyte activation	84	2.06214E-13	2.15288E-09	2.56295E-11	2.311639	41.09639	95	0.02374141	0.05488157	136	6.29876E-07	0.0065759	4.83522E-05	2.120946	21.68843	46	213	0.02504083	0.2159624	ADAM8,CARD11,CCL2,CD1D,CD40,CD5,CDKN1A,ENSG00000267545,FADD,FYN,GATA3,GLI2,GLI3,GRAP2,HES1,HLA-DPA1,HLA-DRA,HLA-DRB5,ICOS,ICOSLG,IGF2,IHH,IKZF1,IL12B,IL15RA,IL1RL2,IL2RA,IL4R,INPP5D,RS2,LCK,LYN,MEF2A,NFATC2,PRKCQ,PRKCZ,PTPN6,SART1,SHH,SRC,SYK,TGFBR2,TNFRSF4,TNIP2,TRAF6,ZNF335
GO Biological Process	GO:0050863	regulation of T cell activation	86	3.08927E-13	3.2252E-09	3.75023E-11	2.28312	42.04773	96	0.02429101	0.05545927	296	0.000130045	1	0.004586726	1.793382	23.41943	42	230	0.02286336	0.1826087	ADAM8,ADORA2A,CARD11,CCL2,CD1D,CD5,ERBB2,FADD,FOXN1,FYN,GATA3,GLI2,GLI3,GRAP2,HES1,HLA-DPA1,HLA-DRA,HLA-DRB5,ICOS,ICOSLG,IGF2,IHH,IKZF1,IL12B,IL1RL2,IL2RA,IL4R,IRF4,LCK,LRRC32,LYN,PRKCQ,PRKCZ,PRNP,PTPN6,SART1,SHH,SRC,SYK,TGFBR2,TRAF6,ZEB1
GO Biological Process	GO:0048565	digestive tract development	88	3.63086E-13	3.79062E-09	4.30752E-11	2.407143	35.727	86	0.02063952	0.04968226	155	1.16099E-06	0.01212069	7.8198E-05	2.539889	11.81154	30	116	0.01633097	0.2586207	ALX4,DACT1,EGFR,EPHB3,FGF9,FGFR3,FOXF1,FOXF2,FOXL1,FOXP4,GATA4,GATA5,GATA6,GLI2,GLI3,HNF1B,ID2,IHH,NODAL,NOTCH1,PTK6,SFRP2,SHH,SMAD2,SMAD3,SOX11,SOX17,TGFBR2,TNF,WNT11

# Ontology	ID	Desc	Binom Rank	Binom P	Binom Bonf P	Binom Fdr Q	Region Fold Enrich	Exp Regions	Obs Regions	Genome Frac	Set Cov	Hyper Rank	Hyper P	Hyper Bonf P	Hyper Fdr Q	Gene Fold Enrich	Exp Genes	Obs Genes	Total Genes	Gene Set Cov	Term Cov	Genes
GO Biological Process	GO:0043467	regulation of generation of precursor metabolites and energy	90	4.21453E-13	4.39996E-09	4.88885E-11	3.873277	10.84353	42	0.006264316	0.02426343	539	0.002359661	1	0.04570475	2.291544	6.109417	14	60	0.007621121	0.2333333	ADRA1B,DDIT4,GCGR,GCK,HDAC4,JGF2,JNS,JRS2,MNF1,NKX1-1,PPARA,PPP1CB,PRDM16,PRKAG2
GO Biological Process	GO:0051249	regulation of lymphocyte activation	102	1.90264E-12	1.98636E-08	1.94741E-10	2.023837	57.81098	117	0.03339744	0.06759099	192	5.38523E-06	0.05622183	0.000292822	1.823425	31.25985	57	307	0.03102885	0.1856678	ADAM8,ADORA2A,CARD11,CCL2,CD1D,CD40,CD5,CDKN1A,ENSG00000267545,ERBB2,FADD,FOXN1,FYN,GATA3,GLI2,GLI3,GRAP2,HES1,HLA-DPA1,HLA-DRA,HLA-DRB5,ICOS,ICOSLG,ID2,IGF2,IHH,IKZF1,IL12B,IL15RA,IL1RL2,IL2RA,IL4R,INHBA,INPP5D,IRF4,IRS2,LCK,LRRC32,LYN,MEF2A,NFATC2,PGLYRP3,PRKCQ,PRKCZ,PRNP,PTPN6,SART1,SHH,SOX11,SRC,SYK,TGFBR2,TNFRSF4,TNIP2,TRAF6,ZEB1,ZNF335
GO Biological Process	GO:0045619	regulation of lymphocyte differentiation	103	2.72511E-12	2.84502E-08	2.76215E-10	2.862232	20.61329	59	0.01190831	0.03408434	245	2.65619E-05	0.2773063	0.001131863	2.305777	11.70972	27	115	0.01469788	0.2347826	ADAM8,CARD11,ERBB2,FOXN1,GATA3,GLI2,GLI3,IDJ2,IHH,IKZF1,IL12B,IL15RA,IL1RL2,IL2RA,IL4R,INHBA,INPP5D,IRF4,LCK,PGLYRP3,PRKCZ,PTPN6,SART1,SHH,SYK,TGFBR2,ZEB1
GO Biological Process	GO:0002696	positive regulation of leukocyte activation	114	5.83387E-12	6.09056E-08	5.34259E-10	2.153977	45.03299	97	0.02601559	0.05603697	130	4.84326E-07	0.005056368	3.88951E-05	2.083222	23.52126	49	231	0.02667392	0.2121212	ADAM8,ADORA2B,CARD11,CCL2,CD1D,CD40,CD5,CDKN1A,ENSG00000267545,FADD,FYN,GATA3,GLI2,GLI3,GRAP2,HES1,HLA-DPA1,HLA-DRA,HLA-DRB5,ICOS,ICOSLG,IGF2,IHH,IKZF1,IL12B,IL15RA,IL1RL1,IL1RL2,IL2RA,IL4R,INPP5D,IRS2,LCK,LYN,MEF2A,NFATC2,PDPN,PRKCQ,PRKCZ,PTPN6,SART1,SHH,SPACA3,SRC,SYK,TGFBR2,TNFRSF4,TNIP2,TRAF6,ZNF335
GO Biological Process	GO:0050867	positive regulation of cell activation	128	1.64219E-11	1.71445E-07	1.33941E-09	2.103368	46.59193	98	0.0269162	0.05661467	143	7.49169E-07	0.007821327	5.46946E-05	2.037532	24.53949	50	241	0.02721829	0.2074689	ADAM8,ADORA2B,CARD11,CCL2,CD1D,CD40,CD5,CDKN1A,ENSG00000267545,FADD,FYN,GATA3,GLI2,GLI3,GRAP2,HES1,HLA-DPA1,HLA-DRA,HLA-DRB5,ICOS,ICOSLG,IGF2,IHH,IKZF1,IL12B,IL15RA,IL1RL1,IL1RL2,IL2RA,IL4R,INPP5D,IRS2,LCK,LYN,MEF2A,NFATC2,PDPN,PRKCQ,PRKCZ,PTPN6,SART1,SHH,SPACA3,SRC,SYK,TGFBR2,TNFRSF4,TNIP2,TRAF6,ZNF335
GO Biological Process	GO:0001890	placenta development	131	1.96906E-11	2.0557E-07	1.56924E-09	2.527738	26.50591	67	0.01531248	0.03870595	396	0.000585164	1	0.01542706	1.935507	13.94984	27	137	0.01469788	0.1970803	ADM,ASCL2,CDKN1C,CEBPA,CEBPB,DLX3,DNAJB6,E2F8,EGFR,EPAS1,GATA2,GJB5,HES1,HS6ST1,HSD17B2,LIF,NODAL,PTK2,RXRA,SETD2,SNAI1,SOCS3,STC2,VDR,VWF,WNT7B,ZFAT
GO Biological Process	GO:0002076	osteoblast development	144	6.44266E-11	6.72613E-07	4.67093E-09	4.802636	5.621912	27	0.003247783	0.01559792	517	0.001873489	1	0.03783216	3.618228	1.934649	7	19	0.003810561	0.3684211	GLI2,HDAC4,LRP5,MSX2,RUNX2,SHH,SMAD3
GO Biological Process	GO:0009612	response to mechanical stimulus	149	9.47862E-11	9.89568E-07	6.6414E-09	2.344463	30.71066	72	0.01774157	0.04159445	250	3.74141E-05	0.3906032	0.001562413	2.129007	14.56078	31	143	0.01687534	0.2167832	ASIC2,BDKRB1,BTG2,CCL2,CD40,CHRNA9,COL1A1,CRADD,DNAH1,ETS1,FADD,FYN,GATA4,GLI2,HPN,IGF2,INHBB,JUN,LCK,PDE2A,PKDREJ,PRKCD,PTK2,SLC12A2,SOST,SOX9,STRBP,TGFBR2,TNF,TNFRSF8,TRPV4
GO Biological Process	GO:0001503	ossification	154	1.55148E-10	1.61974E-06	1.05178E-08	2.069744	44.44995	92	0.02567877	0.05314847	362	0.000374448	1	0.010799	1.794683	20.05925	36	197	0.01959717	0.1827411	ALOX15,ATP6V1B1,COL11A2,COL13A1,COL1A1,COL2A1,CSF1,EGFR,FGF18,FGF9,FGFR3,FOXC1,GLI2,HDAC4,HSPG2,IGF2,IHH,LRP5,MEF2A,MN1,MSX2,PDLIM7,RPL38,RUNX2,RYR1,SHH,SMAD3,SMAD5,SNAI1,SOST,SOX9,TWIST2,WNT11,WWOX,ZBTB40
GO Biological Process	GO:0045580	regulation of T cell differentiation	155	1.66631E-10	1.73962E-06	1.12234E-08	2.872858	17.05618	49	0.00985337	0.02830734	322	0.000220454	1	0.007147639	2.291544	9.164126	21	90	0.01143168	0.2333333	ADAM8,CARD11,ERBB2,FOXN1,GATA3,GLI2,GLI3,IHH,IKZF1,IL12B,IL1RL2,IL2RA,IL4R,IRF4,LCK,PRKCZ,SART1,SHH,SYK,TGFBR2,ZEB1
GO Biological Process	GO:0010817	regulation of hormone levels	156	1.68424E-10	1.75835E-06	1.12715E-08	2.127877	40.41588	86	0.02334828	0.04968226	508	0.001684237	1	0.03461306	1.644224	22.50302	37	221	0.02014154	0.1674208	ACE,ADM,ALDH1A3,ANPEP,CACNA1C,CACNA1H,COMT,CPE,CYP19A1,CYP26B1,DIO3,EDN3,FFAR1,GATA3,HNF1B,HTT,IL1RN,INHBA,LTBP4,LYN,MAFA,NRSA1,PCSK6,PLEKHA1,SCARB1,SCPEP1,SHH,SMAD2,SMPD3,STARD5,SULT1B1,SULT1E1,TBX3,TPO,TRPV4,UGT1A1,WNT4
GO Biological Process	GO:1902105	regulation of leukocyte differentiation	160	2.11637E-10	2.20949E-06	1.38093E-08	2.20064	35.89865	79	0.02073868	0.04563836	207	7.31005E-06	0.07631691	0.000368681	2.056734	19.44831	40	191	0.02177463	0.2094241	ADAM8,C1QC,CARD11,CSF1,ERBB2,FOXN1,GATA2,GATA3,GLI2,GLI3,IDJ2,IHH,IKZF1,IL12B,IL15RA,IL1RL2,IL2RA,IL4R,INHBA,INPP5D,IRF4,JUN,LCK,LIF,LYN,MYC,PGLYRP3,PRDM16,PRKCZ,PTPN6,RUNX1,SART1,SHH,SYK,TGFBR2,TNF,TRAF6,ZBTB46,ZEB1,ZFPM1
GO Biological Process	GO:0061180	mammary gland epithelium development	174	7.98475E-10	8.33608E-06	4.79085E-08	2.585867	20.88274	54	0.01206398	0.03119584	241	2.30008E-05	0.2401281	0.000996382	2.897972	6.211241	18	61	0.009798585	0.295082	CCND1,CEBPB,CSF1,FOXF1,GLI2,IDJ2,LRP5,MGMT,MSX2,NTN1,PHB2,SLC12A2,SRC,TBX3,VDR,WNT4,WNT7B,ZNF703
GO Biological Process	GO:0045582	positive regulation of T cell differentiation	176	1.15687E-09	1.20777E-05	6.86235E-08	3.191204	11.90773	38	0.006879105	0.02195263	308	0.000156695	1	0.00531134	2.709215	5.90577	16	58	0.008709853	0.2758621	ADAM8,GATA3,GLI2,GLI3,IHH,IKZF1,IL12B,IL1RL2,IL2RA,IL4R,LCK,PRKCZ,SART1,SHH,SYK,TGFBR2
GO Biological Process	GO:0001649	osteoblast differentiation	181	1.39374E-09	1.45507E-05	8.03904E-08	2.598336	20.01281	52	0.01156142	0.03004044	485	0.001400021	1	0.03013653	2.196781	7.738595	17	76	0.009254219	0.2236842	COL1A1,FGF9,GLI2,HDAC4,IGF2,IHH,LRP5,MEF2A,MSX2,RUNX2,SHH,SMAD3,SMAD5,SNAI1,TWIST2,WNT11,WWOX
GO Biological Process	GO:0042476	odontogenesis	196	2.49748E-09	2.60737E-05	1.33029E-07	2.308147	27.29462	63	0.01576812	0.03639515	122	3.64521E-07	0.0038056	3.11934E-05	2.777629	10.08054	28	99	0.01524224	0.2828283	ADM,AQP6,BCL2L11,COL1A1,CSF1,DLX3,EDAR,FAM20C,FOXC1,GLI2,GLI3,HDAC1,ID3,INHBA,MSX2,OSR1,PAX9,PPARA,PVRL1,RUNX2,SCN10A,SCN5A,SDC1,SHH,SLC24A4,TBX1,TRAF6,WNT7B
GO Biological Process	GO:0030879	mammary gland development	222	6.77476E-09	7.07285E-05	3.18597E-07	2.021118	39.58206	80	0.02286659	0.04621606	150	9.67976E-07	0.01010567	6.73711E-05	2.474558	12.9316	32	127	0.01741971	0.2519685	B4GALT1,BCL2L11,CCND1,CEBPB,CSF1,ERBB2,FOXF1,GLI2,GLI3,HK2,IDJ2,IGF1R,IRS2,LRP5,MGMT,MSX2,NCOR2,NTN1,PHB2,SLC12A2,SLC29A1,SOX9,SRC,TBX3,TGFA,TGFBR2,VDR,VEGFA,WNT4,WNT7B,XDH,ZNF703
GO Biological Process	GO:0045621	positive regulation of lymphocyte differentiation	224	7.47652E-09	7.80549E-05	3.48459E-07	2.875467	13.91078	40	0.008036271	0.02310803	299	0.000136885	1	0.004779544	2.561975	7.02583	18	69	0.009798585	0.2608696	ADAM8,GATA3,GLI2,GLI3,IHH,IKZF1,IL12B,IL15RA,IL1RL2,IL2RA,IL4R,INPP5D,LCK,PRKCZ,SART1,SHH,SYK,TGFBR2
GO Biological Process	GO:0051145	smooth muscle cell differentiation	235	1.56085E-08	0.000162952	6.93414E-07	2.84144	13.72543	39	0.007929193	0.02253033	407	0.000638917	1	0.01638894	3.000832	3.66565	11	36	0.005988024	0.3055556	ADM,FOXF1,GATA6,HES1,MEF2A,NOTCH1,NTF3,TSHZ3,VEGFA,WNT4,WNT7B
GO Biological Process	GO:0035272	exocrine system development	236	1.66069E-08	0.000173377	7.34646E-07	2.881556	13.18732	38	0.007618324	0.02195263	270	7.38907E-05	0.7714186	0.002857106	3.124833	4.480239	14	44	0.007621121	0.3181818	DAG1,EDAR,EGFR,FGFR1,FOXC1,IGF1R,IGF2,LAMA1,NKX3-1,NRP1,SEMA3C,SHH,TFCP2L1,TNF
GO Biological Process	GO:0072268	pattern specification involved in metanephros development	242	2.03427E-08	0.000212378	8.77593E-07	6.082801	2.630367	16	0.001519565	0.009243212	282	0.000107181	1	0.003967992	9.820904	0.407295	4	4	0.002177463	1	FOXD1,IRX1,IRX2,OSR1
GO Biological Process	GO:0048546	digestive tract morphogenesis	249	2.7874E-08	0.000291004	1.16869E-06	2.49512	18.83677	47	0.01088202	0.02715194	93	2.46262E-08	0.000257098	2.76449E-06	3.81924	5.498476	21	54	0.01143168	0.3888889	DACT1,EGFR,EPHB3,FGFR3,FOXF1,FOXP4,GATA4,GLI2,GLI3,HNF1B,IDJ2,IHH,NODAL,NOTCH1,SFRP2,SHH,SMAD2,SMAD3,SOX11,SOX17,WNT11

# Ontology	ID	Desc	Binom Rank	Binom P	Binom Bonf P	Binom Fdr Q	Region Fold Enrich	Exp Regions	Obs Regions	Genome Frac	Set Cov	Hyper Rank	Hyper P	Hyper Bonf P	Hyper Fdr Q	Gene Fold Enrich	Exp Genes	Obs Genes	Total Genes	Gene Set Cov	Term Cov	Genes
GO Biological Process	GO:0030168	platelet activation	250	2.79939E-08	0.000292257	1.16903E-06	2.003978	37.42556	75	0.02162078	0.04332756	515	0.001771935	1	0.03592039	1.652115	21.79026	36	214	0.01959717	0.1682243	ADRA2A,ADRA2C,APBB1IP,CALM2,CD40,COL1A1,DGKZ,F2RL3,FERMT3,FYN,GNA12,GP1BB,GP9,HGF,JGF2,ITPR3,LCK,LYN,P2RX1,PDGFA,PLCG2,PRKCB,PRKCD,PRKCQ,PRKCZ,PTK2,PTPN1,RAC2,RAP1B,RHOB,SRC,SYK,TRPC7,VAV2,VEGFA,VWF
GO Biological Process	GO:0060070	canonical Wnt receptor signaling pathway	260	3.3827E-08	0.000353154	1.35828E-06	2.37481	21.4754	51	0.01240636	0.02946274	412	0.000652145	1	0.01652522	2.221395	8.553184	19	84	0.01034295	0.2261905	CCND1,DISC1,FAM105B,FZD10,GATA3,GATA4,LRP5,MYC,NKX2-5,PROP1,PTPRU,RYR2,SDC1,SHH,SOX17,WNT11,WNT4,WNT7B,WNT9A
GO Biological Process	GO:0014015	positive regulation of gliogenesis	265	4.52631E-08	0.000472547	1.7832E-06	3.164007	9.797702	31	0.00566014	0.01790872	359	0.000368309	1	0.0107107	3.177351	3.462003	11	34	0.005988024	0.3235294	CXCR4,GFAP,HES1,JD2,LIF,LYN,NKX6-2,NOTCH1,NTF3,PTK2,SHH
GO Biological Process	GO:0097084	vascular smooth muscle cell development	277	6.20386E-08	0.000647683	2.33821E-06	9.146283	1.202674	11	0.000694786	0.006354708	480	0.00135737	1	0.0295228	6.547269	0.610942	4	6	0.002177463	0.6666667	ADM,HES1,NOTCH1,VEGFA
GO Biological Process	GO:0042475	odontogenesis of dentin-containing tooth	279	6.41532E-08	0.00066976	2.40057E-06	2.397646	20.01963	48	0.01156536	0.02772964	235	1.76546E-05	0.1843144	0.000784317	2.766452	7.229477	20	71	0.01088732	0.2816901	ADM,BCL2L11,DLX3,EDAR,FAM20C,FOXC1,GLI2,GLI3,HDAC1,MSX2,PPARA,PVRL1,RUNX2,SCN10A,SCN5A,SHH,SLC24A4,TBX1,TRAF6,WNT7B
GO Biological Process	GO:0043281	regulation of cysteine-type endopeptidase activity involved in apoptotic process	283	6.88935E-08	0.000719248	2.54151E-06	2.075522	31.31743	65	0.0180921	0.03755055	297	0.000130131	1	0.004574302	1.90962	18.32825	35	180	0.0190528	0.1944444	ADORA2A,BCL2L11,BOK,CD44,CRYAA,DNAJB6,DPEP1,F3,FADD,FOXL2,HERPUD1,HGF,IFI6,KLF4,LCK,MGMT,MYC,NGFR,NKX3-1,NODAL,P2RX1,PMAIP1,PRDX5,RPS3,S100A9,SFRP2,SMAD3,SOX7,SRC,TERF1,TNF,UACA,VEGFA,WNT9A,XDH
GO Biological Process	GO:0042036	negative regulation of cytokine biosynthetic process	292	8.96935E-08	0.0009364	3.20685E-06	3.708088	6.472339	24	0.003739075	0.01386482	477	0.001320347	1	0.02889816	3.156719	2.851061	9	28	0.004899292	0.3214286	ASB1,BCL3,INHBA,INHBB,JNPPSD,KLF4,SIGIRR,TRIB2,ZFPM1
GO Biological Process	GO:0002062	chondrocyte differentiation	293	9.02226E-08	0.000941924	3.21476E-06	2.450123	18.36642	45	0.0106103	0.02599653	137	6.67068E-07	0.006964186	5.08335E-05	3.607679	4.989358	18	49	0.009798585	0.3673469	CHST11,COL11A2,COL2A1,CYTL1,FGF18,FGF9,FGFR1,FGFR3,HSPG2,MATN1,MEF2A,MSX2,OSR1,RUNX2,RUNX3,SFRP2,SOX9,SULF2
GO Biological Process	GO:0030099	myeloid cell differentiation	299	9.99248E-08	0.001043215	3.48901E-06	2.067003	30.9627	64	0.01788718	0.03697285	492	0.001462157	1	0.03102626	1.764234	17.00455	30	167	0.01633097	0.1796407	ANK1,CBFA2T3,CEBPA,CEBPG,CSF1,DMTN,EPAS1,EPO,FAM20C,GATA2,GATA3,ID2,INHBA,IRF4,IRF8,JUN,LYN,MEF2A,MMP9,PDE2A,RCOR1,RUNX1,SMAD5,SOX6,TGFBR2,TNF,TRAF6,VEGFA,ZBTB16,ZFPM1
GO Biological Process	GO:0042035	regulation of cytokine biosynthetic process	305	1.2383E-07	0.001292789	4.23865E-06	2.51288	16.71389	42	0.009655627	0.02426343	233	1.67371E-05	0.1747353	0.000749937	2.507465	9.571421	24	94	0.01306478	0.2553191	ASB1,BCL3,CARD11,CEBPB,CEBPG,GATA3,ICOSLG,IL12B,INHBA,INHBB,JNPPSD,IRF4,KLF4,LTB,MAP2K3,PRG3,PRKCQ,SIGIRR,SYK,TNF,TNFRSF8,TRAF6,TRIB2,ZFPM1
GO Biological Process	GO:1902107	positive regulation of leukocyte differentiation	306	1.30677E-07	0.001364268	4.45839E-06	2.292381	21.81138	50	0.01260045	0.02888504	223	1.13944E-05	0.1189577	0.000533443	2.455226	10.58966	26	104	0.01415351	0.25	ADAM8,CSF1,GATA3,GLI2,GLI3,JD2,JHH,JKZF1,IL12B,IL15RA,IL1RL2,IL2RA,IL4R,JNPPSD,JUN,LCK,LIF,PRKCZ,RUNX1,SART1,SHH,SYK,TGFBR2,TNF,TRAF6,ZBTB46
GO Biological Process	GO:0030857	negative regulation of epithelial cell differentiation	333	2.61714E-07	0.002732295	8.20509E-06	3.384365	7.38691	25	0.004267423	0.01444252	356	0.000363836	1	0.01066979	3.682839	2.443767	9	24	0.004899292	0.375	CCND1,DLL1,HES1,MSX2,NODAL,OSR1,TBX3,XDH,ZEB1
GO Biological Process	GO:0045687	positive regulation of glial cell differentiation	345	3.03062E-07	0.00316397	9.17093E-06	3.17527	8.503214	27	0.004912313	0.01559792	521	0.001877192	1	0.0376159	3.273635	2.443767	8	24	0.004354927	0.3333333	CXCR4,HES1,ID2,LIF,NKX6-2,NOTCH1,NTF3,SHH
GO Biological Process	GO:0001709	cell fate determination	348	3.20626E-07	0.00334733	9.61877E-06	2.5563	14.86524	38	0.008587659	0.02195263	373	0.000441771	1	0.01236484	2.946271	4.072945	12	40	0.00653239	0.3	CYP26B1,DLL1,DSCAML1,GATA2,GATA3,GATA6,HES1,JAG1,KLF4,MEF2A,NTF3,SOX17
GO Biological Process	GO:0034698	response to gonadotropin stimulus	375	6.54558E-07	0.006833589	1.82229E-05	3.669871	5.722273	21	0.003305761	0.01213172	421	0.000720465	1	0.01786617	3.399544	2.647414	9	26	0.004899292	0.3461538	EGR3,GATA4,GATA6,INHBA,MAS1,NOTCH1,PAPPA,TOX2,WT1
GO Biological Process	GO:0007440	foregut morphogenesis	413	1.31015E-06	0.01367798	3.31186E-05	4.189094	4.058157	17	0.0023444	0.009820913	183	3.86917E-06	0.04039413	0.000220733	6.547269	1.221883	8	12	0.004354927	0.6666667	FOXF1,FOXP4,GATA4,NOTCH1,SHH,SMAD2,SMAD3,SOX17
GO Biological Process	GO:0007498	mesoderm development	419	1.42074E-06	0.01483256	3.53999E-05	2.039979	26.47086	54	0.01529224	0.03119584	430	0.000802721	1	0.01948931	2.016793	11.40425	23	112	0.01252041	0.2053571	FGFR1,FOXC1,FOXF1,HTT,JKZF1,JNHBA,KLF4,LHX2,MESDC2,MESP2,NODAL,OSR1,POFUT2,RPL38,SETD2,SHH,SMAD2,SMAD3,SNAI1,TBX1,TBX3,TEAD2,VEGFA
GO Biological Process	GO:1902106	negative regulation of leukocyte differentiation	420	1.42556E-06	0.01488287	3.54354E-05	2.505502	13.96926	35	0.00807005	0.02021953	462	0.001082456	1	0.02446069	2.310801	6.924006	16	68	0.008709853	0.2352941	C1QC,GATA2,GLI3,JD2,JHH,JL4R,JNHBA,JNPPSD,LYN,MYC,PGLYRP3,PRDM16,RUNX1,SHH,ZBTB46,ZFPM1
GO Biological Process	GO:0031100	organ regeneration	425	1.52859E-06	0.01595848	3.75494E-05	2.983993	8.713158	26	0.005033598	0.01502022	422	0.000721531	1	0.0178502	2.805972	4.276592	12	42	0.00653239	0.2857143	ADM,CCL2,CCND1,CDKN1A,CEBPA,HGF,LPIN1,MKI67,NNMT,PGF,TGFBR2,UGT1A1
GO Biological Process	GO:0007422	peripheral nervous system development	445	2.44283E-06	0.02550312	5.73104E-05	2.121356	22.15564	47	0.01279933	0.02715194	416	0.000695724	1	0.01745998	2.266362	7.942243	18	78	0.009798585	0.2307692	ASCL2,ASIC2,DAG1,EGR3,ERBB2,GPC1,GSTM3,NDRG1,NFASC,NTF3,PLXNA4,POU3F1,POU3F2,RUNX1,RUNX3,S3TC2,SKI,ZFP57
GO Biological Process	GO:0001942	hair follicle development	451	2.74134E-06	0.02861957	6.3458E-05	2.174544	20.23413	44	0.01168927	0.02541883	190	5.29175E-06	0.05524591	0.000290768	2.805972	7.840419	22	77	0.01197605	0.2857143	ALX4,APCDD1,BARX2,CELSR1,DBI,EDAR,EGFR,FOXN1,HDAC1,JNHBA,MSX2,NGFR,NOTCH1,PDGFA,RUNX1,RUNX3,SHH,SNAI1,SOX18,SOX9,TGM3,TNFRSF19
GO Biological Process	GO:0050810	regulation of steroid biosynthetic process	454	2.87604E-06	0.03002582	6.61362E-05	2.692578	10.77035	29	0.006222037	0.01675332	556	0.002559576	1	0.0480611	2.455226	4.887534	12	48	0.00653239	0.25	ABCG1,APOB,DHCR7,IGF1R,IGF2,INSIG1,INSIG2,NR5A1,SNAI1,TNF,VDR,WNT4
GO Biological Process	GO:0034505	tooth mineralization	464	3.67681E-06	0.03838591	8.27283E-05	4.979439	2.610736	13	0.001508224	0.00751011	268	7.0763E-05	0.7387653	0.002756587	6.547269	0.916413	6	9	0.003266195	0.6666667	COL1A1,FAM20C,MSX2,PPARA,PVRL1,TBX1
GO Biological Process	GO:0042633	hair cycle	468	3.93021E-06	0.04103143	0.000087674	2.14302	20.53177	44	0.01186122	0.02541883	227	1.29638E-05	0.1353418	0.000596219	2.667406	8.247714	22	81	0.01197605	0.2716049	ALX4,APCDD1,BARX2,CELSR1,DBI,EDAR,EGFR,FOXN1,HDAC1,JNHBA,MSX2,NGFR,NOTCH1,PDGFA,RUNX1,RUNX3,SHH,SNAI1,SOX18,SOX9,TGM3,TNFRSF19
GO Biological Process	GO:0060708	spongiotrophoblast differentiation	469	3.98333E-06	0.04158594	8.86694E-05	11.31101	0.618866	7	0.00035752	0.004043905	282	0.000107181	1	0.003967992	9.820904	0.407295	4	4	0.002177463	1	ASCL2,GJB5,LIF,SOCS3
GO Biological Process	GO:0018101	protein citrullination	472	4.37196E-06	0.04564322	9.67017E-05	21.77555	0.229615	5	0.000132649	0.002888504	282	0.000107181	1	0.003967992	9.820904	0.407295	4	4	0.002177463	1	PADI1,PADI2,PADI3,PADI4
GO Biological Process	GO:0060251	regulation of glial cell proliferation	499	6.20706E-06	0.0648017	0.000129863	3.910718	4.09132	16	0.002363559	0.009243212	267	6.91242E-05	0.7216562	0.002702832	4.910452	1.629178	8	16	0.004354927	0.5	ASCL2,GFAP,HES1,LYN,NOTCH1,PTK2,SKI,SOX11

A.5 VISTA Enhancer human elements

VISTA Enhancer IDs:

1-7, 9-28, 30-32, 34-39, 41-44, 46-47, 49-53, 55-60, 62-66, 68-88, 90-113, 115-126, 128-129, 131-138, 140-145, 147-157, 159-178, 180-192, 194-198, 200-225, 227-236, 238-275, 277-286, 288-290, 292-314, 316-324, 326-331, 333-343, 345-359, 361-364, 366-376, 378-379, 381-386, 388-397, 399-404, 406-417, 419-445, 447-448, 450-451, 453-454, 456-457, 459-461, 463-474, 476-480, 482-484, 486-505, 507-509, 511-523, 525-530, 532-534, 536-546, 548-569, 571-590, 592-593, 595-636, 638-641, 643-649, 651-663, 665-669, 671-672, 674-684, 686-693, 695-702, 704-705, 707-708, 710-713, 715-723, 725-731, 733-734, 736-744, 746-750, 752-756, 758-760, 762-765, 767-779, 781-785, 787-813, 815-819, 821-834, 836-856, 858-861, 863-866, 868-874, 876-883, 885-887, 889-891, 893-895, 897-898, 900-902, 904-906, 908-911, 913-948, 950-954, 956-958, 960-962, 964-967, 969-975, 977-978, 980-985, 987-990, 992-993, 995-1008, 1010-1015, 1017-1022, 1024-1027, 1029-1035, 1037-1039, 1041-1043, 1045-1047, 1049-1050, 1052-1060, 1062-1064, 1066-1069, 1071-1072, 1074-1088, 1090-1095, 1097-1132, 1134-1156, 1158-1164, 1166-1170, 1172-1181, 1183-1189, 1191-1218, 1220-1237, 1239-1246, 1248-1258, 1260-1261, 1263-1264, 1266-1268, 1270-1271, 1273-1274, 1276-1278, 1280-1281, 1283-1284, 1286-1287, 1289-1290, 1292-1293, 1295-1296, 1298-1305, 1307-1312, 1314-1316, 1318-1336, 1338-1341, 1343-1346, 1348-1352, 1354-1355, 1357-1364, 1366-1371, 1373-1377, 1379-1380, 1382-1401, 1403-1404, 1406-1407, 1409-1410, 1412-1414, 1416-1419, 1421-1425, 1427-1428, 1430-1438, 1440-1445, 1447-1453, 1455-1457, 1459-1469, 1471-1475, 1477-1489, 1491-1492, 1494-1501, 1503-1510, 1512-1521, 1523-1527, 1529-1535, 1537-1546, 1548-1551, 1553-1555, 1557-1561, 1563-1568, 1570-1575, 1577-1582, 1584-1594, 1596-1597, 1599-1604, 1606-1607, 1609-1617, 1619-1620, 1622-1627, 1629-1636, 1638-1645, 1647-1653, 1655-1668, 1670-1679, 1681-1696, 1698-1699, 1701-1703, 1705-1709, 1711-1712, 1714-1715, 1717-1720, 1722-1726, 1728-1735, 1737-1738, 1740-1748, 1750-1754, 1756-1760, 1762-1764, 1766-1767, 1769-1770, 1772-1773, 1775-1776, 1778-1781, 1783-1788, 1790-1791, 1793-1794, 1796-1798, 1800-1802, 1804-1805, 1807-1811, 1813-1814, 1816-1817, 1819-1820 1822-1823, 1825-1826, 1828-1829, 1831-1834, 1836-1837, 1839-1840, 1842-1843, 1845-1846, 1848-1849, 1851-1852, 1854-1860, 1862-1863, 1865-1867, 1869-1878, 1880-1882, 1884-1887, 1889-1891, 1893-1894, 1896-1897, 1899-1900, 1902-1903, 1905-1906, 1908-1909, 1911-1917, 1919-1920, 1922-1923, 1925-1926, 1928-1929, 1931-1935, 1937-1938, 1940-1941, 1943-1945, 1947-1948, 1950-1953, 1955-1956, 1958-1959, 1961-1963, 1965-1968, 1970-1974, 1976-1977, 1979-1980, 1982-1984, 1986-1992, 1994-1995, 1997-2000, 2002-2003, 2005-2007, 2009-2010, 2012-2013, 2015-2016, 2018-2019, 2021-2023, 2025-2027, 2029-2031, 2033-2034, 2036-2038, 2040-2041, 2043-2045, 2047-2050, 2052-2054, 2056-2073, 2075-2084, 2086-2092, 2094-2098, 2100-2101, 2103-2104, 2106-2107, 2109-2111, 2113-2114, 2116-2119, 2121-2122, 2124-2125, 2127-2128, 2130-2132, 2134-2135, 2137-2139, 2141-2142, 2144-2146, 2148-2149, 2151-2153, 2155-2156, 2158-2160, 2162-2163, 2165-2166, 2168-2169, 2171-2172, 2174-2176, 2178-2179, 2181-2182, 2184-2185, 2187-2189, 2191-2192, 2194-2195, 2197-2198, 2200-2201, 2203-2204, 2206-2207, 2209-2210, 2212-2213, 2215-2216, 2218-2219, 2221-2223, 2225-2227, 2229-2232, 2234-2235, 2237-2238, 2240-2241, 2243-2244, 2246-2249, 2251-2252, 2254-2256, 2258-2259, 2261-2263, 2265-2266, 2268-2269, 2271-2272, 2274-2275, 2277-2278, 2280-2281, 2283-2284, 2286-2287, 2289-2290, 2292-2293, 2295-2298, 2300-2301, 2303-2324, 2326-2327, 2329-2330, 2332-2333, 2335-2336, 2338-2339, 2341-2342, 2344-2346, 2348-2349, 2351-2353, 2355-2357, 2359-2360, 2362-2363, 2365-2366, 2368-2369, 2371-2375, 2377-2378, 2380-2382, 2384-2385, 2387-2388, 2390-2391, 2393-2394, 2396-2397, 2399-2400, 2402-2405, 2407-2408, 2410-2412, 2414-2415, 2417-2420

Roadmap chromHMM map IDs:
E017, E019, E020, E021, E022, E026, E029, E032, E034, E037, E038, E039, E040, E041, E042, E043, E044, E045, E046, E047, E048, E049, E050, E055, E056, E058, E059, E061, E062, E063, E065, E066, E067, E068, E069, E071, E072, E073, E074, E075, E076, E078, E079, E080, E084, E085, E087, E089, E090, E091, E092, E093, E094, E095, E096, E097, E098, E099, E100, E101, E102, E103, E104, E105, E106, E108, E109, E111, E112, E113, E119, E120, E121, E122, E124, E125, E126, E127, E128, E129

Appendix B: Chapter 3 extended data

B.1 Primary and metastatic sample gene mutations

B.1.1 Brain metastasis mutated genes

gene	expr (expression level of this gene, averaged across 91 cell lines)	reptime (DNA replication time of this gene)	hic (chromatin state of this gene)	N_nonsilent (total number of nonsilent mutations)	N_silent (total number of silent mutations)	N_noncoding (total noncoding mutations)	n_nonsilent (Number of nonsilent coding mutations)	n_silent (Number of silent coding mutations)	n_noncoding	nnei (number of neighboring genes)	x (number of mutated bases in the neighboring genes)	X (total number of bases related to the neighboring genes)	pvalue	qvalue	Comment
TP53	2069567	213	34	101937	28149	176913	22	1	0	50	30	9201654	4.44E-16	0	
STK11	2214581	234	29	103818	28644	91971	10	0	0	37	16	5640459	3.45E-13	0	
POLDIP2	1405196	214	10	88110	25344	158895	5	0	0	30	17	5941947	1.01E-07	0	
CRIPAK	1018381	507	41	99660	33099	22671	4	0	0	50	47	15137595	3.19E-07	0	CRIPAK was disregarded as artifact after manual verification
KRAS	259193	512	16	56562	13629	88110	9	0	0	44	26	8442951	1.21E-06	0	
KEAP1	1996822	207	30	145530	42075	61083	8	0	0	50	34	9625572	5.33E-06	0.02	
DCDC1	84163	938	-25	85371	22836	153450	10	2	1	44	73	9966165	1.04E-05	0.03	
KRT76	1444461	422	38	150150	42999	193842	5	1	2	28	35	6582444	9.06E-05	0.21	
CDKN2A	225405	357	-15	78705	20493	44847	3	0	0	50	27	13114563	1.07E-04	0.22	
LPHN3	11170	1494	-18	348744	96558	422037	12	3	1	3	6	1090089	1.62E-04	0.3	
TPRX1	881316	336	39	88968	33990	11286	3	0	0	50	29	11897127	1.75E-04	0.3	
MUC20	907437	365	47	119988	41679	13563	8	0	0	50	57	14690511	2.50E-04	0.38	
COL22A1	5667	846	-22	385044	123519	1283337	11	3	11	16	26	3455661	2.66E-04	0.38	
C1orf173	101813	1070	-14	361614	98637	252648	10	1	0	6	8	2287560	2.99E-04	0.38	
SFXN2	878626	180	27	79068	21219	276111	3	0	0	29	20	7146381	3.05E-04	0.38	
NCR1	840563	573	-15	72897	20262	130581	3	0	1	50	46	10922010	3.37E-04	0.4	
LRP1B	8988	1120	-11	1131009	270831	1726560	26	7	6	3	16	2385669	4.00E-04	0.4	
DIO2	47674	533	-44	72105	21351	64152	3	0	1	50	42	8684907	4.25E-04	0.4	
LAMA2	67993	771	-6	751608	201267	1471833	12	4	3	31	25	8657022	4.29E-04	0.4	
CLRN2	271399	912	37	54087	16302	46332	3	0	0	50	31	10746450	4.30E-04	0.4	
VAV3	132139	554	-11	213015	51810	456588	5	0	2	9	10	2658645	4.46E-04	0.4	
COL4A1	429137	434	12	391314	124872	955152	9	2	2	50	59	16186797	4.64E-04	0.4	
GIMAP5	734961	689	14	72006	20262	48213	3	0	0	50	22	8629203	5.56E-04	0.42	
SULT1C4	361799	483	-2	75240	17127	152559	6	0	0	13	4	3181398	5.68E-04	0.42	
TEX13B	184382	NaN	40	72303	21450	35739	4	0	0	50	38	10956627	5.71E-04	0.42	
KLF7	207002	598	20	70719	20460	106425	2	0	0	42	25	7590957	5.83E-04	0.42	
HRCT1	853975	298	55	25674	8976	11682	2	0	0	50	29	8901090	6.33E-04	0.44	
IL7R	139461	511	5	110616	28974	182556	6	1	2	41	26	6580992	6.72E-04	0.44	
CHD6	378513	502	-1	647196	173712	851994	7	0	0	3	0	1307889	6.81E-04	0.44	
ITM2A	13949	NaN	29	62964	17424	53757	4	0	0	38	41	8862513	6.92E-04	0.44	
CRP	737948	602	-2	52404	14817	39303	3	0	0	50	50	13003023	7.90E-04	0.48	
MESP2	934542	212	37	88077	30525	21285	2	0	0	50	29	9567459	8.53E-04	0.5	
NRN1L	879834	158	50	37488	12804	118602	2	0	0	50	32	11861454	8.94E-04	0.5	
OR6K6	103627	985	1	78936	23232	12474	7	0	0	50	52	10599435	9.01E-04	0.5	
OR10K2	135429	902	5	70488	22275	6831	5	0	0	50	31	7594026	9.61E-04	0.52	
SORCS1	6116	881	-26	296142	84117	491238	8	1	9	31	54	5732859	9.99E-04	0.52	
SEMA5A	45422	891	-2	257730	69861	403128	10	0	5	21	28	4595943	1.03E-03	0.52	
IRF9	974559	220	41	93720	26367	204732	3	0	1	50	34	12144396	1.11E-03	0.55	
C1orf61	2199280	201	27	37785	10824	93951	2	0	1	50	32	10436811	1.16E-03	0.55	
KNTC1	917926	215	37	543147	138171	693297	5	1	2	50	30	9229176	1.17E-03	0.55	
GPR115	176427	691	7	165033	45243	183348	5	0	1	50	44	12885279	1.20E-03	0.55	
UNC93A	254009	704	1	106359	32439	181269	7	1	2	8	12	2023527	1.22E-03	0.55	
PHF2	421433	612	29	261129	72996	301356	3	2	1	35	64	10420773	1.25E-03	0.55	
TXNIP	739922	172	39	93489	25707	118008	3	0	0	50	30	11064273	1.38E-03	0.57	
MTF1	445066	303	33	176253	51645	213147	3	0	0	29	10	5858226	1.39E-03	0.57	
C7orf72	NaN	NaN	NaN	105072	28479	0	4	0	0	50	87	22658328	1.40E-03	0.57	
CHST11	1030269	419	34	83094	22539	114147	2	0	1	50	41	7994547	1.44E-03	0.57	
RHPN2	458398	356	-12	164043	45738	279576	4	0	1	50	37	14440536	1.44E-03	0.57	
ZFHX4	72333	731	8	843843	234366	71280	22	6	1	13	24	2377551	1.49E-03	0.57	
OR6N1	103627	985	1	70752	22110	4059	3	0	0	50	52	10599435	1.50E-03	0.57	

gene	expr (expression level of this gene, averaged across 91 cell lines)	reptime (DNA replication time of this gene)	hic (chromatin state of this gene)	N_nonsilent (total number of nonsilent mutations)	N_silent (total number of silent mutations)	N_noncoding (total noncoding mutations)	n_nonsilent (Number of nonsilent coding mutations)	n_silent (Number of silent coding mutations)	n_noncoding	nnei (number of neighboring genes)	x (number of mutated bases in the neighboring genes)	X (total number of bases related to the neighboring genes)	pvalue	qvalue	Comment
ALMS1	777501	590	22	968022	278586	465498	21	1	2	9	16	3242976	1.60E-03	0.59	
SH3BGRL	137219	NaN	-22	28182	7161	49500	3	0	1	17	18	3295248	1.66E-03	0.6	
NFE2L2	372300	145	36	143649	37917	65934	4	0	0	50	32	11496144	1.69E-03	0.6	
GNRH2	835590	288	54	28215	9108	35541	2	0	0	50	26	8657979	1.76E-03	0.6	
TLR4	58763	944	-8	196284	53988	52569	8	0	0	50	65	10860762	1.78E-03	0.6	
ZIC4	20725	1070	-30	97878	28347	77220	9	1	0	24	42	4662735	1.79E-03	0.6	
KCNQ5	648013	975	15	222651	65637	554004	3	1	2	27	11	6518589	1.95E-03	0.64	
GFRAL	28347	729	-14	98076	22605	165528	3	0	0	8	9	2467872	1.98E-03	0.64	
INF2	651235	241	21	290466	92664	222750	3	1	0	18	10	3657852	1.99E-03	0.64	
GAL	780720	354	49	29832	8778	58212	4	0	1	18	23	4740087	2.09E-03	0.66	
STC2	471196	519	41	71115	20064	86724	3	0	0	29	29	6527697	2.17E-03	0.67	
CELF4	55294	588	-11	118536	30855	134145	2	0	1	50	56	12499773	2.21E-03	0.67	
TGIF2LX	8315	NaN	-35	56331	16137	13464	3	1	0	22	32	2695638	2.24E-03	0.67	
LIN28A	1046693	181	33	49962	13596	61578	2	0	0	50	27	8594520	2.29E-03	0.68	
SNURF	233460	1194	-44	17886	4884	194238	2	0	2	9	24	2742135	2.34E-03	0.68	
OR7D4	398878	582	-50	70587	22374	7326	3	0	0	50	41	11579634	2.57E-03	0.73	
GUCA1C	90728	971	-15	51546	12210	75636	3	0	0	50	61	11311773	2.70E-03	0.74	
COL15A1	199434	148	18	327162	101607	769131	7	4	5	14	14	2746425	2.77E-03	0.74	
FGB	59883	1034	20	119889	29205	127116	4	0	0	27	25	5820969	2.78E-03	0.74	
ATP8A2	331788	648	20	289410	77979	664884	5	0	2	2	4	1333464	2.84E-03	0.74	
PIP5K1A	1409327	184	32	135564	37587	388575	3	0	2	50	33	10648506	2.86E-03	0.74	
POLM	1151124	282	56	115170	36003	137115	3	0	0	15	7	3185127	2.88E-03	0.74	
AMPH	167080	639	7	167937	46893	435996	4	0	4	50	43	12509310	2.91E-03	0.74	
FUT9	76115	1328	-21	85338	21978	24651	5	0	0	25	30	4022403	2.92E-03	0.74	
WBP2	1727802	178	35	63723	16863	112365	2	0	1	50	38	12263097	3.05E-03	0.76	
LRRC69	NaN	NaN	NaN	84315	22011	153549	3	0	1	50	88	22805409	3.06E-03	0.76	
IGFBP5	573046	620	16	64020	18249	53658	2	0	0	50	35	10994082	3.32E-03	0.81	
SIRPA	467799	486	41	117843	35310	145629	3	1	0	38	31	8564226	3.39E-03	0.81	
RNF144A	40925	1103	7	71148	18645	151569	4	2	2	11	13	1404249	3.40E-03	0.81	
SLAIN1	131557	900	38	83127	23199	105930	2	0	0	50	29	12097932	3.49E-03	0.82	
CD79B	1455216	404	25	56067	14718	93654	2	0	0	18	13	4492521	3.57E-03	0.82	
ZNF415	392290	946	-49	135366	33033	79101	3	1	0	47	24	5222646	3.59E-03	0.82	
TMPRSS13	527704	576	45	133749	39897	160083	3	0	0	10	6	2152392	3.60E-03	0.82	
KCNMB2	144490	882	-8	56595	15081	83655	3	0	0	50	96	15868611	3.71E-03	0.83	
MAGEC2	16386	NaN	-29	86460	25014	35541	2	1	0	50	86	9737211	3.79E-03	0.84	
ZNF257	75568	1305	-69	135960	32736	95832	3	0	0	12	8	2350854	3.83E-03	0.84	
OR1B1	407749	646	9	70983	23661	9504	5	1	0	15	12	1629078	3.86E-03	0.84	
FAM184B	NaN	NaN	NaN	254067	67782	0	3	0	0	50	87	22697631	3.95E-03	0.84	
PALM	1914670	200	23	90717	27687	84249	3	0	0	50	32	10735857	3.97E-03	0.84	
FAM46D	102816	NaN	-24	92268	23958	5247	4	0	0	50	126	13383579	4.07E-03	0.84	
NFASC	388939	440	33	328647	91707	600336	9	1	5	50	51	9683520	4.07E-03	0.84	
LCLAT1	140581	648	21	100485	25146	167706	2	0	0	8	4	1740948	4.24E-03	0.85	
KCNU1	7048	474	-36	277035	75009	421146	7	2	5	45	49	6167601	4.29E-03	0.85	
MRGPRX4	631291	436	20	72369	23760	23859	3	0	0	50	48	13521288	4.31E-03	0.85	
SLC6A19	657890	541	24	149556	43593	192555	5	0	0	12	14	3740154	4.42E-03	0.85	
SLC3A2	2342595	181	71	154671	46992	370260	2	1	0	19	23	5345835	4.45E-03	0.85	
ZNF536	106911	926	-73	300531	87450	120582	10	2	1	44	53	7080084	4.45E-03	0.85	
APEX2	245504	NaN	8	120549	35772	44154	3	0	0	50	46	14889369	4.46E-03	0.85	
FAM175B	290823	200	36	101343	25575	173745	3	0	1	50	30	8787768	4.49E-03	0.85	
COL1A2	461036	499	-14	318450	107745	872685	6	0	6	21	30	5540007	4.50E-03	0.85	
MUC1	1801444	194	32	74019	13002	128205	2	0	0	50	39	11328636	4.54E-03	0.85	
RAD21L1	NaN	NaN	NaN	138336	32439	0	7	0	0	50	87	22662288	4.61E-03	0.85	
CNN1	1152381	272	31	71544	19338	93852	2	0	0	50	33	9775986	4.68E-03	0.85	
OR5L1	4183	1091	-36	70785	21879	12177	3	0	0	50	54	6787770	4.74E-03	0.85	
TNR	177437	738	4	320793	91146	481239	7	0	1	20	23	5821101	4.75E-03	0.85	
LAPTM4B	609859	175	39	73986	22836	157509	2	0	0	50	38	11102421	4.92E-03	0.87	
SLC7A13	192785	1147	3	108009	33264	65736	4	0	2	3	6	988746	4.94E-03	0.87	
GDAP1L1	664405	367	19	86823	24651	75735	2	0	0	50	39	13113507	5.04E-03	0.88	
MYO19	416822	556	23	236346	66891	321453	4	1	1	50	32	12863862	5.08E-03	0.88	
CLIC6	284989	406	43	157905	48114	94545	3	0	1	40	32	8920230	5.14E-03	0.88	
CENPM	1335011	323	67	46134	13761	68310	2	0	1	50	40	8363289	5.17E-03	0.88	
FAM47A	7968	NaN	-15	179718	55704	12771	7	0	0	50	94	11714142	5.23E-03	0.88	
TBC1D12	231059	361	22	182325	52899	187308	2	1	0	50	49	11670879	5.35E-03	0.89	
ST8SIA3	298355	599	9	90057	24288	72171	5	0	0	50	46	11077506	5.41E-03	0.89	
OR51G2	65798	1006	-15	70917	22638	5445	5	0	0	50	48	10039887	5.48E-03	0.89	
ADAM19	261911	1079	45	220440	61413	366102	6	0	1	19	9	3373788	5.48E-03	0.89	
MGAM	196155	846	-7	450318	120120	1233738	5	4	10	29	56	8373123	5.51E-03	0.89	
NOP10	542657	234	22	15675	4224	63756	1	0	0	50	31	8773809	5.57E-03	0.89	
SIGLEC8	393294	583	17	114774	36300	139689	2	0	0	22	18	9047379	5.61E-03	0.89	
FHL5	116593	1376	3	71115	15510	97713	5	1	0	50	44	8793378	5.84E-03	0.89	
BNC1	206688	679	11	233871	63426	72270	2	1	0	50	36	10558581	5.92E-03	0.89	
PODXL	236261	220	27	128337	40854	149292	3	1	0	50	27	11022561	5.94E-03	0.89	
TBC1D10C	1269888	177	51	102465	33858	80784	2	0	0	50	26	10275408	5.97E-03	0.89	
PIK3CG	171312	700	32	260271	71280	235323	7	2	0	50	51	13645566	6.02E-03	0.89	
ZNF534	548808	989	-53	161634	40029	62271	5	1	1	44	21	5469750	6.04E-03	0.89	
KCNH8	25033	1018	-14	262713	72303	336204	7	1	1	14	8	2611851	6.07E-03	0.89	
PPY	920082	241	29	22242	7260	14949	2	0	0	50	26	9559506	6.10E-03	0.89	
EEF1G	2695768	217	42	106029	27819	167211	2	0	0	37	20	9070182	6.11E-03	0.89	
RGS21	123814	860	-26	38247	8778	74151	2	1	0	50	90	10697973	6.11E-03	0.89	
PIP4K2B	1198833	207	37	101475	25938	258984	2	0	0	19	3	2802789	6.25E-03	0.91	
GMPR	117199	561	26	82599	23331	161766	3	0	0	10	5	2527404	6.37E-03	0.91	

gene	expr (expression level of this gene, averaged across 91 cell lines)	reptime (DNA replication time of this gene)	hic (chromatin state of this gene)	N_nonsilent (total number of nonsilent mutations)	N_silent (total number of silent mutations)	N_noncoding (total noncoding mutations)	n_nonsilent (Number of nonsilent coding mutations)	n_silent (Number of silent coding mutations)	n_noncoding	nnei (number of neighboring genes)	x (number of mutated bases in the neighboring genes)	X (total number of bases related to the neighboring genes)	pvalue	qvalue	Comment
CDX4	221528	NaN	-13	66231	19107	34155	2	0	0	50	37	8470572	6.38E-03	0.91	
ZNF384	2816983	210	61	136653	38577	173448	2	0	0	46	22	11393613	6.46E-03	0.92	
SERPINB5	200746	166	24	90552	23496	175329	4	0	0	41	35	10112190	6.51E-03	0.92	
ST3GAL4	349521	514	31	80157	21813	175329	2	0	2	35	34	9211455	6.67E-03	0.93	
S100A2	1454773	186	26	23991	5907	50886	1	0	0	50	28	10766151	6.74E-03	0.93	
CDH23	641879	524	41	818994	245553	843876	12	1	7	25	32	6113250	6.75E-03	0.93	
ANAPC13	254024	188	16	18678	4389	60489	1	0	0	50	27	7978542	6.93E-03	0.94	
RBM5	965735	196	35	200904	50952	416592	4	2	1	50	39	10243002	6.95E-03	0.94	
RNF133	53042	1027	6	87747	24420	27720	3	0	0	50	53	9021177	7.01E-03	0.94	
MEGF10	245811	814	6	277629	70752	484407	3	2	2	50	43	10270029	7.01E-03	0.94	
HAVCR1	138350	1207	11	85206	26367	162558	4	1	2	21	23	3379596	7.07E-03	0.94	
CCT6B	268361	579	-14	127644	35409	252450	3	0	0	34	21	6749787	7.14E-03	0.94	
FST	202521	488	-5	83721	21516	98604	3	1	1	43	36	7471662	7.16E-03	0.94	
FAM171B	264318	538	-9	193578	54813	137808	4	1	0	26	21	6067611	7.22E-03	0.94	
NF1	462449	396	26	688017	183876	1185723	10	2	3	6	7	2128269	7.26E-03	0.94	
SCRG1	249791	553	8	24123	6072	50391	2	0	0	50	34	12931809	7.30E-03	0.94	
KRTAP5-5	NaN	NaN	NaN	54780	16104	32769	3	1	0	2	2	491601	7.39E-03	0.94	
ATP5L	963668	192	48	24486	7194	33561	1	0	0	50	35	11066352	7.53E-03	0.95	
OR51V1	109545	1056	-1	73656	21879	11979	3	0	0	50	38	7221984	7.74E-03	0.96	
SCGB2A1	2778889	450	53	23991	5511	78309	2	0	0	50	25	9467535	7.78E-03	0.96	
CKS1B	1612870	224	35	20097	4455	60786	1	0	0	50	38	9994083	7.78E-03	0.96	
ZRANB2	107389	955	-15	84282	21747	207999	3	0	1	32	48	7999134	7.80E-03	0.96	
SPTA1	177547	940	1	590337	148599	1189980	13	5	7	11	20	3881031	7.93E-03	0.97	
KRTAP19-8	200531	972	-45	15312	4092	22770	1	0	0	50	32	6406950	7.99E-03	0.97	
CPM	645106	228	28	107481	27555	226413	2	2	0	50	32	10910790	8.04E-03	0.97	
ARSF	217524	NaN	-11	140085	39402	165627	4	0	1	50	50	11943228	8.10E-03	0.97	
OR10X1	179404	940	1	73689	22539	12969	4	0	0	50	62	12139809	8.23E-03	0.98	
ZNF334	416360	642	19	165132	38709	95139	3	0	1	50	39	9913299	8.35E-03	0.98	
HOXD9	242935	874	15	79200	26235	34254	3	1	0	34	22	4835688	8.36E-03	0.98	
GAS2L3	197479	849	-12	162195	47586	288981	2	1	0	11	14	3506712	8.44E-03	0.98	
ZDHHC13	194027	625	11	152394	38973	165627	3	0	1	50	41	10928874	8.44E-03	0.98	
COX16	500896	549	32	26400	6567	63360	1	0	0	50	36	10954845	8.47E-03	0.98	
SLC26A6	1386275	212	37	181566	55836	299079	4	0	2	42	43	12426117	8.54E-03	0.98	
COBL	133439	509	-22	290334	89232	290961	6	0	3	22	32	6026889	8.67E-03	0.98	
SLC26A7	79500	987	-11	163944	43758	381645	4	1	0	8	2	1583802	8.71E-03	0.98	
CENPW	142596	704	-1	21054	6171	55539	1	0	0	50	52	11385297	8.72E-03	0.98	
ADCY8	45406	771	-7	292215	86559	352737	4	0	2	29	21	8051472	8.74E-03	0.98	
B3GALT1	67289	1010	-14	76659	20460	23067	2	0	0	50	48	10784301	8.76E-03	0.98	
LAMA4	131437	282	-7	452463	120549	826452	7	0	3	8	9	2497242	8.80E-03	0.98	
FAM13C	121813	707	-15	142263	37125	292842	3	1	3	11	20	2976897	8.98E-03	0.98	
TMEM156	435367	152	49	72303	18282	100287	2	0	0	50	35	10239999	9.03E-03	0.98	
EPHA3	27204	NaN	-18	237534	63426	338184	4	0	1	12	16	3019401	9.09E-03	0.98	
TMEM37	203559	497	35	43395	13926	23166	1	0	0	50	20	11436480	9.12E-03	0.98	
C2orf78	1124858	404	43	210804	64119	33264	4	1	0	39	32	7552644	9.15E-03	0.98	
FANCD2	885389	170	41	366861	95172	949806	3	1	0	2	2	1346994	9.15E-03	0.98	
C9orf153	311398	417	8	25377	6303	59202	1	0	0	50	40	10680021	9.24E-03	0.99	
LAD1	857879	209	26	120516	36894	187308	2	0	0	10	1	1898622	9.41E-03	0.99	
PSENEN	1151198	321	30	24783	6699	148797	1	0	0	50	30	8955078	9.48E-03	0.99	
RAB3B	701660	353	22	52569	14157	113157	2	0	0	2	1	824406	9.50E-03	0.99	
ACTRT1	938	NaN	-35	86460	25509	21384	6	0	0	50	83	8353290	9.51E-03	0.99	
ADAMTS2	1496298	677	-4	291852	83853	304821	6	4	3	7	9	1297395	9.66E-03	1	
GLT8D2	1193959	281	37	85767	21747	256608	2	0	2	50	39	9638079	9.74E-03	1	
OR2A2	221024	812	-11	72369	22176	14949	2	0	0	50	42	8058996	9.83E-03	1	
TLX1NB	469702	221	52	27423	9108	7821	2	0	0	50	32	10323918	1.00E-02	1	
MAGEB10	10745	NaN	16	80718	23034	21483	5	0	0	50	55	8103513	1.01E-02	1	
DEFB127	215351	430	25	23793	6303	44253	2	0	0	50	31	10671177	1.02E-02	1	
QRFP	727815	241	22	30360	10230	9603	2	0	0	50	28	9614484	1.03E-02	1	
ANKRD13C	295772	396	-1	130845	34881	220275	2	1	0	50	41	13789281	1.05E-02	1	
NDUFC2	556958	209	38	28347	8085	43956	2	0	0	50	36	10613262	1.06E-02	1	
ATP8B3	2180359	206	29	313863	87978	150480	4	0	0	37	17	6073485	1.06E-02	1	
POLR2F	1370748	203	50	31449	8151	153252	1	0	0	50	25	8988144	1.07E-02	1	
C6orf223	1407986	199	47	61644	16170	31383	2	0	0	50	31	11058795	1.07E-02	1	
CAMKK1	505430	215	30	130977	37125	262845	2	0	0	24	8	4403586	1.07E-02	1	
B2M	673122	428	36	29238	7788	67023	1	0	0	50	41	10412391	1.07E-02	1	
C2orf76	193588	467	26	32142	7755	155133	1	0	1	50	34	10343784	1.07E-02	1	
ZNF491	739660	633	-2	105105	25377	42768	2	0	0	50	52	13730079	1.08E-02	1	
PFN4	269570	385	38	31548	8646	87912	1	1	0	33	22	7447275	1.09E-02	1	
ATRNL1	45264	1002	-20	335940	85008	435897	9	0	0	4	3	1837440	1.09E-02	1	
GP2	243266	848	-5	129162	34980	190773	4	2	1	31	53	7839117	1.09E-02	1	
FAM124B	310772	460	12	111738	33297	19503	2	0	0	50	31	9800868	1.10E-02	1	
FABP2	218603	824	-1	32835	8052	75834	1	0	1	45	30	9478524	1.10E-02	1	
PTPN18	369510	593	29	110187	32472	218988	2	1	1	50	43	11699523	1.10E-02	1	
KIAA0319	417813	373	19	256707	72864	507771	5	1	1	19	18	5733915	1.12E-02	1	
CCL15	779228	525	19	27819	7227	57618	1	0	0	30	11	4639965	1.13E-02	1	
ECM2	579226	474	38	167706	43758	174933	3	2	0	50	37	10618575	1.14E-02	1	
ZNF727	NaN	NaN	NaN	121077	28413	594	2	0	0	50	87	22658856	1.15E-02	1	
TFF2	313101	464	90	31680	8118	69003	1	0	0	50	28	8081469	1.15E-02	1	
SERPINA11	278548	858	13	97911	29304	86130	4	0	0	50	27	7025865	1.16E-02	1	
ZNF443	1754294	593	0	161865	38907	76527	5	0	0	50	58	12144462	1.16E-02	1	
EEFSEC	858998	410	48	137610	42273	136125	2	0	0	28	14	8233434	1.16E-02	1	
PLCL1	627125	907	10	258093	69399	96327	5	0	0	16	12	5104935	1.16E-02	1	

gene	expr (expression level of this gene, averaged across 91 cell lines)	reptime (DNA replication time of this gene)	hic (chromatin state of this gene)	N_nonsilent (total number of nonsilent mutations)	N_silent (total number of silent mutations)	N_noncoding (total noncoding mutations)	n_nonsilent (Number of nonsilent coding mutations)	n_silent (Number of silent coding mutations)	n_noncoding	nnei (number of neighboring genes)	x (number of mutated bases in the neighboring genes)	X (total number of bases related to the neighboring genes)	pvalue	qvalue	Comment
H3F3C	271605	394	20	30459	10131	25344	2	0	0	50	31	10671441	1.16E-02	1	
COL5A2	237205	422	1	351912	114576	864864	7	1	4	5	8	1778205	1.18E-02	1	
C2orf48	881918	220	41	37917	10791	64350	1	0	0	50	31	10441761	1.19E-02	1	
COL27A1	224550	425	3	433158	143220	1139094	6	1	9	50	39	11935473	1.19E-02	1	
IQCF1	1013177	370	39	48840	13926	79695	2	0	1	35	27	7315143	1.19E-02	1	
RNF128	149556	NaN	17	131934	38247	147807	2	0	1	28	12	5230665	1.20E-02	1	
C12orf61	299026	359	3	28413	10989	12474	1	0	0	50	60	12404139	1.22E-02	1	
CELSR2	994135	287	31	668646	213048	475002	8	2	1	50	32	10215216	1.22E-02	1	
E2F4	1109491	150	46	98109	28611	162063	2	0	0	30	24	6296763	1.22E-02	1	
ABCB6	773232	177	36	196251	61248	318582	3	0	0	4	1	1466784	1.23E-02	1	
BMPR1B	119865	642	-15	121572	31779	180873	4	0	0	20	10	4288020	1.23E-02	1	
KRT1	1441383	464	29	150084	44847	161865	2	0	0	7	3	1380192	1.24E-02	1	
CLLU1OS	168575	582	19	24981	6699	86229	2	0	0	38	30	7775757	1.24E-02	1	
TPP2	195408	521	20	301521	80916	487872	2	0	1	26	18	6147933	1.24E-02	1	
DDX60L	134820	238	-11	417846	103389	332640	7	1	0	26	18	5271057	1.25E-02	1	
CARKD	504689	278	33	90816	29073	160578	4	0	0	7	4	1951752	1.29E-02	1	
GMNN	422850	362	40	51678	13068	91476	1	0	0	50	36	14339622	1.29E-02	1	
FYB	171734	367	-6	200772	52866	283140	4	0	3	50	53	11048004	1.30E-02	1	
GALK2	315919	528	8	111837	30426	271458	2	0	1	19	15	4903239	1.30E-02	1	
TTR	294098	541	20	34419	10725	78804	1	0	0	50	47	10751235	1.31E-02	1	
SH2D1B	363646	391	1	32505	8184	110880	2	0	0	50	27	11091795	1.32E-02	1	
C7orf69	NaN	NaN	NaN	30228	7293	188694	2	0	0	50	87	22825836	1.32E-02	1	
CD3G	963668	192	48	45738	11187	150876	3	1	0	50	35	11066352	1.32E-02	1	
METAP1	218494	549	21	94182	24717	125235	2	1	1	50	38	8595873	1.33E-02	1	
CRHR2	484075	433	32	99561	27159	163647	3	0	2	42	48	9545910	1.35E-02	1	
VAMP4	415434	718	3	36828	8118	129492	1	0	1	50	42	10240659	1.35E-02	1	
FAM169B	149919	473	-53	46827	12474	86130	3	1	0	50	36	7778199	1.36E-02	1	
ARID3C	731413	298	51	94149	30789	94347	4	1	0	21	13	3686298	1.37E-02	1	
SLC39A8	275900	441	4	116193	33099	135432	2	0	0	17	5	2663199	1.37E-02	1	
ABCB1	287293	822	0	305943	85206	675081	6	1	5	42	42	9502053	1.38E-02	1	
FLVCR1	326908	445	13	128634	40062	174834	2	1	0	50	40	11324742	1.39E-02	1	
TMEM159	338008	549	39	38016	11286	110979	3	0	0	50	38	11031306	1.40E-02	1	
CTNNA2	10595	1614	-24	199650	55572	430155	8	4	8	23	24	2127213	1.40E-02	1	
THY1	1054477	263	40	37719	11583	59301	1	0	0	50	43	10144332	1.41E-02	1	
C7	211302	644	5	204567	52833	222156	7	1	1	21	12	3707715	1.44E-02	1	
TYMP	1075525	207	42	109989	37818	94347	2	1	0	39	25	6406059	1.45E-02	1	
CNTN6	6599	831	-23	247830	66495	393426	4	1	4	24	42	5500242	1.46E-02	1	
OIT3	635147	291	38	128964	36366	303039	2	1	0	9	5	2878194	1.46E-02	1	
SYN3	259538	520	-72	137742	39567	369072	3	0	1	5	1	809490	1.46E-02	1	
FGF22	1832767	218	10	37917	13563	4455	1	0	0	50	44	10190697	1.46E-02	1	
AZIN1	689560	171	54	109989	27324	200574	2	0	0	37	25	6620658	1.46E-02	1	
EFHD2	554165	212	29	57255	15708	31383	1	0	0	50	30	11749188	1.46E-02	1	
LZIC	757819	183	29	47553	11550	149292	1	0	0	30	17	6973032	1.47E-02	1	
ZNF408	965022	176	48	165990	52899	91971	2	0	0	50	24	9483243	1.47E-02	1	
LBP	563754	350	47	115731	33165	332937	2	0	2	50	52	12004971	1.48E-02	1	
PAEP	327501	539	37	44913	11616	102861	1	0	1	41	32	9455820	1.48E-02	1	
MCF2	68546	NaN	-2	256773	59928	462627	3	1	3	19	30	5290362	1.50E-02	1	
OR6S1	709113	577	11	74118	24486	13266	2	0	0	50	37	10965438	1.50E-02	1	
SPNS1	1189140	257	55	120516	41349	213840	2	0	1	8	3	2078934	1.50E-02	1	
TMEM216	1394229	286	57	35673	10560	57915	1	0	0	50	39	8492418	1.50E-02	1	
PABPC1	1141833	276	41	153021	41613	245718	3	1	0	23	18	4625016	1.51E-02	1	
RBM10	819569	NaN	38	223476	61743	205821	4	0	0	9	6	1745865	1.53E-02	1	
OR2L8	27938	1015	-28	71544	21417	11484	3	0	0	50	115	10001772	1.54E-02	1	
H1F0	1385064	203	66	44484	13431	8613	1	0	0	50	32	8970786	1.55E-02	1	
LRRC36	1158788	125	49	180708	51447	282645	3	1	1	50	21	8889639	1.58E-02	1	
DTX1	562320	387	44	141867	45738	66627	2	0	1	50	42	11314149	1.58E-02	1	
TMEM116	816439	317	43	59532	16302	253044	1	0	1	50	26	10836177	1.59E-02	1	
CSRNP3	104412	947	2	138237	37389	102267	4	0	0	28	23	4357617	1.59E-02	1	
SRI	157359	454	23	50358	12705	163350	1	0	1	50	33	9996723	1.61E-02	1	
FABP1	180956	612	6	31944	7656	101079	2	0	0	18	12	4397052	1.61E-02	1	
MS4A6E	433906	899	24	34848	10494	78111	2	0	1	50	35	10181325	1.62E-02	1	
SFN	1022828	188	30	57618	16533	15345	1	0	0	50	28	10071600	1.62E-02	1	
CABP5	993450	433	26	42801	10857	124344	1	0	0	27	40	9522480	1.62E-02	1	
STIM2	116073	934	37	184272	48972	236907	2	0	1	50	34	12996489	1.63E-02	1	
SPHKAP	93607	615	-9	398673	111078	239283	10	0	3	37	31	7896273	1.63E-02	1	
HGF	57326	604	17	180312	44517	318978	6	0	5	16	20	3079428	1.64E-02	1	
HECW1	496063	617	13	382008	106359	553410	8	2	3	39	33	10128855	1.64E-02	1	
SNX22	743794	251	43	47157	13035	154143	1	0	1	50	42	11209374	1.64E-02	1	
ZIM3	311003	924	-64	115170	26895	100287	3	0	0	21	14	3409428	1.64E-02	1	
DLEU7	136985	649	31	35343	12969	9999	2	0	0	50	40	11715528	1.64E-02	1	
FAM107B	256093	327	36	73524	19239	172161	1	0	0	50	25	12005499	1.65E-02	1	
LMAN1L	764533	195	46	123156	38709	263637	2	0	0	28	19	5682567	1.65E-02	1	
PILRA	1430637	204	54	72303	20361	177606	1	0	0	50	18	10276464	1.65E-02	1	
C8orf44	372262	213	38	38049	10263	23562	3	0	0	50	40	10151559	1.66E-02	1	
PARP15	419536	311	44	166815	46233	163548	3	0	1	50	36	11426052	1.67E-02	1	
CACNA1G	1559131	196	19	573111	174042	424512	5	0	2	22	26	6452424	1.68E-02	1	
PKHD1L1	134643	732	6	1014354	276606	928818	13	0	7	7	9	1984356	1.68E-02	1	
CTCFL	181464	562	10	160908	40260	257400	3	2	0	30	22	7800012	1.69E-02	1	
NPPC	893534	439	33	27951	10164	0	1	0	0	15	13	3266241	1.70E-02	1	
OR6B3	315906	569	21	74745	23760	5049	2	0	0	35	27	7668012	1.70E-02	1	
FAM135B	6187	989	-37	330693	94710	381843	6	4	1	16	19	2921490	1.71E-02	1	

gene	expr (expression level of this gene, averaged across 91 cell lines)	reptime (DNA replication time of this gene)	hic (chromatin state of this gene)	N_nonsilent (total number of nonsilent mutations)	N_silent (total number of silent mutations)	N_noncoding (total noncoding mutations)	n_nonsilent (Number of nonsilent coding mutations)	n_silent (Number of silent coding mutations)	n_noncoding	nnei (number of neighboring genes)	x (number of mutated bases in the neighboring genes)	X (total number of bases related to the neighboring genes)	pvalue	qvalue	Comment
IL4R	227872	560	25	194172	56298	197307	3	1	0	50	27	11256234	1.72E-02	1	
CNST	416573	346	4	174702	46167	253242	3	0	0	8	5	2525721	1.72E-02	1	
BBS2	543900	285	20	172260	48510	355113	3	0	2	37	35	10417011	1.73E-02	1	
IL32	1626425	234	44	47355	11748	121968	1	0	0	50	34	8999529	1.74E-02	1	
KRT72	1280474	464	36	121143	34683	184041	4	1	0	50	45	10747110	1.74E-02	1	
RAP1B	439623	356	37	46035	11682	234333	1	0	0	9	6	3968118	1.75E-02	1	
JAKMIP2	102729	642	-4	199353	49038	491535	3	1	1	26	26	7497600	1.75E-02	1	
OR10R2	177013	929	7	76065	23628	10098	2	0	0	50	34	9270492	1.76E-02	1	
DCLK3	374894	617	-14	152394	41547	93159	4	2	0	40	40	6248616	1.76E-02	1	
PDCD1	559742	487	29	66495	21120	44451	3	0	0	50	41	13566795	1.76E-02	1	
AADACL3	145424	454	23	82995	23034	100386	2	0	0	50	34	10079883	1.77E-02	1	
ITGB8	207764	550	30	186450	47388	230175	2	2	1	16	30	5170374	1.77E-02	1	
ANKRD22	175425	289	26	46629	12375	141669	1	0	1	50	49	11520696	1.78E-02	1	
MND1	202295	238	27	52668	11484	104049	1	0	0	45	28	10220166	1.79E-02	1	
RIPPLY2	108781	870	-22	31317	8184	52569	1	0	0	44	54	7860369	1.79E-02	1	
AVL9	371440	595	10	155892	42801	293238	2	2	1	50	36	10380414	1.80E-02	1	
ACCS	345807	680	3	121044	33396	231363	2	0	2	30	24	6157701	1.82E-02	1	
HTR3A	175442	586	1	122793	36102	159687	2	0	1	47	42	10095129	1.82E-02	1	
CDH2	43239	931	-16	215094	60225	304722	6	1	0	5	7	1562352	1.82E-02	1	
G6PC2	173915	724	7	85041	22473	108207	2	0	0	50	39	11009460	1.83E-02	1	
VSNL1	135361	1255	-2	47124	11088	60192	1	0	0	50	41	9539838	1.83E-02	1	
RFXAP	191525	731	-15	63195	18678	30690	2	0	0	50	63	11623953	1.84E-02	1	
ADM	860704	273	12	42933	13497	40194	2	0	0	50	42	12499278	1.87E-02	1	
C6orf203	214917	417	22	59136	15510	62766	1	0	0	50	39	11443905	1.87E-02	1	
VTI1A	232217	459	7	53757	13761	150282	1	1	0	40	27	9751170	1.88E-02	1	
OR5D14	4183	1091	-36	72006	21450	10593	2	0	0	50	54	6787770	1.88E-02	1	
C7orf66	75279	982	-24	27918	7326	24255	1	0	0	50	122	13500234	1.89E-02	1	
VPS13B	209453	862	27	972576	265221	1498563	5	0	1	4	2	2884233	1.89E-02	1	
ITPR1	290645	395	19	668118	175065	776754	4	0	1	6	4	1976172	1.90E-02	1	
TSPAN3	319089	441	40	60852	16962	89001	1	0	0	50	38	11354541	1.90E-02	1	
USP36	854363	214	27	264462	76494	319176	3	1	0	6	1	1347324	1.91E-02	1	
MRGPRX1	561250	695	15	72501	23628	18513	1	0	0	50	24	10293261	1.91E-02	1	
CHP2	517081	598	33	48081	12804	109890	1	0	2	44	32	7759191	1.92E-02	1	
MATN2	616574	272	28	232518	58839	295416	3	0	0	4	0	1178100	1.92E-02	1	
KRTAP21-2	30244	1055	-70	19437	5313	18612	1	0	0	50	54	5871228	1.93E-02	1	
C12orf74	208773	509	-1	44220	14190	32967	1	0	0	50	43	9185748	1.93E-02	1	
TMEM158	NaN	NaN	NaN	63921	25476	0	1	0	0	50	87	22655325	1.93E-02	1	
OCIAD1	257328	372	63	64812	15180	166419	1	0	1	50	36	11510037	1.93E-02	1	
MOXD1	188578	673	1	147576	39336	206712	3	0	1	31	34	6677814	1.94E-02	1	
RAVER2	269434	456	-10	161238	44979	186714	3	0	0	4	1	999867	1.94E-02	1	
CLDN9	1443513	234	49	47058	17688	14850	1	0	0	50	34	7878585	1.95E-02	1	
CCDC152	NaN	NaN	NaN	65703	13398	26730	1	0	0	50	87	22669977	1.95E-02	1	
BNIP1	483076	397	34	65604	17556	133848	2	0	0	50	25	11290653	1.96E-02	1	
CD7	2856070	193	24	53691	19074	27324	1	0	0	50	31	8206605	1.96E-02	1	
PKD2L1	1071291	198	56	192093	53229	342144	5	0	1	15	7	3144636	1.97E-02	1	
ZNF880	526380	960	-32	138930	34122	28710	4	1	0	41	20	5237331	1.97E-02	1	
ULBP2	423805	245	34	58641	16500	57618	1	0	1	45	33	9748794	1.98E-02	1	
ZNF71	174439	849	-68	114180	31746	42273	6	1	0	50	56	9272010	2.00E-02	1	
TBX18	28205	623	-12	140118	43230	118701	5	0	1	50	42	10869375	2.01E-02	1	
TCEAL4	490345	NaN	31	52965	11979	20592	1	0	0	50	47	12728265	2.01E-02	1	
TMTC1	99753	510	-15	185823	51084	359568	4	2	4	8	18	2052138	2.01E-02	1	
CPT1A	619314	303	45	189618	51348	330957	3	1	1	50	36	11406285	2.01E-02	1	
ZFC3H1	259684	393	-11	479853	124641	640431	5	2	0	19	9	4524597	2.02E-02	1	
ORM2	206775	425	4	49401	12573	78705	1	0	0	50	40	11248182	2.02E-02	1	
MAL	171487	647	-1	35871	11253	67815	2	1	0	35	28	7072758	2.02E-02	1	
USH2A	82180	967	-25	1231560	342639	1447380	28	7	9	22	44	5588319	2.02E-02	1	
ATP1B1	231221	416	19	73557	18711	142164	1	0	0	50	25	9747507	2.03E-02	1	
KCTD11	1911735	206	36	50952	18249	79497	1	0	0	50	41	9311907	2.04E-02	1	
GRPEL2	422727	380	28	52734	15576	66528	1	0	0	33	29	7632009	2.04E-02	1	
FGF23	295315	860	33	58179	17457	56430	2	1	0	9	9	1446555	2.04E-02	1	
WRNIP1	473112	285	31	153978	46200	106722	2	0	0	50	32	11223729	2.04E-02	1	
FAM19A4	263741	842	-15	34683	8778	113553	2	0	0	50	44	9375432	2.06E-02	1	
DHTKD1	341646	227	25	218658	61116	352935	3	0	0	21	11	4764870	2.07E-02	1	
ARL6IP6	275277	716	6	52173	16434	130977	2	0	1	50	33	11525481	2.07E-02	1	
USP13	361278	589	12	210507	54219	440847	3	0	0	3	0	1028412	2.07E-02	1	
SV2B	457413	624	1	164835	42669	248391	2	0	0	6	0	1287033	2.07E-02	1	
SNTG1	2777	1189	-32	127446	33330	309276	4	1	4	50	68	8116977	2.08E-02	1	
FTMT	70878	920	-16	55374	16995	28215	5	1	0	3	2	377256	2.09E-02	1	
SMARCA4	1984564	190	40	401148	112563	477774	4	3	4	23	22	4678014	2.09E-02	1	
SDPR	158359	484	4	99198	27720	26829	2	0	0	50	24	9766449	2.09E-02	1	
RPL10L	10737	1057	-48	50160	14091	17622	3	0	0	50	36	6371904	2.10E-02	1	
TMEM121	689089	298	41	70026	25014	0	1	0	0	50	33	10567854	2.10E-02	1	
ST8SIA4	24350	1122	36	85371	23331	130086	1	0	0	50	34	13261809	2.11E-02	1	
THSD7B	11902	824	-16	379533	99330	343233	6	4	2	15	13	2495493	2.12E-02	1	
LIFR	159760	271	-2	265188	68442	343827	4	0	1	20	11	3792492	2.12E-02	1	
KRTAP15-1	14979	1035	-68	32142	9042	13959	3	0	0	50	53	5785395	2.13E-02	1	
FBN2	226750	605	9	715605	174900	1309671	5	2	4	14	10	4253106	2.13E-02	1	
MAPK1	458136	299	47	87648	22737	142956	1	0	0	50	23	10436316	2.14E-02	1	
LAMC3	815585	293	36	370656	108306	378675	5	0	2	29	21	6484500	2.14E-02	1	
TIMP1	775214	NaN	38	50292	13662	61776	1	0	0	50	46	9595476	2.15E-02	1	
C15orf60	294157	472	16	63426	18447	73458	1	0	0	50	40	11174361	2.15E-02	1	

gene	expr (expression level of this gene, averaged across 91 cell lines)	reptime (DNA replication time of this gene)	hic (chromatin state of this gene)	N_nonsilent (total number of nonsilent mutations)	N_silent (total number of silent mutations)	N_noncoding (total noncoding mutations)	n_nonsilent (Number of nonsilent coding mutations)	n_silent (Number of silent coding mutations)	n_noncoding	nnei (number of neighboring genes)	x (number of mutated bases in the neighboring genes)	X (total number of bases related to the neighboring genes)	pvalue	qvalue	Comment
ERCC5	196444	475	33	282612	75669	303534	3	0	0	17	8	5557497	2.15E-02	1	
SLC25A33	749351	268	33	75570	22440	166617	1	0	0	45	25	9391173	2.15E-02	1	
TBX20	154022	926	5	105039	30690	112266	2	1	0	50	47	8878023	2.15E-02	1	
IDH3A	395288	314	50	89100	24057	224037	1	0	0	50	29	12140172	2.16E-02	1	
KCNK9	62115	562	5	86856	25311	20493	4	0	0	50	37	10379193	2.17E-02	1	
CPA6	232126	890	15	106656	27390	227304	3	3	0	50	47	9404472	2.17E-02	1	
CLDND2	368509	610	-9	39600	11682	63459	1	0	0	50	34	7407642	2.18E-02	1	
NASP	873800	161	27	196284	51414	232947	2	0	0	13	4	3117246	2.18E-02	1	
SURF6	1846659	416	36	83523	25773	106326	3	0	0	50	17	8506707	2.19E-02	1	
HCN1	80822	1235	-24	207009	60588	130383	5	3	1	25	24	3434607	2.20E-02	1	
TMEM196	187539	874	-4	40656	12111	40194	1	0	0	50	63	11188056	2.20E-02	1	
C5orf20	625030	509	47	54945	17820	47619	1	0	0	50	59	12268641	2.22E-02	1	
KCNA1	260795	903	12	114444	32868	14850	2	0	0	50	40	8627619	2.23E-02	1	
CD200R1L	240755	616	6	64614	18150	127413	1	0	0	50	47	12100572	2.23E-02	1	
XBP1	571394	219	-10	94677	21648	90684	1	0	0	44	29	12635436	2.23E-02	1	
CLEC3B	479578	326	41	47883	13398	16236	2	0	0	50	39	12362493	2.24E-02	1	
CTAG2	2090264	NaN	31	56892	20130	22275	2	0	0	50	34	11248314	2.25E-02	1	
CCDC39	182757	1031	2	233244	54054	166518	4	0	2	50	39	8316693	2.26E-02	1	
STAT2	2869106	184	52	206085	56067	459756	2	0	0	17	10	5153412	2.26E-02	1	
EFCAB3	344431	362	8	121374	29007	221859	2	0	1	50	39	11045793	2.28E-02	1	
COL14A1	262056	839	-5	430650	121671	978714	4	2	2	12	15	3422958	2.28E-02	1	
KRTAP5-10	586193	626	50	46728	13761	33066	1	0	0	50	44	8285970	2.29E-02	1	
LRRCC1	99309	1450	-7	254199	59730	286704	3	0	0	8	2	1870539	2.30E-02	1	
KRTAP4-3	996799	468	-7	46860	11352	15741	1	0	0	50	41	7496379	2.30E-02	1	
RAET1L	402308	382	51	58344	16302	60687	2	0	0	50	28	10818027	2.32E-02	1	
SAMD5	97528	850	-17	39336	12738	32175	1	0	0	50	65	9088233	2.32E-02	1	
RFPL4B	96263	432	0	61182	17226	19008	3	1	0	7	6	842424	2.33E-02	1	
ZNF821	468594	163	28	87087	25080	110682	1	0	0	50	37	13491753	2.34E-02	1	
CWC22	35642	927	7	223278	53823	89199	4	0	1	50	46	8411238	2.34E-02	1	
GALNTL5	309129	476	12	107679	27357	188100	2	1	0	50	34	10307682	2.35E-02	1	
THEMIS	191535	553	-9	160677	43956	113652	2	0	0	48	48	12005928	2.36E-02	1	
PACS1	2403161	236	49	229614	65802	453222	2	1	1	50	21	8966199	2.36E-02	1	
PSME2	974559	220	41	60522	15114	268290	1	0	0	15	8	4114341	2.37E-02	1	
AGPAT4	74993	874	-5	91080	24651	172458	2	1	1	50	65	13084896	2.37E-02	1	
ANXA1	247950	411	-15	85734	22077	224730	1	0	1	50	31	11170863	2.37E-02	1	
CXorf22	3828	NaN	-13	234399	61908	216810	5	1	1	48	98	13851090	2.38E-02	1	
ZBTB22	1456953	262	30	140547	48444	100881	3	0	0	50	17	10161030	2.38E-02	1	
POU3F1	448782	374	32	99594	34848	0	1	0	0	50	25	11025564	2.41E-02	1	
MSH5	2509866	183	32	204336	58509	456291	2	0	3	39	18	7192086	2.42E-02	1	
RIMS2	198819	1035	-8	336864	92400	825858	7	1	9	20	30	4985508	2.42E-02	1	
FAM105A	514082	380	28	86922	21879	132957	1	0	0	28	15	7102854	2.43E-02	1	
F13A1	44522	696	22	175956	47289	306405	6	1	1	14	14	2842422	2.45E-02	1	
VEGFC	145765	585	-14	101376	26037	117513	3	0	0	50	31	12427668	2.45E-02	1	
NLK	1261475	410	-3	125631	35145	130185	2	0	0	50	43	8639268	2.45E-02	1	
COL6A2	742075	538	10	266112	79794	378378	6	1	3	31	25	6232776	2.45E-02	1	
KRTAP13-4	14979	1035	-68	36795	10824	16038	3	0	0	50	53	5785395	2.45E-02	1	
TNFAIP3	315946	256	51	188133	49962	149292	2	2	0	13	4	2800611	2.47E-02	1	
MMP3	468929	506	-21	114972	30756	168597	4	1	0	16	19	4199811	2.48E-02	1	
ZNF808	546511	1005	-53	215853	53823	87615	2	0	0	13	7	1861926	2.48E-02	1	
CCNYL1	330620	210	26	88209	22275	180180	1	0	1	50	38	12637119	2.49E-02	1	
KLK12	490795	632	22	66132	15939	81279	1	0	0	50	33	8896173	2.49E-02	1	
SMARCAL1	590685	384	31	225753	64416	329670	3	1	0	50	44	15223263	2.49E-02	1	
STX7	180943	618	12	65373	16005	157707	1	0	1	50	34	8822847	2.51E-02	1	
OR2AK2	NaN	NaN	NaN	75999	23694	7623	3	0	0	50	87	22661166	2.52E-02	1	
CDH12	127633	1075	-20	187638	52437	504207	8	3	1	16	24	3287988	2.52E-02	1	
DDX59	371965	276	13	146751	40161	132363	3	2	0	4	3	794244	2.53E-02	1	
PRUNE	1617972	165	35	105996	31614	196020	2	0	0	50	34	11880033	2.53E-02	1	
FOXG1	52980	1173	-46	110484	35046	8316	4	0	0	50	64	7431699	2.54E-02	1	
CSPG4	578030	231	39	515427	178266	128007	8	1	0	50	41	10901319	2.56E-02	1	
B3GALTL	183118	606	27	122892	30855	257598	3	1	3	10	12	2640561	2.56E-02	1	
PSTPIP1	327336	436	33	102663	26730	92664	2	0	0	50	42	10413975	2.59E-02	1	
GIMAP6	682986	689	20	67551	20064	58608	2	0	0	50	20	9094569	2.59E-02	1	
NKX2-1	58829	624	-18	91575	28809	28215	2	0	0	18	19	3174336	2.60E-02	1	
WDR86	608657	190	39	86130	27819	37818	1	0	0	50	32	10681209	2.60E-02	1	
EVX1	473864	394	13	90651	31317	41679	1	0	0	50	48	14908971	2.60E-02	1	
ATE1	233024	427	-3	137808	34056	207999	2	0	1	33	23	5818758	2.61E-02	1	
MTMR1	288573	NaN	9	160446	43296	245520	3	0	1	13	9	2607660	2.62E-02	1	
GPR137C	356488	313	14	99660	30426	84249	2	1	1	9	5	1601886	2.62E-02	1	
SERPINA3	316239	858	10	99231	28281	103455	2	1	1	21	20	3594855	2.62E-02	1	
CHMP4C	158801	681	12	56496	14586	73260	2	0	0	50	50	9493407	2.63E-02	1	
TSFM	1160849	344	39	87252	26202	180972	1	0	0	34	17	7141959	2.63E-02	1	
ADIPOR2	259057	403	29	92796	24915	230175	1	0	1	50	53	16179636	2.64E-02	1	
ELMO1	63863	859	16	180378	44154	534204	3	1	4	20	20	3949341	2.64E-02	1	
ZIC1	20725	1070	-30	103356	30492	40788	3	1	0	24	42	4662735	2.65E-02	1	
GNB1L	690032	204	10	75768	24024	153846	2	1	0	50	29	12734799	2.66E-02	1	
CCBE1	257439	491	24	97152	27885	195228	1	0	0	44	26	11400048	2.66E-02	1	
DCDC2	362257	711	-12	124113	31218	203940	2	0	0	10	3	1598916	2.66E-02	1	
CGNL1	189519	561	10	311157	82962	309276	3	0	0	7	2	1451010	2.66E-02	1	
KLK1	655673	589	25	62832	17259	108999	1	0	1	50	47	11102355	2.67E-02	1	
ETV5	780547	397	36	123255	33264	224631	2	0	1	24	19	4498791	2.67E-02	1	
OR2T3	112460	903	-20	72864	21879	9900	6	1	0	13	27	2165724	2.67E-02	1	

gene	expr (expression level of this gene, averaged across 91 cell lines)	reptime (DNA replication time of this gene)	hic (chromatin state of this gene)	N_nonsilent (total number of nonsilent mutations)	N_silent (total number of silent mutations)	N_noncoding (total noncoding mutations)	n_nonsilent (Number of nonsilent coding mutations)	n_silent (Number of silent coding mutations)	n_noncoding	nnei (number of neighboring genes)	x (number of mutated bases in the neighboring genes)	X (total number of bases related to the neighboring genes)	pvalue	qvalue	Comment
YBX1	650376	461	21	77385	21912	106920	1	0	0	50	38	10345434	2.68E-02	1	
ACSM2A	181483	823	6	138270	38544	298386	3	1	0	37	34	6826809	2.69E-02	1	
MRPL44	314020	346	23	77715	22572	65142	1	0	0	50	45	11487036	2.70E-02	1	
FKBP10	1760409	175	35	136026	40689	171864	2	0	0	50	32	11082423	2.70E-02	1	
OR4K1	420010	827	-51	72600	20460	20592	6	0	1	15	9	1567137	2.71E-02	1	
SLFN11	761918	456	11	211365	58113	78210	6	0	0	44	31	7577922	2.71E-02	1	
NACA2	323585	358	-8	49764	14784	19206	2	0	0	50	37	10345170	2.72E-02	1	
UCHL5	288979	433	-18	83259	18711	216810	3	0	2	50	41	9980817	2.74E-02	1	
C9	168073	424	-8	135927	34551	230373	4	0	2	50	52	10397475	2.74E-02	1	
TMEM208	1109491	150	46	41877	11781	89892	2	0	0	50	40	9716520	2.75E-02	1	
PSMA6	452722	188	40	60126	15609	151074	2	1	0	50	36	10297485	2.75E-02	1	
HDAC11	370030	199	39	83193	24321	149193	1	0	1	50	38	10638738	2.75E-02	1	
ATF2	347113	641	24	120450	34584	224928	1	0	1	50	32	14193366	2.76E-02	1	
RAB6C	342211	671	4	58575	17160	20493	2	0	0	50	36	10259337	2.76E-02	1	
GABRA6	12083	986	-6	108273	29733	183348	2	0	0	19	17	3385503	2.77E-02	1	
DTNA	116061	552	-20	188133	50160	422235	5	1	2	5	12	1859748	2.77E-02	1	
CBLN4	236365	941	-74	45969	14817	51381	1	0	0	50	53	7581783	2.79E-02	1	
LAMA1	241772	228	-37	736824	201498	1118502	7	5	6	19	27	5189382	2.79E-02	1	
GRP	280782	512	53	36069	9174	47223	1	1	0	11	8	2068275	2.79E-02	1	
MCF2L2	398145	606	24	273372	69465	585684	4	0	0	11	2	2909775	2.81E-02	1	
C8orf74	110362	481	9	68739	20262	20889	1	0	0	50	48	10482186	2.81E-02	1	
LMAN2L	921353	219	45	86559	23727	149490	2	0	0	48	30	12812283	2.81E-02	1	
FSTL1	382340	375	23	77121	18414	220275	1	0	2	48	45	12390510	2.82E-02	1	
GPR116	168285	392	0	321354	86625	382041	3	0	0	5	0	1339239	2.83E-02	1	
UNC5CL	134733	565	32	121044	36267	128700	2	0	1	50	45	13799082	2.84E-02	1	
CACNA1S	522609	490	31	449295	124311	868725	3	2	0	17	17	5031873	2.84E-02	1	
DMRTB1	552240	385	25	77319	25938	89100	3	0	0	47	31	9936333	2.85E-02	1	
WDR12	596746	392	31	104082	26598	206415	1	0	0	50	41	13852014	2.85E-02	1	
EOMES	165387	429	1	157641	48378	100683	2	1	0	42	40	9841128	2.86E-02	1	
ZBTB44	658636	276	18	107250	29568	265815	1	0	0	37	19	8507037	2.87E-02	1	
DUSP13	223189	298	40	111507	34716	91080	1	0	0	50	25	10665336	2.88E-02	1	
GH2	1446139	380	27	81675	18810	68112	1	0	0	28	18	6480144	2.89E-02	1	
GNAI3	1022796	187	33	87087	21516	156123	1	0	0	50	26	8311083	2.89E-02	1	
OR2Y1	831824	515	18	69960	22704	7326	2	0	0	50	21	8025435	2.91E-02	1	
OR1L4	409923	553	12	71247	21417	13761	3	0	0	50	37	8945640	2.92E-02	1	
CD200	260383	621	20	69828	20163	127512	1	0	1	50	38	10628343	2.93E-02	1	
CDX2	309835	614	47	71049	23001	29601	3	0	0	50	39	8698305	2.93E-02	1	
PPP1R7	1363901	233	35	88407	22572	216315	1	0	0	50	42	11568876	2.93E-02	1	
ACOT1	NaN	NaN	NaN	94809	31317	29700	1	0	0	50	87	22690866	2.93E-02	1	
CCDC91	96900	922	-5	109395	26433	200475	2	0	0	45	40	9137700	2.94E-02	1	
SERPINB12	190401	339	20	98241	24519	164736	3	0	2	50	40	10739256	2.95E-02	1	
MATN1	286775	493	27	115962	34617	71379	3	0	0	50	35	11441958	2.95E-02	1	
DOK2	402326	470	12	94380	30063	55143	1	1	0	50	42	12455355	2.96E-02	1	
SIGLEC15	384767	380	42	74778	24915	72369	1	0	0	50	36	10780110	2.97E-02	1	
NEUROD2	1164935	239	28	86097	28050	297	2	0	0	50	17	8814003	2.97E-02	1	
C8orf47	630526	229	39	87153	25014	28611	1	0	0	50	43	11915079	2.97E-02	1	
TLN2	315754	568	15	601623	175428	1010889	6	1	2	38	27	9814530	2.98E-02	1	
UGT3A1	155890	614	-4	129492	36432	119295	3	0	1	17	28	4447674	2.98E-02	1	
OR6C74	411183	816	-1	71478	21384	9405	3	0	0	50	43	9315537	2.98E-02	1	
SERTAD1	1010633	202	34	52503	18084	10098	2	0	0	50	29	8091765	2.98E-02	1	
WIPF2	1491292	165	30	99363	34386	150777	1	0	0	50	33	11135916	2.99E-02	1	
ING2	204597	227	15	67386	16467	13464	1	0	0	50	27	7757904	3.00E-02	1	
OR8B3	149616	857	0	72237	20823	4158	3	0	0	50	40	11261943	3.01E-02	1	
SLC38A4	518041	415	-10	132990	35310	274032	2	0	1	31	23	9085890	3.01E-02	1	
INADL	274287	469	-3	432003	119823	844767	5	0	2	17	16	5416521	3.01E-02	1	
APPL1	427002	389	33	175527	43857	339471	4	0	0	11	2	1758966	3.01E-02	1	
PIP4K2A	196977	577	25	99429	25014	182655	1	0	0	43	26	9347646	3.02E-02	1	
CPZ	92046	471	35	153318	44979	110286	4	1	0	50	37	12452847	3.02E-02	1	
AP1G2	1250373	553	35	184239	57519	251064	3	0	0	28	26	7017912	3.02E-02	1	
KRT85	936407	213	42	119262	34782	175032	1	0	0	48	28	11495979	3.02E-02	1	
TCTE3	233648	447	3	48642	12045	90783	1	0	1	10	5	2273535	3.03E-02	1	
FAM160A2	466636	728	4	221661	75438	204237	4	1	0	27	35	6966663	3.05E-02	1	
GLCCI1	179344	405	24	126027	39501	217998	2	1	0	50	39	11542410	3.05E-02	1	
CCDC34	336836	378	-21	97548	23826	102861	1	0	0	50	44	12754632	3.05E-02	1	
GJB7	222729	596	22	52965	13959	87219	2	0	0	35	26	7757145	3.06E-02	1	
HSD17B6	2637730	154	47	74547	21483	74943	2	0	0	50	33	12552804	3.06E-02	1	
UBQLN4	2005405	213	39	140580	42372	195228	2	0	0	28	24	5302869	3.08E-02	1	
COL9A1	126322	633	5	225753	70059	606276	4	1	3	25	25	4712136	3.08E-02	1	
TMEM215	610744	401	-18	54450	16038	19800	2	0	0	50	42	11926464	3.08E-02	1	
PDYN	461602	543	28	59235	17292	43758	2	0	0	50	37	11593560	3.09E-02	1	
OR5T1	10317	969	-43	75075	22044	16830	4	0	0	50	33	6056226	3.10E-02	1	
SPOCK3	7931	1043	-11	109098	24849	232353	4	0	0	3	3	645480	3.11E-02	1	
CSMD3	9197	1295	-29	896247	239085	1281852	23	3	9	1	12	1568985	3.12E-02	1	
MID1IP1	91133	NaN	31	42504	12144	9405	1	0	0	10	5	1908489	3.12E-02	1	
ACOT2	606300	169	39	108702	35838	66825	3	0	1	50	39	11067969	3.12E-02	1	
ATOH8	1105419	181	46	70785	25641	35145	1	1	0	39	34	7703916	3.13E-02	1	
LCT	493727	403	28	452133	127017	358875	6	0	1	12	13	3944358	3.13E-02	1	
ZDHHC9	370406	NaN	69	87417	24156	198792	1	0	1	50	44	11146575	3.13E-02	1	
DIO3	1315103	577	-26	68739	21846	16236	1	0	0	50	41	8389590	3.13E-02	1	
EFNB2	167697	1267	2	79398	21582	98901	1	0	0	50	32	8254653	3.14E-02	1	
GIMAP2	682986	689	20	80355	20823	31086	2	0	0	50	20	9094569	3.14E-02	1	

gene	expr (expression level of this gene, averaged across 91 cell lines)	reptime (DNA replication time of this gene)	hic (chromatin state of this gene)	N_nonsilent (total number of nonsilent mutations)	N_silent (total number of silent mutations)	N_noncoding (total noncoding mutations)	n_nonsilent (Number of nonsilent coding mutations)	n_silent (Number of silent coding mutations)	n_noncoding	nnei (number of neighboring genes)	x (number of mutated bases in the neighboring genes)	X (total number of bases related to the neighboring genes)	pvalue	qvalue	Comment
PDE7B	283639	392	6	112002	26697	244530	1	0	1	50	24	9599832	3.15E-02	1	
SUMF1	155099	172	10	88836	25905	178794	2	0	0	6	6	1393887	3.15E-02	1	
PGBD5	286328	478	10	108603	29403	175725	1	0	0	50	34	11711040	3.17E-02	1	
GPX5	247624	576	27	55143	12573	134937	2	0	0	25	21	5967456	3.17E-02	1	
NETO1	32720	916	-26	132792	33330	200475	3	1	2	50	72	9572310	3.17E-02	1	
GJA8	NaN	NaN	NaN	99363	30129	13860	1	0	0	50	87	22673838	3.18E-02	1	
GLYCTK	1113485	229	49	120648	36564	105732	2	0	0	50	41	11459943	3.19E-02	1	
ZFYVE26	484591	344	49	598158	172458	880407	5	1	3	39	23	8039427	3.19E-02	1	
DEFB119	591819	594	-83	41283	9702	100287	1	0	0	14	8	2038212	3.20E-02	1	
PRR25	1187310	273	39	87087	33297	4059	1	0	0	50	39	10013520	3.20E-02	1	
DPP10	7721	1178	-22	211266	52272	494901	3	0	1	3	1	819753	3.20E-02	1	
GFRA3	1219794	273	43	94413	27456	99495	1	0	0	50	40	11051898	3.20E-02	1	
KRTAP6-3	22916	1048	-79	25608	7557	7227	2	0	0	50	58	6825720	3.21E-02	1	
BNIP2	521844	172	33	104808	28446	119493	1	0	0	50	42	12631311	3.21E-02	1	
HPGDS	129148	311	14	49599	11781	96228	1	0	0	11	11	2729430	3.22E-02	1	
ATXN3	356163	415	27	91707	19965	212256	1	0	0	31	37	10085592	3.22E-02	1	
FBXO22	550701	339	47	98043	27786	115236	1	0	0	50	45	12513270	3.22E-02	1	
LTB4R	915114	220	38	76296	28941	22077	2	0	0	50	29	9879969	3.23E-02	1	
SLC25A10	3745109	205	30	68508	21186	137214	1	0	1	50	36	7552149	3.24E-02	1	
ITGAD	839076	558	11	277035	79761	486684	4	3	1	31	31	7093449	3.25E-02	1	
LHFPL3	324628	895	5	55011	16170	212256	2	1	0	45	47	10953756	3.25E-02	1	
DNAJB6	318088	174	22	82335	21615	175032	1	0	2	18	20	5944884	3.25E-02	1	
FAM78B	320493	733	-12	60357	17853	41184	3	0	1	19	10	2990526	3.27E-02	1	
OR5M11	NaN	NaN	NaN	69663	21120	9900	2	0	0	50	87	22660869	3.30E-02	1	
LIPH	953031	366	31	109362	28446	212058	1	0	1	42	21	8096253	3.31E-02	1	
GPT2	482346	362	12	124311	35673	229185	1	1	0	50	40	13824096	3.34E-02	1	
OR8K3	10317	969	-43	71577	21186	10197	3	0	0	50	33	6056226	3.34E-02	1	
C3orf30	197435	589	-3	126159	34320	70686	2	0	1	50	63	10346754	3.35E-02	1	
CSNK1A1L	156140	631	31	79563	20823	22176	3	0	0	50	32	10961610	3.36E-02	1	
ZNF385D	10169	1033	-21	92532	27852	152559	2	1	0	14	10	1988448	3.36E-02	1	
NDUFS3	989876	214	28	63393	18282	152163	2	1	0	50	42	11482218	3.37E-02	1	
KLHL21	628180	259	38	134838	44154	35838	1	0	0	50	30	12260391	3.38E-02	1	
HARBI1	914919	176	39	80949	23397	46134	2	0	0	50	30	11716980	3.38E-02	1	
WAS	824559	NaN	27	116292	37653	95238	1	0	1	50	38	12783144	3.39E-02	1	
RIMBP2	225009	632	-35	246576	72897	280368	5	2	1	14	19	2896278	3.39E-02	1	
ELK1	772893	NaN	41	96393	33000	56628	3	1	0	50	56	10857132	3.40E-02	1	
HERPUD1	820472	125	47	93192	26004	160182	1	0	2	28	21	7367514	3.40E-02	1	
SESN2	766835	170	29	112992	33429	170478	1	0	1	26	14	6684975	3.40E-02	1	
HAPLN2	2049505	180	37	78144	25113	25740	1	0	0	50	44	9602835	3.41E-02	1	
C5orf45	1610057	359	38	81345	23397	123651	1	0	0	30	14	5458662	3.42E-02	1	
PRELP	529605	314	28	87450	27093	43065	2	0	0	50	32	11705991	3.44E-02	1	
DCT	123751	906	0	132132	35673	155727	2	0	2	50	49	8842284	3.44E-02	1	
ASCL1	21203	964	3	54054	16731	8514	1	0	0	50	65	8763843	3.44E-02	1	
NCKAP5	36731	826	18	446160	128238	206514	6	2	2	50	51	10173504	3.44E-02	1	
PRDM2	68342	892	25	407946	112497	245520	3	2	1	27	25	5382003	3.45E-02	1	
MESDC2	265620	516	-2	55308	15477	52767	2	0	0	50	44	10055727	3.45E-02	1	
ALX1	35916	925	-34	77847	20658	66825	5	0	1	33	43	4225749	3.45E-02	1	
MTX1	1801444	194	32	106623	34848	120879	1	0	1	50	39	11328636	3.46E-02	1	
ISL2	289294	229	39	84546	24552	101673	1	0	0	50	34	8422359	3.47E-02	1	
USP6	1391000	384	20	340098	89166	662409	5	0	0	0	0	751575	3.48E-02	1	
KIAA0196	489847	520	25	281457	73755	532026	3	2	0	41	24	8090808	3.50E-02	1	
ARMCX3	329999	NaN	8	87747	25113	26928	1	0	0	50	47	11831952	3.50E-02	1	
SSBP2	159943	864	23	90288	23562	224730	1	0	1	50	59	14653617	3.50E-02	1	
HEATR5B	614688	281	16	485166	144078	594990	4	1	0	26	13	6054048	3.50E-02	1	
OR2A12	222080	812	-11	71115	21153	8415	1	0	0	50	42	8058996	3.50E-02	1	
NR4A1	468067	234	41	135861	44418	85041	1	1	0	50	23	9882015	3.50E-02	1	
TMEM184A	595899	243	38	96855	29073	68409	1	0	0	50	40	10459812	3.52E-02	1	
PCDH8	248822	638	27	235389	83490	36531	4	1	0	50	31	11032890	3.52E-02	1	
AVPR2	2328608	NaN	27	84084	29172	37026	1	0	0	50	40	9142419	3.52E-02	1	
SLC22A10	1307417	572	17	126588	37950	188496	2	0	1	46	35	8721174	3.53E-02	1	
ICAM1	1729517	200	31	120615	40062	93753	1	0	0	50	41	13225080	3.53E-02	1	
GRAMD1B	684637	439	15	181203	45804	352341	2	0	1	22	20	5568486	3.54E-02	1	
H2BFWT	439642	NaN	30	40227	13035	34848	1	1	0	9	9	2107446	3.54E-02	1	
GGPS1	513285	188	22	72105	18480	92862	3	0	0	50	41	11564223	3.55E-02	1	
APLF	227462	617	22	124377	31251	225225	4	0	0	3	3	1352505	3.55E-02	1	
HLF	147479	526	-54	69498	19602	43362	1	0	0	50	38	7158921	3.57E-02	1	
CLMN	207561	612	30	238524	64119	253638	3	0	0	6	0	1383888	3.58E-02	1	
ZP4	14616	1051	-30	128337	36894	212157	6	0	2	29	42	5035107	3.58E-02	1	
GPR119	301748	NaN	37	75603	24585	22077	1	0	0	50	44	10685829	3.59E-02	1	
FAM63A	1621279	172	30	122331	35475	215226	1	0	0	50	34	11522049	3.59E-02	1	
NSUN6	74263	512	-4	112728	30822	203148	2	0	0	13	17	4948416	3.59E-02	1	
MAPK8IP2	1073094	203	43	193512	64482	74547	2	0	1	38	26	6883107	3.59E-02	1	
SUOX	2080397	229	31	123750	39600	68706	1	0	0	50	30	10661112	3.60E-02	1	
C9orf43	390186	297	36	113982	28380	323730	2	0	3	33	24	6801168	3.61E-02	1	
RAG1	97499	412	-17	243540	66924	19206	6	0	0	22	20	3915747	3.62E-02	1	
CD300E	876565	484	14	48345	14421	68409	2	0	0	50	36	6843276	3.62E-02	1	
WDR81	1447501	189	32	442002	144474	142461	3	0	1	50	32	10718301	3.63E-02	1	
FAM153A	1267336	553	18	80487	19404	269577	1	0	2	32	22	5592213	3.64E-02	1	
MST1	1551582	271	48	173844	48906	177408	3	2	0	19	28	5222745	3.64E-02	1	
MYCBP2	236609	713	59	1117116	305019	1390455	9	0	3	3	4	2039697	3.64E-02	1	
B3GAT1	270422	686	5	76230	25245	44253	2	0	0	50	44	12631542	3.66E-02	1	

gene	expr (expression level of this gene, averaged across 91 cell lines)	reptime (DNA replication time of this gene)	hic (chromatin state of this gene)	N_nonsilent (total number of nonsilent mutations)	N_silent (total number of silent mutations)	N_noncoding (total noncoding mutations)	n_nonsilent (Number of nonsilent coding mutations)	n_silent (Number of silent coding mutations)	n_noncoding	nnei (number of neighboring genes)	x (number of mutated bases in the neighboring genes)	X (total number of bases related to the neighboring genes)	pvalue	qvalue	Comment
CDH18	4731	1195	-17	189024	49863	230076	8	2	1	6	7	1006005	3.66E-02	1	
ADAMTS12	373086	875	7	380820	102201	526383	10	1	3	20	20	3897465	3.66E-02	1	
TWF2	1114737	181	52	83754	23364	130185	2	0	0	50	25	10774797	3.67E-02	1	
RCSD1	520615	405	14	97449	28776	118503	1	1	0	50	35	11325204	3.67E-02	1	
POU6F1	496692	176	40	69927	21747	93456	2	0	0	50	33	10164660	3.67E-02	1	
SPRY1	98850	613	21	74679	20757	26532	1	0	0	38	37	7313790	3.68E-02	1	
NPY1R	74427	1197	-18	90420	24321	26928	4	0	0	21	22	2825658	3.68E-02	1	
KLF14	241201	399	35	70983	25443	8118	2	0	0	50	33	11284515	3.69E-02	1	
GRB7	1448852	226	36	126786	37059	210771	1	0	0	50	41	13457433	3.70E-02	1	
SRRD	163136	553	-46	81180	22572	138204	1	0	0	18	7	3265383	3.70E-02	1	
MPP3	1053443	192	27	142263	38907	285516	3	0	1	50	25	10362990	3.70E-02	1	
ZNF229	725569	528	-49	196746	49764	81081	5	0	0	50	49	9180402	3.71E-02	1	
SMPD3	794719	279	46	150711	46497	79794	2	1	0	50	32	11471130	3.72E-02	1	
UQCRB	217849	671	12	28314	6138	97020	2	1	1	7	5	1211199	3.73E-02	1	
LRRC3B	5998	918	-24	60357	16863	22176	4	1	0	16	13	2252052	3.74E-02	1	
OSBPL8	586978	492	3	219351	54087	383526	3	0	0	10	7	3319041	3.75E-02	1	
MPEG1	542222	686	9	163713	49236	19305	4	1	0	19	10	3055404	3.75E-02	1	
OR52M1	190769	879	-1	70686	23661	6930	2	0	0	50	27	9708765	3.76E-02	1	
PCDHA1	861651	606	2	216117	70290	425205	6	2	0	19	24	5586141	3.76E-02	1	
NFKBIL1	2517285	220	35	87516	27126	93060	2	0	0	44	19	7548288	3.78E-02	1	
SLC25A20	1790488	140	38	72171	20691	167508	2	0	0	50	34	11170137	3.79E-02	1	
HOXA7	356045	385	28	53493	15510	31680	2	0	0	50	50	11428395	3.79E-02	1	
CACNA1B	286606	675	29	550374	162426	590337	6	1	2	24	15	5168823	3.80E-02	1	
SYCE1L	NaN	NaN	NaN	60588	15543	28908	2	0	0	50	87	22674300	3.80E-02	1	
PCDHA4	861651	606	2	213840	69894	373824	5	2	0	50	57	13104894	3.80E-02	1	
GIMAP7	563029	671	24	71643	18150	152460	2	0	0	50	27	8938710	3.81E-02	1	
ENO3	1416161	231	30	104478	29271	195525	2	0	0	34	12	7072395	3.81E-02	1	
OR51B2	100115	1056	1	72006	21351	6930	1	0	0	50	36	6689562	3.82E-02	1	
WDR38	1125524	207	41	74910	22011	143847	1	0	0	29	13	3692865	3.82E-02	1	
ZNF683	1021101	217	22	116127	36333	55539	1	0	0	50	33	10105359	3.82E-02	1	
CTNNAL1	337834	433	14	178563	47058	265221	2	0	0	15	10	3341778	3.84E-02	1	
SLC37A3	354350	587	28	133617	33693	254430	1	1	0	43	23	9274254	3.84E-02	1	
SLC9A2	75782	858	-5	190476	55341	226512	3	3	0	23	31	6244062	3.84E-02	1	
SERPINB10	133459	522	19	96657	23727	120186	3	1	1	3	9	1659603	3.85E-02	1	
C1QTNF7	144115	540	20	68607	20394	46134	2	0	0	50	39	10076088	3.85E-02	1	
CAPNS1	978272	326	6	66396	17754	179982	1	0	0	18	15	3463845	3.85E-02	1	
AQP2	1033911	220	48	60621	21549	65934	2	0	0	50	53	11249469	3.87E-02	1	
TM6SF2	940952	232	17	89595	26433	132462	2	0	1	50	35	10338042	3.88E-02	1	
PFKFB3	560233	399	38	127875	34188	349470	1	1	2	32	25	9077574	3.89E-02	1	
LRRC7	298237	695	-26	367752	98736	509355	8	2	1	1	3	648714	3.90E-02	1	
TNFRSF21	206419	389	1	151602	45210	76824	1	0	0	50	35	13528515	3.91E-02	1	
RYR3	140241	806	-13	1176516	311157	1312443	17	4	5	2	9	2191167	3.94E-02	1	
ANKRD18A	NaN	NaN	NaN	242583	58476	28413	3	0	0	50	87	22716738	3.94E-02	1	
C14orf37	285319	683	3	180642	52305	172557	2	2	0	49	49	11560494	3.94E-02	1	
OR5J2	10101	1007	-38	71412	21351	7524	4	0	0	50	63	7813641	3.94E-02	1	
KCMF1	926772	202	23	90387	25443	60588	2	0	0	50	31	11210166	3.95E-02	1	
TMEM53	563553	252	45	63261	20295	69399	2	0	0	50	38	10290423	3.95E-02	1	
APPL2	233879	387	30	163284	42141	366102	2	1	0	9	7	2400420	3.95E-02	1	
C15orf43	706600	569	31	55407	12804	102663	2	0	1	48	49	10869012	3.97E-02	1	
DDX4	331341	486	34	181071	46629	389466	2	0	1	7	3	1827903	3.97E-02	1	
ESM1	205256	603	13	44187	11748	67518	2	0	1	50	34	8733879	3.98E-02	1	
ADSSL1	681504	208	29	133452	39699	191466	1	0	0	50	25	9203238	4.00E-02	1	
GFI1B	1887171	461	34	79266	21417	70983	2	0	0	50	28	9563928	4.00E-02	1	
ADCY2	60911	1014	-22	262020	72006	544995	4	2	1	9	10	3140313	4.02E-02	1	
NCAPD3	270385	601	14	361647	97020	657954	3	0	3	18	11	4307820	4.02E-02	1	
DCHS2	63258	1024	3	846714	249711	388377	10	4	2	50	35	7043058	4.03E-02	1	
UGDH	658220	165	48	119163	32208	183150	1	0	1	50	31	9520830	4.03E-02	1	
POLDIP3	914105	189	69	98967	29733	195525	1	0	0	25	15	5019135	4.03E-02	1	
KIAA1109	325000	770	8	1196646	323202	1385604	7	0	0	2	0	1809159	4.03E-02	1	
PLK4	230666	331	9	233475	60852	228492	2	0	0	7	2	2007291	4.03E-02	1	
C17orf85	505430	215	30	150645	38742	122364	1	0	1	50	23	9764106	4.03E-02	1	
AMIGO1	1056159	272	37	112431	34683	12573	2	0	0	50	39	8315637	4.03E-02	1	
MSN	325257	NaN	47	141306	35112	152163	1	0	1	50	36	12486936	4.04E-02	1	
ZNF141	427337	756	5	115236	27225	138501	1	0	0	48	47	12544752	4.04E-02	1	
HOXA13	473864	394	13	87450	28677	39204	2	0	0	50	48	14908971	4.05E-02	1	
CXCL16	1272092	222	28	63525	20031	145035	2	0	1	50	42	10747869	4.06E-02	1	
PTPN22	326668	406	15	199683	50292	354222	3	1	2	31	30	6848523	4.06E-02	1	
CDCA7	255854	559	28	108966	28545	231165	1	0	0	28	16	6122853	4.06E-02	1	
UBXN7	1165640	231	37	119031	30657	227205	1	0	3	50	23	7682235	4.08E-02	1	
SMR3B	264684	915	25	18150	6402	54648	1	0	0	50	34	10912242	4.09E-02	1	
BARD1	310874	948	11	185559	49467	178596	4	2	2	22	21	4265877	4.09E-02	1	
STK33	314196	551	13	126687	30624	203346	3	1	0	50	44	12410640	4.10E-02	1	
C11orf65	405487	579	12	78474	18150	146718	2	1	1	2	2	677424	4.11E-02	1	
OR4K5	431654	854	-49	73557	22473	12969	3	0	0	50	30	5391639	4.11E-02	1	
PRLR	247066	854	17	148335	39864	187110	2	0	0	19	11	4319073	4.12E-02	1	
RPS9	717623	458	-6	45639	14058	53163	2	0	0	50	41	7262574	4.12E-02	1	
NUAK2	664542	178	30	144804	44385	144837	1	1	0	50	28	10476708	4.12E-02	1	
PAX7	520847	598	28	130416	38973	105831	4	0	0	23	35	5204496	4.13E-02	1	
FBXO40	289905	613	25	166749	45309	67419	2	0	1	50	41	12737835	4.13E-02	1	
ZNF584	2179183	270	-4	100122	26400	92961	1	0	1	50	34	8286366	4.14E-02	1	
BCAT2	1700740	181	39	93291	27588	152955	3	0	0	50	38	10650255	4.15E-02	1	

gene	expr (expression level of this gene, averaged across 91 cell lines)	reptime (DNA replication time of this gene)	hic (chromatin state of this gene)	N_nonsilent (total number of nonsilent mutations)	N_silent (total number of silent mutations)	N_noncoding (total noncoding mutations)	n_nonsilent (Number of nonsilent coding mutations)	n_silent (Number of silent coding mutations)	n_noncoding	nnei (number of neighboring genes)	x (number of mutated bases in the neighboring genes)	X (total number of bases related to the neighboring genes)	pvalue	qvalue	Comment
DAAM1	156241	405	32	263901	66462	447282	2	0	1	26	18	6816843	4.16E-02	1	
HNRNPH1	1600541	229	28	111012	27588	202653	1	0	0	45	34	9163407	4.17E-02	1	
DST	403209	446	23	1617792	410520	1886049	13	3	4	17	15	6311151	4.17E-02	1	
MEF2B	1260413	215	35	97614	25245	82962	1	0	1	50	33	7855914	4.17E-02	1	
MRGPRG	NaN	NaN	NaN	62436	23496	0	2	0	0	50	87	22653345	4.18E-02	1	
FAM183A	1068006	294	30	33594	8085	40887	1	1	0	6	12	1704516	4.18E-02	1	
GABPB2	1614665	165	40	105237	31086	169884	2	0	0	50	41	11451231	4.19E-02	1	
RPUSD4	601483	341	26	89133	25905	150678	2	0	0	10	1	2349666	4.19E-02	1	
BRMS1	1963418	182	55	72567	18414	177210	1	0	2	16	10	3195093	4.19E-02	1	
GRWD1	1418188	209	41	103851	31284	84744	1	0	1	37	42	9941217	4.19E-02	1	
DOK3	1120311	347	41	137313	47124	155232	1	0	0	50	31	10521258	4.19E-02	1	
ITGA4	233378	518	20	252417	65076	473418	3	1	2	50	45	10723812	4.20E-02	1	
CIAO1	904852	213	43	81180	22176	126423	2	0	0	17	9	3927957	4.20E-02	1	
ANKRD30B	NaN	NaN	NaN	344190	83391	213840	5	1	0	50	88	22927080	4.21E-02	1	
RUSC1	1668953	177	44	214797	70521	306603	3	0	3	50	35	9246930	4.21E-02	1	
UBXN11	1021101	217	22	124377	36102	203544	1	0	0	40	18	6773910	4.21E-02	1	
OR7A10	963401	546	14	70653	21417	6237	1	0	0	50	47	8932671	4.21E-02	1	
ADIPOR1	421206	389	30	89067	25377	148896	2	0	0	50	38	11299266	4.22E-02	1	
FAM187B	767246	472	7	85800	24882	21780	1	0	0	50	38	9104766	4.25E-02	1	
RABGGTA	915114	220	38	134508	40722	166815	2	0	0	50	29	9879969	4.25E-02	1	
CD1B	128145	833	4	78705	22473	135828	3	0	2	38	39	7754043	4.25E-02	1	
TGM4	466410	380	31	165066	43725	351846	1	0	1	50	25	10787238	4.25E-02	1	
CLNK	214087	1111	17	109164	25377	185526	3	0	4	17	12	2032404	4.25E-02	1	
ZNF713	757446	584	-6	104841	24552	63558	1	0	0	35	25	7379163	4.26E-02	1	
SIX3	133317	566	23	74712	24585	13761	2	0	0	50	29	8858553	4.27E-02	1	
RET	454068	563	-19	263175	78474	298485	5	2	1	26	18	6053322	4.28E-02	1	
DMWD	929036	203	43	150645	51612	86427	1	0	1	49	28	10525284	4.28E-02	1	
DTNB	367255	356	44	154440	39897	222354	2	0	1	20	19	4068768	4.29E-02	1	
ARHGAP8	NaN	NaN	NaN	116391	33297	169884	3	1	1	5	7	1611159	4.30E-02	1	
RALGAPB	531076	440	45	357060	98439	657954	3	1	1	8	10	2367024	4.30E-02	1	
PLEK	208260	586	23	86394	21021	214830	2	0	0	41	22	8739654	4.31E-02	1	
DNASE1L1	2277674	NaN	40	71247	21516	116325	1	0	1	43	41	6493212	4.31E-02	1	
ACTG2	1036341	284	45	90519	24816	150975	2	1	0	50	39	12268608	4.33E-02	1	
HCLS1	289905	613	25	118272	31317	268488	2	0	0	24	14	7151364	4.33E-02	1	
NRSN1	349698	825	-21	45573	13431	46629	2	0	1	50	44	8010783	4.34E-02	1	
TMX2	922344	354	11	71379	19800	214434	2	0	0	27	15	4568949	4.35E-02	1	
NAT8	876583	590	16	51942	16170	17127	2	0	0	50	49	10737837	4.36E-02	1	
KCNJ6	253563	541	36	99561	27555	76329	2	0	0	50	30	10546206	4.37E-02	1	
SAGE1	246199	NaN	26	218031	58080	543213	3	1	2	9	6	1918224	4.37E-02	1	
PBX4	818567	284	-1	88374	25872	152559	1	0	0	33	24	6561687	4.37E-02	1	
DPEP1	1092575	211	38	97119	28809	110781	2	0	0	50	22	8042100	4.37E-02	1	
SHISA2	96652	584	12	67386	21120	10494	2	0	0	50	48	11044506	4.37E-02	1	
OR10G2	658946	624	20	69234	23034	11880	2	0	0	50	46	10861587	4.39E-02	1	
TTN	172693	713	23	8594982	2438271	4496283	69	13	25	12	47	9463311	4.40E-02	1	
PENK	411547	624	-16	63723	16863	40788	3	0	0	50	55	9494826	4.40E-02	1	
AADACL2	91028	611	-5	95073	25905	85635	1	0	0	50	72	13951806	4.40E-02	1	
CCR8	159083	507	38	83061	23067	12573	2	0	0	50	36	12743544	4.41E-02	1	
ARMC3	164960	449	12	210342	55968	283536	2	0	1	12	10	3575187	4.41E-02	1	
LPPR4	100692	868	-15	177342	51942	132264	4	0	1	25	27	6136647	4.43E-02	1	
TAF15	620019	235	8	143781	38478	277794	1	0	0	33	22	9943824	4.44E-02	1	
R3HDM1	561193	558	20	264660	71544	396297	3	0	0	13	7	3274953	4.44E-02	1	
MNDA	97345	985	2	98043	25509	123255	3	0	1	50	45	9281514	4.44E-02	1	
USP38	182738	678	3	243474	69861	160479	2	1	0	50	49	11486805	4.45E-02	1	
AKAP4	56301	NaN	5	202884	53031	97218	3	0	0	45	65	11746482	4.45E-02	1	
CNKSR1	896535	217	30	169455	50721	277299	4	0	1	50	45	11312763	4.45E-02	1	
SAMSN1	157937	898	56	91740	22308	145035	2	1	0	50	23	9401436	4.45E-02	1	
AP1G1	460697	NaN	24	199881	54549	482724	2	0	2	3	3	1885356	4.46E-02	1	
SPAST	378148	197	28	147345	42240	267993	1	0	0	50	41	13517526	4.46E-02	1	
ITGB3	965617	188	19	188760	51315	309375	2	0	1	50	27	10033353	4.47E-02	1	
ZNF585B	248427	984	-56	183876	46002	88605	3	0	0	32	23	4264755	4.47E-02	1	
MRP63	447305	240	43	24024	6963	12573	1	0	0	50	28	11035035	4.47E-02	1	
RFWD3	577167	151	31	183117	51810	332145	4	1	0	50	38	13143966	4.47E-02	1	
ETV1	46434	989	-16	123255	29700	189288	2	0	0	3	1	834306	4.48E-02	1	
RARB	216662	812	-14	107580	28545	171567	1	0	1	37	23	6351675	4.48E-02	1	
PHOX2B	308280	376	22	71082	23265	41481	2	0	1	1	2	581790	4.49E-02	1	
RC3H2	407115	430	26	288288	81180	350559	3	0	1	28	19	6805887	4.51E-02	1	
SLC26A5	355264	620	17	180906	51348	577863	2	1	1	40	36	8507400	4.51E-02	1	
BCLAF1	286851	153	21	219351	58542	176418	6	1	1	25	26	6522054	4.51E-02	1	
ARHGAP31	265987	585	23	333564	99957	237699	2	1	2	50	29	11527197	4.52E-02	1	
GIPR	950605	194	45	109989	33858	169092	2	0	1	50	25	10107834	4.52E-02	1	
RASGRF2	214755	766	29	298287	79695	490644	3	0	1	28	23	7597326	4.52E-02	1	
FTCD	742075	538	10	126258	40458	145233	1	0	0	45	31	9059721	4.52E-02	1	
SHROOM3	570950	418	17	454443	142626	295119	5	0	1	50	24	11541816	4.53E-02	1	
PANK3	413558	432	32	88110	24453	137016	1	0	0	33	36	7001412	4.54E-02	1	
HYAL3	843769	206	51	95172	30162	144738	2	0	1	50	24	9486708	4.54E-02	1	
PPP1R8	717145	218	26	82962	24156	164043	2	0	0	39	33	10784235	4.56E-02	1	
DPF2	4112809	242	50	95964	24618	191565	1	0	0	27	17	4643199	4.57E-02	1	
ZSWIM2	278175	629	-5	153417	38247	143352	2	1	1	19	16	3823281	4.57E-02	1	
ULK3	764533	195	46	113619	32802	129888	1	0	1	50	43	11111595	4.59E-02	1	
FBXL2	496632	352	15	103389	28479	264627	1	0	1	16	15	5461665	4.59E-02	1	
ELOVL3	1041024	187	45	64812	16863	94347	2	0	0	50	51	13762914	4.59E-02	1	

gene	expr (expression level of this gene, averaged across 91 cell lines)	reptime (DNA replication time of this gene)	hic (chromatin state of this gene)	N_nonsilent (total number of nonsilent mutations)	N_silent (total number of silent mutations)	N_noncoding (total noncoding mutations)	n_nonsilent (Number of nonsilent coding mutations)	n_silent (Number of silent coding mutations)	n_noncoding	nnei (number of neighboring genes)	x (number of mutated bases in the neighboring genes)	X (total number of bases related to the neighboring genes)	pvalue	qvalue	Comment
BPHL	444075	435	41	69498	19800	181764	1	0	0	14	12	2786223	4.60E-02	1	
OR2B3	155147	738	1	72831	20823	13959	2	0	0	50	49	11674542	4.60E-02	1	
SH3TC2	350600	254	15	300828	88341	366894	2	2	1	50	35	11788062	4.61E-02	1	
ZNF646	1398193	247	40	418605	126588	24156	4	0	0	22	18	4718010	4.62E-02	1	
PPM1H	298536	426	0	120087	36432	209583	1	0	0	42	41	11235378	4.62E-02	1	
C22orf39	650407	300	16	37719	10791	48312	1	0	0	50	32	12676356	4.64E-02	1	
OR1S1	637499	828	-25	74613	22110	11583	3	0	0	50	28	6430974	4.64E-02	1	
LUZP2	1913	1282	-32	86031	21384	189783	2	0	4	31	63	6639402	4.65E-02	1	
CREB3L4	2028800	222	34	93522	27654	131472	1	0	0	23	10	3401706	4.65E-02	1	
UBR7	325787	222	35	107052	23430	142857	1	0	2	50	40	9672630	4.70E-02	1	
TP53I13	760947	353	26	88407	31185	66231	1	0	1	16	11	3825756	4.71E-02	1	
WSCD2	451405	644	29	133518	37356	134739	6	0	1	26	44	5961747	4.71E-02	1	
DND1	774470	356	31	80223	26301	119691	2	0	0	31	27	7515123	4.71E-02	1	
WDR78	334231	692	-14	209649	54879	323730	3	0	0	4	2	740091	4.72E-02	1	
IRX4	387770	635	11	116655	39369	57618	3	0	0	50	45	10502943	4.73E-02	1	
LILRA2	819283	601	-4	112860	33858	147312	1	0	0	50	54	11953491	4.74E-02	1	
WSCD1	283312	593	-11	133254	40590	159588	3	0	0	21	13	5277063	4.75E-02	1	
DYX1C1	298473	405	1	112959	25245	275913	1	0	1	21	14	5545485	4.75E-02	1	
TFAP2C	280847	442	-44	104742	31581	130482	1	0	0	27	26	6350223	4.77E-02	1	
TGFBR3	284037	252	1	202389	56595	318384	2	1	0	47	23	11291742	4.77E-02	1	
GABPA	534601	728	31	110154	28545	175923	1	0	0	27	10	3983331	4.79E-02	1	
PSEN2	428197	373	1	106656	30657	205029	1	1	1	50	46	9999495	4.79E-02	1	
SLC39A12	94785	360	-9	164934	45342	230175	5	1	2	11	8	2136882	4.80E-02	1	
STK31	367849	441	14	251394	61050	535095	2	0	1	22	12	4575945	4.80E-02	1	
CHFR	944795	674	10	167475	46365	268290	5	0	2	50	40	10839642	4.81E-02	1	
ATP11A	585697	416	45	294756	78870	588654	3	0	1	7	5	2061081	4.81E-02	1	
MAP2K3	415581	517	-43	85173	22539	245322	2	1	1	50	33	9786150	4.82E-02	1	
P2RY10	14218	NaN	50	79596	21780	16335	2	0	0	50	28	9215415	4.82E-02	1	
KIF1A	1094176	446	29	405372	115071	436194	3	1	4	44	44	10374243	4.82E-02	1	
KRT27	836683	269	-10	108372	31218	110484	3	1	0	50	33	8393055	4.82E-02	1	
LRP3	406033	270	-8	175098	56463	104247	2	0	0	17	9	5002107	4.83E-02	1	
DYNC1LI2	804191	145	10	118503	32868	238689	1	0	0	17	13	4696296	4.86E-02	1	
OSBPL2	990360	252	9	118470	29139	272052	1	1	0	28	17	5464503	4.86E-02	1	
OR52I1	212267	935	-3	73656	22770	16632	3	0	0	50	56	11358732	4.86E-02	1	
HGFAC	547964	412	46	153549	46827	131175	2	0	0	25	21	6526410	4.86E-02	1	
ADCYAP1R1	364216	631	7	115434	29997	310365	1	2	1	50	41	9698733	4.89E-02	1	
CUEDC1	873797	222	6	92301	26202	154836	2	0	0	45	25	9021210	4.89E-02	1	
XPNPEP2	358809	NaN	48	163581	44814	391545	2	0	0	3	2	1265847	4.91E-02	1	
TIMM44	960059	320	23	109824	29469	187209	1	0	1	28	15	5414211	4.91E-02	1	
MMP17	364888	206	17	139260	43890	93951	4	1	0	50	35	10174890	4.91E-02	1	
CASQ2	287179	639	8	99792	22968	210474	3	0	1	50	37	10069752	4.91E-02	1	
SLC35D1	387355	673	-4	85338	24750	172161	1	1	1	50	38	7091601	4.92E-02	1	
ACTL6B	1884430	247	60	103587	28380	161469	2	0	2	4	4	1022637	4.92E-02	1	
RHOBTB2	740323	241	48	176913	51282	161568	2	0	1	50	44	12230295	4.93E-02	1	
SMEK2	597106	382	25	206151	52635	264528	1	1	0	45	18	9607191	4.93E-02	1	
PPP1R3C	276001	206	16	75471	19371	22671	2	0	0	50	24	7899804	4.93E-02	1	
VPS36	365031	318	37	95733	24552	213444	1	0	0	19	20	4794504	4.93E-02	1	
DBF4	202733	758	14	164934	39897	214335	2	0	2	11	9	2501664	4.94E-02	1	
CHCHD6	250457	604	39	57420	15642	187011	1	1	1	7	4	1586178	4.95E-02	1	
PPP2R3C	640994	321	45	113355	26235	248886	1	0	0	23	21	5761965	4.95E-02	1	
IGFN1	572258	333	23	850377	260304	199386	11	1	1	38	16	7819449	4.95E-02	1	
AGPAT9	186399	436	21	104742	28809	251559	1	0	0	25	17	5488329	4.96E-02	1	
OR5B2	475497	835	-29	71643	20823	13662	3	0	0	50	45	6548916	4.97E-02	1	
EBAG9	134329	623	8	52866	13068	135531	2	0	0	12	14	3192948	4.97E-02	1	
NELL2	79685	930	-19	220902	54120	352242	5	1	2	6	10	1528461	4.98E-02	1	
MUC21	2483101	261	36	126324	42867	56727	2	0	0	50	23	9463311	5.00E-02	1	

B.1.2 Primary tumor mutated genes

gene	expr (expression level of this gene, averaged across 91 cell lines)	reptime (DNA replication time of this gene)	hic (chromatin state of this gene)	N_nonsilent (total number of nonsilent mutations)	N_silent (total number of silent mutations)	N_noncoding (total noncoding mutations)	n_nonsilent (Number of nonsilent coding mutations)	n_silent (Number of silent coding mutations)	n_noncoding	nnei (number of neighboring genes)	x (number of mutated bases in the neighboring genes)	X (total number of bases related to the neighboring genes)	pvalue	qvalue
TP53	2069567	213	34	92670	25590	0	16	1	0	50	6	1554750	4.91E-12	0.00
STK11	2214581	234	29	94380	26040	0	6	0	0	50	4	1726320	1.44E-06	0.01
KRAS	259193	512	16	51420	12390	0	8	0	0	50	5	1729200	2.24E-05	0.14
KEAP1	1996822	207	30	132300	38250	0	5	0	0	50	0	1609110	3.07E-04	1.00
RNGTT	202476	672	37	133950	32910	0	3	0	0	50	6	1783860	6.96E-04	1.00
TPRX1	881316	336	39	80880	30900	0	3	0	0	50	5	1896480	1.95E-03	1.00
CPM	645106	228	28	97710	25050	0	2	0	0	50	2	1965330	2.12E-03	1.00
SPATA22	365088	515	14	80910	20250	0	4	0	0	50	8	1985640	2.81E-03	1.00
MESP2	934542	212	37	80070	27750	0	2	0	0	50	4	1767960	2.90E-03	1.00
SFXN2	878626	180	27	71880	19290	0	2	0	0	50	5	1713720	3.06E-03	1.00
MAFA	2154403	237	32	71970	23610	0	2	0	0	50	6	1838130	3.08E-03	1.00
C4orf36	268001	603	36	26190	7110	0	2	0	0	50	5	1365180	3.23E-03	1.00
TGIF2LX	8315	NaN	-35	51210	14670	0	3	0	0	50	21	2031900	3.43E-03	1.00
KRT76	1444461	422	38	136500	39090	0	3	0	0	50	2	1885470	3.49E-03	1.00
RNASE10	464497	395	17	46170	12330	0	2	0	0	50	4	1802160	4.51E-03	1.00
OR10K2	135429	902	5	64080	20250	0	5	1	0	50	12	1390710	5.11E-03	1.00
OLFM4	207557	842	2	108690	30720	0	4	1	0	50	4	1750740	5.34E-03	1.00
FAM183A	1068006	294	30	30540	7350	0	2	0	0	50	9	1710630	5.49E-03	1.00
RAB3B	701660	353	22	47790	12870	0	2	0	0	50	7	2165040	5.99E-03	1.00
FRMPD4	191024	NaN	22	282390	80580	0	7	0	0	45	1	1508100	6.24E-03	1.00
POLDIP2	1405196	214	10	80100	23040	0	2	0	0	50	10	1896060	6.36E-03	1.00
GFRA4	942842	353	32	63450	18990	0	2	0	0	50	3	1795380	6.46E-03	1.00
CLRN2	271399	912	37	49170	14820	0	2	0	0	50	4	1483380	7.33E-03	1.00
CHST11	1030269	419	34	75540	20490	0	2	0	0	50	6	1260720	7.80E-03	1.00
FGB	59883	1034	20	108990	26550	0	4	0	0	50	6	1611660	7.97E-03	1.00
CNKSR1	896535	217	30	154050	46110	0	4	0	0	50	7	2225460	9.47E-03	1.00
GATA6	230941	437	-1	121620	41460	0	2	0	0	50	8	1839930	1.07E-02	1.00
FGF23	295315	860	33	52890	15870	0	2	0	0	50	5	1444080	1.13E-02	1.00
LASP1	1255539	159	41	58920	13980	0	2	0	0	50	7	1908780	1.18E-02	1.00
SERPINF1	1447501	189	32	89190	26460	0	2	0	0	50	3	1818510	1.22E-02	1.00
IRF5	845100	361	49	109560	32370	0	2	0	0	50	15	2133870	1.30E-02	1.00
COL4A1	429137	434	12	355740	113520	0	8	1	0	50	15	2283390	1.31E-02	1.00
TMPRSS13	527704	576	45	121590	36270	0	2	0	0	50	8	1728210	1.35E-02	1.00
FAM53A	902379	350	33	81870	26940	0	2	0	0	50	4	1788150	1.36E-02	1.00
STC2	471196	519	41	64650	18240	0	2	0	0	50	8	1975170	1.49E-02	1.00
MYO19	416822	556	23	214860	60810	0	5	1	0	50	8	2328660	1.49E-02	1.00
NOP10	542657	234	22	14250	3840	0	1	0	0	50	7	1587990	1.61E-02	1.00
OR51Q1	99042	1030	1	65670	20550	0	4	0	0	50	12	1670160	1.66E-02	1.00
LAMA2	67993	771	-6	683280	182970	0	8	1	0	50	5	2244420	1.66E-02	1.00
OGFRL1	132568	459	-14	98820	25380	0	2	0	0	50	14	1695150	1.68E-02	1.00
NFASC	388939	440	33	298770	83370	0	8	1	0	50	5	1684920	1.73E-02	1.00
RGAG4	582710	NaN	53	119190	34710	0	2	1	0	50	7	1577940	1.75E-02	1.00
TYMP	1075525	207	42	99990	34380	0	3	1	0	50	10	1804890	1.86E-02	1.00
GLT8D2	1193959	281	37	77970	19770	0	2	0	0	50	5	1621200	1.86E-02	1.00
CKS1B	1612870	224	35	18270	4050	0	1	0	0	50	5	1558020	1.89E-02	1.00
UGT3A1	155890	614	-4	117720	33120	0	4	0	0	50	11	1919850	1.89E-02	1.00
ANAPC13	254024	188	16	16980	3990	0	1	0	0	50	6	1512930	1.96E-02	1.00
NPPC	893534	439	33	25410	9240	0	1	0	0	50	4	1887300	2.01E-02	1.00
CCDC39	182757	1031	2	212040	49140	0	5	0	0	41	6	1240950	2.03E-02	1.00
NAAA	483325	429	44	78360	23070	0	2	0	0	50	8	2033250	2.07E-02	1.00
B3GALT1	67289	1010	-14	69690	18600	0	2	0	0	50	9	1879680	2.08E-02	1.00
CENPW	142596	704	-1	19140	5610	0	1	0	0	50	9	1784250	2.13E-02	1.00
S100A2	1454773	186	26	21810	5370	0	1	0	0	50	5	1632870	2.17E-02	1.00
KIAA0040	NaN	NaN	NaN	21510	5490	0	1	0	0	50	21	3346410	2.18E-02	1.00
EEFSEC	858998	410	48	125100	38430	0	2	0	0	28	1	1277910	2.19E-02	1.00
IL7R	139461	511	5	100560	26340	0	4	1	0	50	11	1529070	2.20E-02	1.00
KRTAP19-8	200531	972	-45	13920	3720	0	1	0	0	50	11	1462380	2.21E-02	1.00
QRSL1	227264	174	40	114660	31770	0	2	0	0	50	2	1419240	2.21E-02	1.00
GALP	500375	480	-61	27150	6240	0	2	0	0	50	19	2389110	2.22E-02	1.00
NPNT	262174	361	7	133650	36540	0	3	0	0	50	3	1540290	2.23E-02	1.00
LRRC69	NaN	NaN	NaN	76650	20010	0	3	0	0	50	21	3360930	2.24E-02	1.00
DCDC1	84163	938	-25	77610	20760	0	8	1	0	8	3	247950	2.26E-02	1.00
ZNF257	75568	1305	-69	123600	29760	0	4	0	0	46	16	1329420	2.26E-02	1.00
ADAMTS2	1496298	677	-4	265320	76230	0	4	3	0	50	21	2188800	2.27E-02	1.00

gene	expr (expression level of this gene, averaged across 91 cell lines)	reptime (DNA replication time of this gene)	hic (chromatin state of this gene)	N_nonsilent (total number of nonsilent mutations)	N_silent (total number of silent mutations)	N_noncoding (total noncoding mutations)	n_nonsilent (Number of nonsilent coding mutations)	n_silent (Number of silent coding mutations)	n_noncoding	nnei (number of neighboring genes)	x (number of mutated bases in the neighboring genes)	X (total number of bases related to the neighboring genes)	pvalue	qvalue
ZNF556	1287128	291	27	99030	25620	0	2	0	0	50	5	1833330	2.30E-02	1.00
C9orf153	311398	417	8	23070	5730	0	1	0	0	50	7	1791630	2.38E-02	1.00
MMP24	1290513	170	45	121020	33960	0	2	0	0	50	4	1955970	2.40E-02	1.00
FAM122C	293030	NaN	27	83130	22260	0	2	0	0	50	10	2254710	2.42E-02	1.00
CRP	737948	602	-2	47640	13470	0	3	0	0	50	21	1945950	2.45E-02	1.00
ACTRT1	938	NaN	-35	78600	23190	0	6	0	0	50	21	2031900	2.45E-02	1.00
PALM	1914670	200	23	82470	25170	0	3	0	0	50	9	1833660	2.46E-02	1.00
EPS15	568927	251	11	206970	52680	0	2	0	0	50	10	2305590	2.49E-02	1.00
DIO2	47674	533	-44	65550	19410	0	2	0	0	50	10	1790580	2.51E-02	1.00
ATAD1	335357	169	45	80250	20730	0	2	0	0	50	6	1538640	2.52E-02	1.00
GNRH2	835590	288	54	25650	8280	0	1	0	0	50	4	1473780	2.60E-02	1.00
APLF	227462	617	22	113070	28410	0	4	0	0	50	3	1522830	2.60E-02	1.00
SLC7A13	192785	1147	3	98190	30240	0	2	0	0	50	4	1489320	2.63E-02	1.00
DDX59	371965	276	13	133410	36510	0	2	2	0	34	7	1244190	2.64E-02	1.00
FAM136A	671278	370	38	30960	7470	0	2	0	0	50	6	2155140	2.65E-02	1.00
TRHR	151352	616	18	84360	23910	0	2	0	0	50	5	1411650	2.66E-02	1.00
FAM13C	121813	707	-15	129330	33750	0	2	1	0	24	7	1240050	2.66E-02	1.00
SV2B	457413	624	1	149850	38790	0	3	1	0	50	13	1714740	2.67E-02	1.00
BCLAF1	286851	153	21	199410	53220	0	6	1	0	50	7	2176560	2.74E-02	1.00
WAS	824559	NaN	27	105720	34230	0	2	0	0	10	0	705810	2.76E-02	1.00
SERPINB8	114875	671	26	83640	20760	0	3	0	0	50	14	1656630	2.78E-02	1.00
NFE2L2	372300	145	36	130590	34470	0	3	0	0	50	6	1607550	2.78E-02	1.00
OR2J2	162844	808	12	64560	20040	0	2	1	0	50	10	1581120	2.80E-02	1.00
VAV3	132139	554	-11	193650	47100	0	2	0	0	50	6	1810650	2.84E-02	1.00
UBQLN4	2005405	213	39	127800	38520	0	2	0	0	50	4	1834410	2.85E-02	1.00
PITPNM3	255838	313	8	208920	61170	0	3	1	0	50	3	1750260	2.87E-02	1.00
DTX1	562320	387	44	128970	41580	0	2	0	0	50	4	1839990	2.87E-02	1.00
IQCF6	NaN	NaN	NaN	22860	6660	0	2	0	0	50	21	3347580	2.88E-02	1.00
C2orf48	881918	220	41	34470	9810	0	1	0	0	50	6	2067060	2.89E-02	1.00
NRN1L	879834	158	50	34080	11640	0	1	0	0	50	6	2002650	2.91E-02	1.00
RNF128	149556	NaN	17	119940	34770	0	2	0	0	50	4	1729140	2.92E-02	1.00
ZNF720	NaN	NaN	NaN	28950	6780	0	1	0	0	50	21	3347700	2.95E-02	1.00
SIVA1	711772	247	36	37410	11370	0	1	0	0	50	2	1586850	2.97E-02	1.00
CADM1	29879	993	5	95580	27270	0	3	0	0	50	15	2404260	2.98E-02	1.00
OR6K6	103627	985	1	71760	21120	0	5	0	0	50	11	1755810	3.01E-02	1.00
INHBA	42008	671	-27	90540	25110	0	3	0	0	50	8	1554330	3.01E-02	1.00
HLA-DRB1	1074109	595	29	57540	16530	0	2	0	0	50	12	1529160	3.05E-02	1.00
AMPH	167080	639	7	152670	42630	0	3	0	0	50	6	1828950	3.08E-02	1.00
ACOT13	421216	362	24	30960	9090	0	1	0	0	50	6	1677390	3.10E-02	1.00
ST8SIA3	298355	599	9	81870	22080	0	4	0	0	50	10	1926000	3.10E-02	1.00
TMEM37	203559	497	35	39450	12660	0	1	0	0	50	3	1753200	3.11E-02	1.00
KCNH4	2129732	206	30	211920	68340	0	2	0	0	46	6	1738020	3.13E-02	1.00
C17orf97	248876	591	1	20430	7290	0	2	0	0	50	9	1526670	3.17E-02	1.00
HOXD9	242935	874	15	72000	23850	0	2	0	0	50	10	1641570	3.29E-02	1.00
E2F4	1109491	150	46	89190	26010	0	2	0	0	50	7	1607460	3.30E-02	1.00
ADM	860704	273	12	39030	12270	0	2	0	0	50	5	2037090	3.33E-02	1.00
CERCAM	1928973	188	64	126840	38220	0	3	0	0	50	6	2002530	3.34E-02	1.00
SEMA5A	45422	891	-2	234300	63510	0	8	0	0	50	8	1804170	3.39E-02	1.00
SIRPD	422390	424	46	42090	12630	0	2	0	0	50	3	1758540	3.42E-02	1.00
MPV17	1245306	182	52	38880	11430	0	1	0	0	50	6	1907820	3.44E-02	1.00
ZNF285	715346	528	-44	129540	31110	0	5	1	0	50	9	1496160	3.57E-02	1.00
CLNK	214087	1111	17	99240	23070	0	2	0	0	50	6	1565970	3.59E-02	1.00
TCEAL4	490345	NaN	31	48150	10890	0	1	0	0	50	3	1863570	3.60E-02	1.00
PSME2	974559	220	41	55020	13740	0	1	0	0	50	3	2070150	3.65E-02	1.00
OR5D14	4183	1091	-36	65460	19500	0	2	0	0	50	13	1760160	3.66E-02	1.00
MAGEF1	1301642	485	36	64050	19110	0	1	0	0	50	2	2218980	3.66E-02	1.00
C7orf69	NaN	NaN	NaN	27480	6630	0	2	0	0	50	21	3347550	3.66E-02	1.00
TMEM132C	NaN	NaN	NaN	231120	71190	0	9	2	0	50	23	3412110	3.67E-02	1.00
TMEM216	1394229	286	57	32430	9600	0	1	0	0	50	6	1510200	3.69E-02	1.00
C2orf78	1124858	404	43	191640	58290	0	4	0	0	50	7	1677330	3.71E-02	1.00
MID1IP1	91133	NaN	31	38640	11040	0	1	0	0	50	7	1900410	3.71E-02	1.00
GAS2L3	197479	849	-12	147450	43260	0	2	0	0	50	15	1800750	3.73E-02	1.00
SUOX	2080397	229	31	112500	36000	0	2	0	0	50	6	1710000	3.76E-02	1.00
DCX	131531	NaN	13	96690	26160	0	2	0	0	50	8	1739040	3.82E-02	1.00

gene	expr (expression level of this gene, averaged across 91 cell lines)	reptime (DNA replication time of this gene)	hic (chromatin state of this gene)	N_nonsilent (total number of nonsilent mutations)	N_silent (total number of silent mutations)	N_noncoding (total noncoding mutations)	n_nonsilent (Number of nonsilent coding mutations)	n_silent (Number of silent coding mutations)	n_noncoding	nnei (number of neighboring genes)	x (number of mutated bases in the neighboring genes)	X (total number of bases related to the neighboring genes)	pvalue	qvalue
TPX2	890615	237	23	165210	42510	0	2	0	0	50	1	1236300	3.86E-02	1.00
FSCN3	411207	619	-7	106320	30390	0	4	0	0	50	4	1268130	3.87E-02	1.00
PIP5K1A	1409327	184	32	123240	34170	0	2	0	0	50	6	1825080	3.92E-02	1.00
LILRA6	717623	458	-6	101970	30690	0	5	2	0	50	11	1519230	3.93E-02	1.00
ABCB5	164218	648	4	277020	75690	0	5	0	0	14	0	580380	3.94E-02	1.00
SMARCA4	1984564	190	40	364680	102330	0	4	0	0	50	2	1875030	3.95E-02	1.00
PDCD1	559742	487	29	60450	19200	0	2	0	0	50	3	2176230	4.02E-02	1.00
TTR	294098	541	20	31290	9750	0	1	0	0	50	13	1871130	4.02E-02	1.00
SLC41A3	210410	610	32	127440	40230	0	2	0	0	38	3	1347270	4.03E-02	1.00
MAG	767180	472	8	134250	41790	0	3	0	0	50	6	1417530	4.05E-02	1.00
XPNPEP2	358809	NaN	48	148710	40740	0	2	0	0	50	5	1953540	4.06E-02	1.00
TCTE3	233648	447	3	44220	10950	0	1	0	0	50	7	1982550	4.07E-02	1.00
MAL	171487	647	-1	32610	10230	0	3	0	0	50	10	1648200	4.08E-02	1.00
SPOCK3	7931	1043	-11	99180	22590	0	4	0	0	50	10	1982340	4.08E-02	1.00
LIN28A	1046693	181	33	45420	12360	0	1	0	0	50	3	1528410	4.08E-02	1.00
TOP3A	941112	223	13	216990	60030	0	2	0	0	50	10	2035800	4.10E-02	1.00
PLSCR2	122649	611	-16	50430	12480	0	1	0	0	50	9	2355930	4.14E-02	1.00
KCNQ5	648013	975	15	202410	59670	0	2	1	0	50	8	1673790	4.16E-02	1.00
HCN1	80822	1235	-24	188190	55080	0	5	1	0	50	11	1679400	4.17E-02	1.00
PLOD3	1022626	258	39	159510	46500	0	4	0	0	50	11	1660560	4.17E-02	1.00
CD7	2856070	193	24	48810	17340	0	1	0	0	50	4	1673100	4.25E-02	1.00
KRTAP21-2	30244	1055	-70	17670	4830	0	1	0	0	50	18	1216110	4.26E-02	1.00
UBXN7	1165640	231	37	108210	27870	0	2	0	0	50	5	1401510	4.27E-02	1.00
THY1	1054477	263	40	34290	10530	0	1	0	0	50	14	1978110	4.33E-02	1.00
ALMS1	777501	590	22	880020	253260	0	9	1	0	35	6	1611060	4.35E-02	1.00
SRI	157359	454	23	45780	11550	0	1	0	0	50	4	1533240	4.37E-02	1.00
AP1G2	1250373	553	35	167490	52290	0	2	0	0	50	5	2143230	4.41E-02	1.00
MMP3	468929	506	-21	104520	27960	0	2	0	0	50	12	2256900	4.43E-02	1.00
TNFRSF4	2356409	196	27	58980	18420	0	1	0	0	50	6	2254080	4.47E-02	1.00
FAM161A	758678	249	34	143820	36630	0	2	0	0	50	4	1614750	4.50E-02	1.00
PAEP	327501	539	37	40830	10560	0	1	0	0	50	7	1590510	4.51E-02	1.00
RBM10	819569	NaN	38	203160	56130	0	2	0	0	50	1	1396890	4.51E-02	1.00
ALDH1B1	272380	622	3	107130	33090	0	2	0	0	50	7	1612920	4.56E-02	1.00
SERTAD1	1010633	202	34	47730	16440	0	2	0	0	50	3	1525590	4.57E-02	1.00
KCNMB2	144490	882	-8	51450	13710	0	2	0	0	50	26	2557530	4.63E-02	1.00
VAMP4	415434	718	3	33480	7380	0	1	0	0	50	12	1646910	4.64E-02	1.00
C6orf223	1407986	199	47	56040	14700	0	1	0	0	50	6	1959930	4.68E-02	1.00
GMNN	422850	362	40	46980	11880	0	1	0	0	50	9	2033520	4.69E-02	1.00
DCLK3	374894	617	-14	138540	37770	0	3	1	0	50	5	1303440	4.70E-02	1.00
UCHL3	197113	362	37	52500	12930	0	1	0	0	50	5	1732170	4.72E-02	1.00
PAX7	520847	598	28	118560	35430	0	4	0	0	27	3	1114590	4.72E-02	1.00
GCH1	673378	380	40	61860	17070	0	1	0	0	50	6	2181480	4.73E-02	1.00
ZNF334	416360	642	19	150120	35190	0	4	0	0	50	11	1480080	4.74E-02	1.00
PLA1A	288657	595	42	99660	27510	0	4	0	0	50	6	1570710	4.75E-02	1.00
TMEM40	487514	176	35	53550	13590	0	1	0	0	50	4	1712040	4.77E-02	1.00
USP6	1391000	384	20	309180	81060	0	3	0	0	16	0	982350	4.79E-02	1.00
FUT9	76115	1328	-21	77580	19980	0	4	0	0	50	22	1709760	4.79E-02	1.00
MAGEC2	16386	NaN	-29	78600	22740	0	2	1	0	50	14	1840110	4.87E-02	1.00
PITX3	1020192	169	44	61890	21000	0	1	0	0	50	7	2184330	4.90E-02	1.00
GAL	780720	354	49	27120	7980	0	2	0	0	50	13	1807080	4.92E-02	1.00
KLK1	655673	589	25	57120	15690	0	1	0	0	50	7	2076750	4.93E-02	1.00
VSNL1	135361	1255	-2	42840	10080	0	1	0	0	50	7	1504530	4.94E-02	1.00
SLC9A1	886940	336	27	171480	52800	0	2	1	0	50	4	1795920	4.99E-02	1.00

B.2 MEV pathway enrichment analyses

B.2.1 KEGG results

MEVs: 2490 genes recognized by DAVID (116 not recognized by DAVID)
KEGG enrichment results (q < 0.05):

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
KEGG_PATHWAY	hsa04510:Focal adhesion	55	2.209	2.68E-07	FLT1, ITGB5, TNC, PIK3CD, ELK1, MYLK3, PIK3CG, ARHGAP35, IGF1R, MYLK, IBSP, CCND1, ITGB8, ITGAV, VAV3, ITGA4, HGF, ACTN1, PGF, VAV2, COL2A1, PIK3CA, COL4A1, COL6A2, MYL2, COL4A3, COL4A6, COL4A5, ITGA6, TLN2, COL6A5, DOCK1, CRK, SOS2, LAMA5, TNXB, LAMA2, LAMA1, COL11A1, LAMA4, LAMA3, THBS2, RELN, PDGFD, PDGFC, FLNC, PAK2, JUN, COL24A1, BRAF, COL1A1, COL1A2, COL5A1, COL5A2, BCAR1	907	206	6879	2.0249462	7.56E-05	3.79E-05	3.48E-05
KEGG_PATHWAY	hsa04512:ECM-receptor interaction	31	1.245	2.69E-07	LAMA5, TNXB, LAMA2, ITGB5, LAMA1, LAMA4, COL11A1, LAMA3, TNC, THBS2, SV2C, RELN, IBSP, DAG1, ITGB8, ITGAV, COL24A1, ITGA4, GP1BA, COL1A1, COL2A1, COL1A2, COL5A1, COL4A1, COL6A2, COL4A3, COL5A2, COL4A6, COL4A5, ITGA6, COL6A5	907	87	6879	2.7024674	7.58E-05	3.79E-05	3.48E-05
KEGG_PATHWAY	hsa05146:Amoebiasis	32	1.285	9.18E-06	SERPINB10, LAMA5, ITGAM, LAMA2, LAMA1, LAMA4, COL11A1, LAMA3, ITGB2, PIK3CD, C8A, PIK3CG, MUC2, GNA11, COL24A1, ACTN1, NFKB1, COL1A1, PLCB3, IL6, COL2A1, PLCB4, COL1A2, COL5A1, PIK3CA, COL4A1, COL4A3, COL5A2, COL4A6, COL4A5, PLCB2, RAB5A	907	106	6879	2.2896133	0.002586448	8.24E-04	7.57E-04
KEGG_PATHWAY	hsa04974:Protein digestion and absorption	28	1.124	1.31E-05	KCNK5, COL17A1, COL15A1, COL11A1, SLC3A2, SLC8A2, KCNN4, SLC6A19, COL24A1, COL22A1, SLC8A3, COL1A1, SLC9A3, COL2A1, COL1A2, COL5A1, COL4A1, KCNQ1, COL6A2, COL4A3, COL5A2, COL4A6, COL9A1, COL4A5, COL21A1, COL9A3, COL6A5, COL9A2	907	88	6879	2.4132004	0.003696736	8.24E-04	7.57E-04
KEGG_PATHWAY	hsa04810:Regulation of actin cytoskeleton	51	2.048	1.46E-05	CHRM2, CYFIP1, NCKAP1, CHRM3, ITGAM, ITGB5, ITGB2, PIK3CD, FGF1, MYLK3, PIK3CG, ARHGAP35, MYLK, FGF9, ITGAX, PIP4K2A, ITGB8, ITGAV, GIT1, VAV3, MAP2K2, ITGA4, ACTN1, VAV2, TIAM1, MRAS, PIK3CA, ITGAD, MYL2, ITGA6, DOCK1, CRK, SOS2, IQGAP1, IQGAP2, FGD1, FGD3, PDGFD, PDGFC, FGF20, PAK2, FGF22, BRAF, BAIAP2, SSH1, PIKFYVE, DIAPH2, FGF13, FGF12, BCAR1, FGFR1	907	210	6879	1.8419121	0.004112741	8.24E-04	7.57E-04
KEGG_PATHWAY	hsa04015:Rap1 signaling pathway	46	1.847	5.19E-04	CSF1R, FLT1, ITGAM, ITGB2, GNAI3, ADCY3, PIK3CD, FGF1, ADCY8, RASGRP2, FYB, PIK3CG, IGF1R, GRIN2A, SIPA1L1, FGF9, PDGFD, PDGFC, FGF20, PLCE1, PLCG1, FGF22, MAP2K2, HGF, BRAF, GRIN2B, PGF, GNAO1, APBB1IP, TIAM1, PLCB3, PLCB4, MRAS, ADCY9, PIK3CA, PARD3, ID1, TLN2, FGF13, RAPGEF5, FGF12, PLCB2, CRK, BCAR1, FGFR1, EPHA2	907	210	6879	1.6613325	0.136101201	0.024376951	0.02238876
KEGG_PATHWAY	hsa04151:PI3K-Akt signaling pathway	67	2.691	8.54E-04	CHRM2, FLT1, ITGB5, TNC, PIK3CD, FGF1, PIK3CG, IGF1R, GYS2, STK11, TCL1A, IBSP, FGF9, CCND1, PPP2R5E, CREB3L4, CREB3L2, ITGB8, ITGAV, JAK2, MAP2K2, ITGA4, CHUK, HGF, PGF, COL2A1, PIK3CA, COL4A1, COL6A2, COL4A3, COL4A6, COL4A5, ITGA6, COL6A5, TP53, SOS2, EPHA2, CREB5, CSF1R, LAMA5, TNXB, LAMA2, LAMA1, COL11A1, LAMA4, LAMA3, THBS2, RELN, RXRA, PDGFD, PDGFC, FGF20, FGF22, PCK2, COL24A1, PPP2R3A, MTOR, NFKB1, COL1A1, IL6, COL1A2, COL5A1, PPP2R2B, COL5A2, FGF13, FGF12, FGFR1	907	345	6879	1.4729016	0.214111001	0.03440526	0.031599157
KEGG_PATHWAY	hsa00562:Inositol phosphate metabolism	20	0.803	0.0016	MTMR1, MTMR2, ISYNA1, PIK3CD, PIK3C2G, OCRL, PIK3C2A, PIK3CG, MTMR7, INPP4B, PIKFYVE, PLCB3, PLCB4, PIK3CA, INPP5J, PIP4K2A, PLCE1, PIK3C3, PLCG1, PLCB2	907	71	6879	2.1364349	0.3629678	0.050063859	0.045980637
KEGG_PATHWAY	hsa05218:Melanoma	20	0.803	0.0016	RB1, MAP2K2, CDKN2A, HGF, PIK3CD, BRAF, FGF1, PIK3CG, IGF1R, PIK3CA, FGF9, CCND1, PDGFD, PDGFC, FGF20, FGF13, FGF12, TP53, FGF22, FGFR1	907	71	6879	2.1364349	0.3629678	0.050063859	0.045980637

B.2.2 Gene lists

Focal Adhesion (from DAVID - https://david.ncifcrf.gov/kegg.jsp?path=hsa04510$Focal%20adhesion&termId=550028744&source=kegg):

ACTB	COL24A1	FLNC	ITGB6	MYL9	PPP1R12C	SOS2
ACTG1	COL27A1	FLT1	ITGB7	MYLK	PRKCA	SPP1
ACTN1	COL2A1	FLT4	ITGB8	MYLK2	PRKCB	SRC
ACTN2	COL3A1	FN1	JUN	MYLK3	PRKCG	THBS1
ACTN3	COL4A1	FYN	KDR	MYLK4	PTEN	THBS2
ACTN4	COL4A2	GRB2	LAMA1	MYLPF	PTK2	THBS3
AKT1	COL4A3	GSK3B	LAMA2	PARVA	PXN	THBS4
AKT2	COL4A4	HGF	LAMA3	PARVB	RAC1	TLN1
AKT3	COL4A5	HRAS	LAMA4	PARVG	RAC1	TLN2
ARHGAP35	COL4A6	IBSP	LAMA5	PDGFA	RAC1	TNC
ARHGAP5	COL5A1	IGF1	LAMB1	PDGFB	RAC1	TNN
BAD	COL5A2	IGF1R	LAMB2	PDGFC	RAC1	TNR
BCAR1	COL5A3	ILK	LAMB3	PDGFD	RAC1	TNXB
BCL2	COL6A1	ITGA1	LAMB4	PDGFRA	RAC1	VASP
BIRC2	COL6A2	ITGA10	LAMC1	PDGFRB	RAC2	VAV1
BIRC3	COL6A3	ITGA11	LAMC2	PDPK1	RAC3	VAV2
BRAF	COL6A5	ITGA2	LAMC3	PGF	RAF1	VAV3
CAPN2	COL6A6	ITGA2B	MAP2K1	PIK3CA	RAP1A	VCL
CAV1	COMP	ITGA3	MAPK1	PIK3CB	RAP1B	VEGFA
CAV2	CRK	ITGA4	MAPK10	PIK3CD	RAPGEF1	VEGFB
CAV3	CRKL	ITGA5	MAPK3	PIK3CG	RASGRF1	VEGFC
CCND1	CTNNB1	ITGA6	MAPK8	PIK3R1	RELN	VEGFD
CCND2	DIAPH1	ITGA7	MAPK9	PIK3R2	RHOA	VTN
CCND3	DOCK1	ITGA8	MET	PIK3R3	ROCK1	VWF
CDC42	EGF	ITGA9	MYL10	PIK3R5	ROCK2	XIAP
CHAD	EGFR	ITGAV	MYL12A	PPP1CA	SHC1	ZYX
COL11A1	ELK1	ITGB1	MYL12B	PPP1CB	SHC2	
COL11A2	ERBB2	ITGB3	MYL2	PPP1CC	SHC3	
COL1A1	FLNA	ITGB4	MYL5	PPP1R12A	SHC4	
COL1A2	FLNB	ITGB5	MYL7	PPP1R12B	SOS1	

ECM-receptor interaction (from DAVID - https://david.ncifcrf.gov/kegg.jsp?path=hsa04512$ECM-receptor%20interaction&termId=550028745&source=kegg):

AGRN	COL2A1	COL6A1	GP6	ITGA4	ITGB6	LAMB4
CD36	COL3A1	COL6A2	GP9	ITGA5	ITGB7	LAMC1
CD44	COL4A1	COL6A3	HMMR	ITGA6	ITGB8	LAMC2
CD47	COL4A2	COL6A5	HSPG2	ITGA7	LAMA1	LAMC3
CHAD	COL4A3	COL6A6	IBSP	ITGA8	LAMA2	RELN
COL11A1	COL4A4	COMP	ITGA1	ITGA9	LAMA3	SDC1
COL11A2	COL4A5	DAG1	ITGA10	ITGAV	LAMA4	SDC4
COL1A1	COL4A6	FN1	ITGA11	ITGB1	LAMA5	SPP1
COL1A2	COL5A1	GP1BA	ITGA2	ITGB3	LAMB1	SV2A
COL24A1	COL5A2	GP1BB	ITGA2B	ITGB4	LAMB2	SV2B
COL27A1	COL5A3	GP5	ITGA3	ITGB5	LAMB3	SV2C

THBS1	THBS3	TNC	TNR	VTN
THBS2	THBS4	TNN	TNXB	VWF

Amoebiasis (from DAVID - https://david.ncifcrf.gov/kegg.jsp?path=hsa05146$Amoebiasis&termId=550028844&source=kegg):

ACTN1	COL1A2	CXCL8	IL6	NFKB1	PRKACG	SERPINB4
ACTN2	COL24A1	FN1	ITGAM	NOS2	PRKCA	SERPINB6
ACTN3	COL27A1	GNA11	ITGB2	PIK3CA	PRKCB	SERPINB9
ACTN4	COL2A1	GNA14	LAMA1	PIK3CB	PRKCG	TGFB1
ADCY1	COL3A1	GNA15	LAMA2	PIK3CD	PTK2	TGFB2
ARG1	COL4A1	GNAL	LAMA3	PIK3CG	RAB5A	TGFB3
ARG2	COL4A2	GNAQ	LAMA4	PIK3R1	RAB5B	TLR2
C8A	COL4A3	GNAS	LAMA5	PIK3R2	RAB5C	TLR4
C8B	COL4A4	HSPB1	LAMB1	PIK3R3	RAB7A	TNF
C8G	COL4A5	IFNG	LAMB2	PIK3R5	RAB7B	VCL
C9	COL4A6	IL10	LAMB3	PLCB1	RELA	
CASP3	COL5A1	IL12A	LAMB4	PLCB2	SERPINB1	
CD14	COL5A2	IL12B	LAMC1	PLCB3	SERPINB10	
COL11A1	COL5A3	IL1B	LAMC2	PLCB4	SERPINB13	
COL11A2	CSF2	IL1R1	LAMC3	PRKACA	SERPINB2	
COL1A1	CTSG	IL1R2	MUC2	PRKACB	SERPINB3	

Protein digestion and absorption (from DAVID - https://david.ncifcrf.gov/kegg.jsp?path=hsa04974$Protein%20digestion%20and%20absorption&termId=550028816&source=kegg):

ACE2	COL10A1	COL24A1	COL6A1	CPB2	PGA3	SLC3A2
ATP1A1	COL11A1	COL27A1	COL6A2	CTRL	PGA4	SLC6A19
ATP1A2	COL11A2	COL2A1	COL6A3	DPP4	PGA5	SLC7A7
ATP1A3	COL12A1	COL3A1	COL6A5	ELN	PRCP	SLC7A8
ATP1A4	COL13A1	COL4A1	COL6A6	FXYD2	PRSS1	SLC7A9
ATP1B1	COL14A1	COL4A2	COL7A1	KCNE3	PRSS2	SLC8A1
ATP1B2	COL15A1	COL4A3	COL9A1	KCNJ13	PRSS3	SLC8A2
ATP1B3	COL17A1	COL4A4	COL9A2	KCNK5	SLC15A1	SLC8A3
ATP1B4	COL18A1	COL4A5	COL9A3	KCNN4	SLC16A10	SLC9A3
CELA2A	COL1A1	COL4A6	CPA1	KCNQ1	SLC1A1	XPNPEP2
CELA2B	COL1A2	COL5A1	CPA2	MEP1A	SLC1A5	
CELA3A	COL21A1	COL5A2	CPA3	MEP1B	SLC38A2	
CELA3B	COL22A1	COL5A3	CPB1	MME	SLC3A1	

Regulation of actin cytoskeleton (from DAVID - https://david.ncifcrf.gov/kegg.jsp?path=hsa04810$Regulation%20of%20actin%20cytoskeleton&termId=550028785&source=kegg):

ABI2	ACTN4	ARHGEF12	ARPC2	BCAR1	CFL1	CHRM5
ACTB	APC	ARHGEF4	ARPC3	BDKRB1	CFL2	CRK
ACTG1	APC2	ARHGEF6	ARPC4	BDKRB2	CHRM1	CRKL
ACTN1	ARAF	ARHGEF7	ARPC5	BRAF	CHRM2	CYFIP1
ACTN2	ARHGAP35	ARPC1A	ARPC5L	BRK1	CHRM3	CYFIP2
ACTN3	ARHGEF1	ARPC1B	BAIAP2	CDC42	CHRM4	DIAPH1

DIAPH2	FGF22	IQGAP3	ITGB6	MYLPF	PIK3R1	RHOA
DIAPH3	FGF23	ITGA1	ITGB7	NCKAP1	PIK3R2	ROCK1
DOCK1	FGF3	ITGA10	ITGB8	NCKAP1L	PIK3R3	ROCK2
EGF	FGF4	ITGA11	KRAS	NRAS	PIK3R5	RRAS
EGFR	FGF5	ITGA2	LIMK1	PAK1	PIKFYVE	RRAS2
ENAH	FGF6	ITGA2B	LIMK2	PAK2	PIP4K2A	SCIN
EZR	FGF7	ITGA3	MAP2K1	PAK3	PIP4K2B	SLC9A1
F2	FGF8	ITGA4	MAP2K2	PAK4	PIP4K2C	SOS1
F2R	FGF9	ITGA5	MAPK1	PAK5	PIP5K1A	SOS2
FGD1	FGFR1	ITGA6	MAPK3	PAK6	PIP5K1B	SRC
FGD3	FGFR2	ITGA7	MOS	PDGFA	PIP5K1C	SSH1
FGF1	FGFR3	ITGA8	MRAS	PDGFB	PPP1CA	SSH2
FGF10	FGFR4	ITGA9	MSN	PDGFC	PPP1CB	SSH3
FGF11	FN1	ITGAD	MYL10	PDGFD	PPP1CC	TIAM1
FGF12	GIT1	ITGAE	MYL12A	PDGFRA	PPP1R12A	TMSB4X
FGF13	GNA12	ITGAL	MYL12B	PDGFRB	PPP1R12B	TMSB4Y
FGF14	GNA13	ITGAM	MYL2	PFN1	PPP1R12C	VAV1
FGF16	GNG12	ITGAV	MYL5	PFN2	PTK2	VAV2
FGF17	GSN	ITGAX	MYL7	PFN3	PXN	VAV3
FGF18	HRAS	ITGB1	MYL9	PFN4	RAC1	VCL
FGF19	INS	ITGB2	MYLK	PIK3CA	RAC2	WAS
FGF2	INSRR	ITGB3	MYLK2	PIK3CB	RAC3	WASF1
FGF20	IQGAP1	ITGB4	MYLK3	PIK3CD	RAF1	WASF2
FGF21	IQGAP2	ITGB5	MYLK4	PIK3CG	RDX	WASL

Rap1 signaling pathway (from DAVID - https://david.ncifcrf.gov/kegg.jsp?path=hsa04015$Rap1%20signaling%20pathway&termId=550028704&source=kegg):

ACTB	BCAR1	EFNA4	FGF23	GRIN1	LCP2	NGFR
ACTG1	BRAF	EFNA5	FGF3	GRIN2A	LPAR1	NRAS
ADCY1	CALM1	EGF	FGF4	GRIN2B	LPAR2	P2RY1
ADCY2	CALM2	EGFR	FGF5	HGF	LPAR3	PARD3
ADCY3	CALM3	EPHA2	FGF6	HRAS	LPAR4	PARD6A
ADCY4	CALML3	F2R	FGF7	ID1	LPAR5	PARD6B
ADCY5	CALML5	F2RL3	FGF8	IGF1	MAGI1	PARD6G
ADCY6	CALML6	FARP2	FGF9	IGF1R	MAGI2	PDGFA
ADCY7	CDC42	FGF1	FGFR1	INS	MAGI3	PDGFB
ADCY8	CDH1	FGF10	FGFR2	INSR	MAP2K1	PDGFC
ADCY9	CNR1	FGF11	FGFR3	ITGA2B	MAP2K2	PDGFD
ADORA2A	CRK	FGF12	FGFR4	ITGAL	MAP2K3	PDGFRA
ADORA2B	CRKL	FGF13	FLT1	ITGAM	MAP2K6	PDGFRB
AFDN	CSF1	FGF14	FLT4	ITGB1	MAPK1	PFN1
AKT1	CSF1R	FGF16	FPR1	ITGB2	MAPK11	PFN2
AKT2	CTNNB1	FGF17	FYB	ITGB3	MAPK12	PFN3
AKT3	CTNND1	FGF18	GNAI1	KDR	MAPK13	PFN4
ANGPT1	DOCK4	FGF19	GNAI2	KIT	MAPK14	PGF
ANGPT2	DRD2	FGF2	GNAI3	KITLG	MAPK3	PIK3CA
ANGPT4	EFNA1	FGF20	GNAO1	KRAS	MET	PIK3CB
APBB1IP	EFNA2	FGF21	GNAQ	KRIT1	MRAS	PIK3CD
ARAP3	EFNA3	FGF22	GNAS	LAT	NGF	PIK3CG

PIK3R1	PLCE1	PRKD2	RALGDS	RAPGEF5	SIPA1	TIAM1
PIK3R2	PLCG1	PRKD3	RAP1A	RAPGEF6	SIPA1L1	TLN1
PIK3R3	PRKCA	RAC1	RAP1B	RASGRP2	SIPA1L2	TLN2
PIK3R5	PRKCB	RAC2	RAP1GAP	RASGRP3	SIPA1L3	VASP
PLCB1	PRKCG	RAC3	RAPGEF1	RASSF5	SKAP1	VEGFA
PLCB2	PRKCI	RAF1	RAPGEF2	RGS14	SRC	VEGFB
PLCB3	PRKCZ	RALA	RAPGEF3	RHOA	TEK	VEGFC
PLCB4	PRKD1	RALB	RAPGEF4	RRAS	THBS1	VEGFD

PI3K-Akt signaling pathway (from DAVID - https://david.ncifcrf.gov/kegg.jsp?path=hsa04151$PI3K-Akt%20signaling%20pathway&termId=550028729&source=kegg):

AKT1	COL27A1	EGF	FLT1	IFNA10	ITGA2B	LPAR4
AKT2	COL2A1	EGFR	FLT4	IFNA13	ITGA3	LPAR5
AKT3	COL3A1	EIF4B	FN1	IFNA14	ITGA4	LPAR6
ANGPT1	COL4A1	EIF4E	FOXO3	IFNA16	ITGA5	MAP2K1
ANGPT2	COL4A2	EIF4E1B	G6PC	IFNA17	ITGA6	MAP2K2
ANGPT4	COL4A3	EIF4E2	G6PC2	IFNA2	ITGA7	MAPK1
ATF2	COL4A4	EIF4EBP1	G6PC3	IFNA21	ITGA8	MAPK3
ATF4	COL4A5	EPHA2	GH1	IFNA4	ITGA9	MCL1
ATF6B	COL4A6	EPO	GH2	IFNA5	ITGAV	MDM2
BAD	COL5A1	EPOR	GHR	IFNA6	ITGB1	MET
BCL2	COL5A2	F2R	GNB1	IFNA7	ITGB3	MLST8
BCL2L1	COL5A3	FASLG	GNB2	IFNA8	ITGB4	MTCP1
BCL2L11	COL6A1	FGF1	GNB3	IFNAR1	ITGB5	MTOR
BRCA1	COL6A2	FGF10	GNB4	IFNAR2	ITGB6	MYB
C8orf44-	COL6A3	FGF11	GNB5	IFNB1	ITGB7	MYC
SGK3	COL6A5	FGF12	GNG10	IGF1	ITGB8	NFKB1
CASP9	COL6A6	FGF13	GNG11	IGF1R	JAK1	NGF
CCND1	COMP	FGF14	GNG12	IKBKB	JAK2	NGFR
CCND2	CREB1	FGF16	GNG13	IKBKG	JAK3	NOS3
CCND3	CREB3	FGF17	GNG2	IL2	KDR	NR4A1
CCNE1	CREB3L1	FGF18	GNG3	IL2RA	KIT	NRAS
CCNE2	CREB3L2	FGF19	GNG4	IL2RB	KITLG	OSM
CD19	CREB3L3	FGF2	GNG5	IL2RG	KRAS	OSMR
CDC37	CREB3L4	FGF20	GNG7	IL3	LAMA1	PCK1
CDK2	CREB5	FGF21	GNG8	IL3RA	LAMA2	PCK2
CDK4	CRTC2	FGF22	GNGT1	IL4	LAMA3	PDGFA
CDK6	CSF1	FGF23	GNGT2	IL4R	LAMA4	PDGFB
CDKN1A	CSF1R	FGF3	GRB2	IL6	LAMA5	PDGFC
CDKN1B	CSF3	FGF4	GSK3B	IL6R	LAMB1	PDGFD
CHAD	CSF3R	FGF5	GYS1	IL7	LAMB2	PDGFRA
CHRM1	CSH1	FGF6	GYS2	IL7R	LAMB3	PDGFRB
CHRM2	CSH2	FGF7	HGF	INS	LAMB4	PDPK1
CHUK	DDIT4	FGF8	HRAS	INSR	LAMC1	PGF
COL11A1	EFNA1	FGF9	HSP90AA1	IRS1	LAMC2	PHLPP1
COL11A2	EFNA2	FGFR1	HSP90AB1	ITGA1	LAMC3	PHLPP2
COL1A1	EFNA3	FGFR2	HSP90B1	ITGA10	LPAR1	PIK3AP1
COL1A2	EFNA4	FGFR3	IBSP	ITGA11	LPAR2	PIK3CA
COL24A1	EFNA5	FGFR4	IFNA1	ITGA2	LPAR3	PIK3CB

PIK3CD	PPP2R1B	PPP2R5E	RELN	SPP1	TLR4	VTN
PIK3CG	PPP2R2A	PRKAA1	RHEB	STK11	TNC	VWF
PIK3R1	PPP2R2B	PRKAA2	RPS6	SYK	TNN	YWHAB
PIK3R2	PPP2R2C	PRKCA	RPS6KB1	TCL1A	TNR	YWHAE
PIK3R3	PPP2R2D	PRL	RPS6KB2	TCL1B	TNXB	YWHAG
PIK3R5	PPP2R3A	PRLR	RPTOR	TEK	TP53	YWHAH
PKN1	PPP2R3B	PTEN	RXRA	THBS1	TSC1	YWHAQ
PKN2	PPP2R3C	PTK2	SGK1	THBS2	TSC2	YWHAZ
PKN3	PPP2R5A	RAC1	SGK2	THBS3	VEGFA	
PPP2CA	PPP2R5B	RAF1	SGK3	THBS4	VEGFB	
PPP2CB	PPP2R5C	RBL2	SOS1	THEM4	VEGFC	
PPP2R1A	PPP2R5D	RELA	SOS2	TLR2	VEGFD	

Inositol phosphate metabolism (from DAVID - https://david.ncifcrf.gov/kegg.jsp?path=hsa00562$Inositol%20phosphate%20metabolism&termId=550028641&source=kegg):

ALDH6A1	INPP5E	MINPP1	OCRL	PIK3CD	PLCB2	PLCZ1
CDIPT	INPP5J	MIOX	PI4K2A	PIK3CG	PLCB3	PTEN
IMPA1	INPP5K	MTM1	PI4K2B	PIKFYVE	PLCB4	SYNJ1
IMPA2	INPPL1	MTMR1	PI4KA	PIP4K2A	PLCD1	SYNJ2
IMPAD1	IPMK	MTMR14	PI4KB	PIP4K2B	PLCD3	TPI1
INPP1	IPPK	MTMR2	PIK3C2A	PIP4K2C	PLCD4	
INPP4A	ISYNA1	MTMR3	PIK3C2B	PIP5K1A	PLCE1	
INPP4B	ITPK1	MTMR4	PIK3C2G	PIP5K1B	PLCG1	
INPP5A	ITPKA	MTMR6	PIK3C3	PIP5K1C	PLCG2	
INPP5B	ITPKB	MTMR7	PIK3CA	PIP5KL1	PLCH1	
INPP5D	ITPKC	MTMR8	PIK3CB	PLCB1	PLCH2	

Melanoma (from DAVID - https://david.ncifcrf.gov/kegg.jsp?path=hsa05218$Melanoma&termId=550028868&source=kegg):

AKT1	CDKN2A	FGF14	FGF4	KRAS	PDGFC	PIK3R5
AKT2	E2F1	FGF16	FGF5	MAP2K1	PDGFD	PTEN
AKT3	E2F2	FGF17	FGF6	MAP2K2	PDGFRA	RAF1
ARAF	E2F3	FGF18	FGF7	MAPK1	PDGFRB	RB1
BAD	EGF	FGF19	FGF8	MAPK3	PIK3CA	TP53
BRAF	EGFR	FGF2	FGF9	MDM2	PIK3CB	
CCND1	FGF1	FGF20	FGFR1	MET	PIK3CD	
CDH1	FGF10	FGF21	HGF	MITF	PIK3CG	
CDK4	FGF11	FGF22	HRAS	NRAS	PIK3R1	
CDK6	FGF12	FGF23	IGF1	PDGFA	PIK3R2	
CDKN1A	FGF13	FGF3	IGF1R	PDGFB	PIK3R3	

B.2.3 Individual patient MEVs

Master PatientID (BM)	Focal Adhesion genes with MEVs	ECM-receptor interaction genes with MEVs	Amoebiasis genes with MEVs	Protein digestion and absorption	Regulation of actin cytoskeleton	Rap1 signaling pathway	PI3K-Akt signaling pathway	Inositol phosphate metabolism	Melanoma		Pathways with mutations	MEV genes in pathways
19	CCND1 COL11A1 COL2A1 COL4A3 COL5A1 COL6A5 CRK FLNC ITGB5 ITGB8 LAMA2 LAMA5 PDGFD PGF THBS2 TLN2 TNC VAV3	COL11A1 COL2A1 COL4A3 COL5A1 COL6A5 GP1BA ITGB5 ITGB8 LAMA2 LAMA5 THBS2 TNC	C8A COL11A1 COL2A1 COL4A3 COL5A1 GNA11 ITGAM ITGB2 LAMA2 LAMA5 MUC2 PLCB3	COL11A1 COL15A1 COL2A1 COL4A3 COL5A1 COL6A5 COL9A1 KCNK5 SLC6A19 SLC8A2 SLC9A3	CRK FGF12 FGF13 ITGAD ITGAM ITGB2 ITGB5 ITGB8 PDGFD PIP4K2A TIAM1 VAV3	APBB1IP CRK FGF12 FGF13 GRIN2B ITGAM ITGB2 PDGFD PGF PLCB3 RAPGEF5 RASGRP2 TIAM1 TLN2	CCND1 COL11A1 COL2A1 COL4A3 COL5A1 COL6A5 FGF12 FGF13 ITGB5 ITGB8 LAMA2 LAMA5 PDGFD PGF PPP2R2B THBS2 TNC TP53	INPP4B ISYNA1 PIP4K2A PLCB3	CCND1 CDKN2A FGF12 FGF13 PDGFD TP53		Focal Adhesion ECM-receptor interaction Amoebiasis Protein digestion and absorption Regulation of actin cytoskeleton Rap1 signaling pathway PI3K-Akt signaling pathway Inositol phosphate metabolism Melanoma 9/9	45/760 (5.92%)
4	ACTN1 TNXB	TNXB	ACTN1		ACTN1		TNXB				Focal Adhesion ECM-receptor interaction Amoebiasis Regulation of actin cytoskeleton PI3K-Akt signaling pathway 5/9	2/31 (6.45%)
20	ARHGAP35 COL4A1 ITGA4 LAMA2 PIK3CD RELN	COL4A1 ITGA4 LAMA2 RELN	COL4A1 LAMA2 NFKB1 PIK3CD	COL4A1 COL9A2	ARHGAP35 CHRM2 FGF9 ITGA4 PAK2 PIK3CD	FGF9 GNAO1 PARD3 PIK3CD	CHRM2 COL4A1 FGF9 ITGA4 LAMA2 NFKB1 PIK3CD PPP2R3A RELN	PIK3CD	FGF9 PIK3CD		Focal Adhesion ECM-receptor interaction Amoebiasis Protein digestion and absorption Regulation of actin cytoskeleton Rap1 signaling pathway PI3K-Akt signaling pathway Inositol phosphate metabolism Melanoma 9/9	14/238 (5.88%)
5	COL5A2 COL6A5 IGF1R LAMA4 MYLK3	COL5A2 COL6A5 LAMA4	COL5A2 LAMA4 NFKB1 PLCB4	COL15A1 COL22A1 COL5A2 COL6A5 COL9A1	FGF22 MYLK3 NCKAP1	FGF22 GRIN2A ID1 IGF1R PLCB4 SIPA1L1	CHRM2 COL4A1 FGF9 ITGA4 LAMA2 NFKB1 PIK3CD PPP2R3A RELN	PIK3C2G PIK3C3 PLCB4	CDKN2A FGF22 IGF1R TP53		Focal Adhesion ECM-receptor interaction Amoebiasis Protein digestion and absorption Regulation of actin cytoskeleton Rap1 signaling pathway PI3K-Akt signaling pathway Inositol phosphate metabolism Melanoma 9/9	27/355 (7.61%)
21	LAMA1 RELN	LAMA1 RELN	LAMA1		ITGAX		LAMA1 RELN STK11				Focal Adhesion ECM-receptor interaction Amoebiasis Regulation of actin cytoskeleton PI3K-Akt signaling pathway 5/9	4/109 (3.67%)
34											0/9	0/15 (0%)

Master PatientID (BM)	Focal Adhesion genes with MEVs	ECM-receptor interaction genes with MEVs	Amoebiasis genes with MEVs	Protein digestion and absorption	Regulation of actin cytoskeleton	Rap1 signaling pathway	PI3K-Akt signaling pathway	Inositol phosphate metabolism	Melanoma		Pathways with mutations	MEV genes in pathways
30	COL4A5	COL4A5	COL4A5 SERPINB10	COL4A5		GRIN2A	COL4A5 PPP2R5E				Focal Adhesion ECM-receptor interaction Amoebiasis Protein digestion and absorption Rap1 signaling pathway PI3K-Akt signaling pathway 6/9	4/64 (6.25%)
7	VAV2		PLCB2	COL21A1	CYFIP1 DIAPH2 FGF1 FGF13 FGF20 GIT1 VAV2	FGF1 FGF13 FGF20 PLCB2	FGF1 FGF13 FGF20	INPP5J OCRL PLCB2	FGF1 FGF13 FGF20		Focal Adhesion Amoebiasis Protein digestion and absorption Regulation of actin cytoskeleton Rap1 signaling pathway PI3K-Akt signaling pathway Inositol phosphate metabolism Melanoma 8/9	11/200 (5.5%)
8						PLCG1	STK11	PLCG1	CDKN2A		Rap1 signaling pathway PI3K-Akt signaling pathway Inositol phosphate metabolism Melanoma 4/9	3/31 (9.68%)
31	COL6A2	COL6A2		COL6A2 KCNN4 SLC3A2			COL6A2 CREB3L2 TP53		TP53		Focal Adhesion ECM-receptor interaction Protein digestion and absorption PI3K-Akt signaling pathway Melanoma 5/9	5/29 (17.24%)
45											0/9	0/4 (0%)
47	ELK1										Focal Adhesion 1/9	1/8 (12.5%)
50				COL9A3			CREB5	MTMR1			Protein digestion and absorption PI3K-Akt signaling pathway Inositol phosphate metabolism 3/9	3/60 (5%)
42	RELN	RELN					RELN				Focal Adhesion ECM-receptor interaction PI3K-Akt signaling pathway 3/9	1/24 (4.17%)
43	COL1A2 HGF IGF1R VAV3	COL1A2	COL1A2	COL1A2	VAV3	CSF1R FYB HGF IGF1R	COL1A2 CREB3L4 CSF1R HGF IGF1R		HGF IGF1R		Focal Adhesion ECM-receptor interaction Amoebiasis Protein digestion and absorption Regulation of actin cytoskeleton Rap1 signaling pathway PI3K-Akt signaling pathway Melanoma 8/9	7/55 (12.73%)

Master PatientID (BM)	Focal Adhesion genes with MEVs	ECM-receptor interaction genes with MEVs	Amoebiasis genes with MEVs	Protein digestion and absorption	Regulation of actin cytoskeleton	Rap1 signaling pathway	PI3K-Akt signaling pathway	Inositol phosphate metabolism	Melanoma		Pathways with mutations	MEV genes in pathways
44	LAMA4	LAMA4	LAMA4				LAMA4				Focal Adhesion ECM-receptor interaction Amoebiasis PI3K-Akt signaling pathway 4/9	1/27 (3.70%)
33											0/9	0/23 (0%)
9											0/9	0/6 (0%)
10	LAMA1 PDGFC	LAMA1	LAMA1		PDGFC	PDGFC	LAMA1 PDGFC PPP2R2B		PDGFC		Focal Adhesion ECM-receptor interaction Amoebiasis Regulation of actin cytoskeleton Rap1 signaling pathway PI3K-Akt signaling pathway Melanoma 7/9	3/38 (7.89%)
32			PLCB4			PLCB4	TCL1A	PIK3C2G PLCB4			Amoebiasis Rap1 signaling pathway PI3K-Akt signaling pathway Inositol phosphate metabolism 4/9	3/18 (16.67%)
11		SV2C									ECM-receptor interaction 1/9	1/20 (5%)
16	HGF			SLC6A19	FGD3	HGF	HGF		HGF		Focal Adhesion Protein digestion and absorption Regulation of actin cytoskeleton Rap1 signaling pathway PI3K-Akt signaling pathway Melanoma 6/9	3/54 (5.56%)
17	COL1A1 IBSP PIK3CA PIK3CG SOS2	COL1A1 DAG1 IBSP	COL1A1 IL6 PIK3CA PIK3CG	COL17A1 COL1A1 SLC8A2	PIK3CA PIK3CG PIKFYVE SOS2	ADCY3 ADCY9 PIK3CA PIK3CG	COL1A1 IBSP IL6 PCK2 PIK3CA PIK3CG SOS2 TP53	PIK3C2A PIK3CA PIK3CG PIKFYVE	PIK3CA PIK3CG RB1 TP53		Focal Adhesion ECM-receptor interaction Amoebiasis Protein digestion and absorption Regulation of actin cytoskeleton Rap1 signaling pathway PI3K-Akt signaling pathway Inositol phosphate metabolism Melanoma 9/9	16/373 (4.29%)
18				COL22A1	BAIAP2						Protein digestion and absorption Regulation of actin cytoskeleton 2/9	2/18 (11.11%)

Master PatientID (BM)	Focal Adhesion genes with MEVs	ECM-receptor interaction genes with MEVs	Amoebiasis genes with MEVs	Protein digestion and absorption	Regulation of actin cytoskeleton	Rap1 signaling pathway	PI3K-Akt signaling pathway	Inositol phosphate metabolism	Melanoma		Pathways with mutations	MEV genes in pathways
2	COL5A1 COL6A5 FLNC	COL5A1 COL6A5	COL5A1	COL5A1 COL6A5	FGD1 IQGAP1		COL5A1 COL6A5 JAK2				Focal Adhesion ECM-receptor interaction Amoebiasis Protein digestion and absorption Regulation of actin cytoskeleton PI3K-Akt signaling pathway 6/9	6/133 (4.51%)
3	ARHGAP35 BCAR1 BRAF COL24A1 COL4A6 FLT1 HGF	COL24A1 COL4A6	COL24A1 COL4A6 MUC2 RAB5A	COL24A1 COL4A6 COL9A1	ARHGAP35 BCAR1 BRAF CHRM3 FGFR1 IQGAP2 MAP2K2 MRAS	BCAR1 BRAF FGFR1 FLT1 HGF MAP2K2 MRAS PLCE1	CHUK COL24A1 COL4A6 FGFR1 FLT1 HGF MAP2K2 RXRA	PLCE1	BRAF FGFR1 HGF MAP2K2		Focal Adhesion ECM-receptor interaction Amoebiasis Protein digestion and absorption Regulation of actin cytoskeleton Rap1 signaling pathway PI3K-Akt signaling pathway Inositol phosphate metabolism Melanoma 9/9	18/302 (5.96%)
22											0/9	0/4 (0%)
23	COL11A1 COL2A1	COL11A1 COL2A1	COL11A1 COL2A1	COL11A1 COL2A1 KCNQ1 SLC8A3			COL11A1 COL2A1 GYS2	INPP4B MTMR2	RB1		Focal Adhesion ECM-receptor interaction Amoebiasis Protein digestion and absorption PI3K-Akt signaling pathway Inositol phosphate metabolism Melanoma 7/9	8/124 (6.45%)
6	ITGA6 ITGAV LAMA3 MYL2	ITGA6 ITGAV LAMA3	LAMA3		ITGA6 ITGAV MYL2	ADCY8	ITGA6 ITGAV LAMA3	MTMR7			Focal Adhesion ECM-receptor interaction Amoebiasis Regulation of actin cytoskeleton Rap1 signaling pathway PI3K-Akt signaling pathway Inositol phosphate metabolism 7/9	6/46 (13.04%)
24	DOCK1 ITGA6 JUN MYLK	ITGA6		COL22A1	DOCK1 ITGA6 MYLK SSH1	EPHA2 GNAI3	EPHA2 ITGA6 TP53	MTMR1	TP53		Focal Adhesion ECM-receptor interaction Protein digestion and absorption Regulation of actin cytoskeleton Rap1 signaling pathway PI3K-Akt signaling pathway Inositol phosphate metabolism Melanoma 8/9	10/125 (8%)
TOTAL NUMBER OF SAMPLES WITH AT LEAST 1 MEV IN PATHWAY:	20/30	18/30	16/30	15/30	15/30	14/30	22/30	12/30	13/30			
TOTAL NUMBER OF GENES IN PATHWAY:	206	87	106	88	210	210	345	71	71			

B.3 Differentially expressed genes

B.3.1 Genes with a significant gain of expression in the metastasis compared to the primary

Genes are listed in order from greatest change in expression to least.

GFAP	CADM2	HPCAL4	MAST1	TMEM63C	LOC101927905
PLP1	CNTN2	KIF5C	GRIN1	KCTD21	DCST1-AS1
MAG	ZDHHC22	PRR15L	OLIG1	FUT2	CNIH2
ATP1B2	TMEM151B	ADGRB2	NDRG4	LDLRAD3	TACC2
NCAN	SLC17A7	GRM8	GRM1	KIRREL3	ENPEP
AMER2	TIMP4	KCNJ3	PACSIN1	RAB40B	GPR75
SLC1A2	SCN3B	SALRNA2	CADM3	NGEF	PCLO
GAP43	LINC00599	PHYHIPL	MLXIPL	TMEM139	BRSK2
PMP2	ST8SIA5	PSCA	PLA2G16	PON1	CIT
ZIC1	GRIN2B	SYNGR1	GPR156	LOC105447645	MCF2L2
SLC12A5	SYN2	GNAO1	JPH4	ZIC2	NUDT17
MOBP	C1QL1	PTPRZ1	PTGES	LEAP2	LOC105369635
ATP1A2	XKRX	KCNT1	FAM71F2	NUDT16P1	DOC2A
SH3TC2	CACNG7	SYBU	GREB1	CDK5R1	LINC01806
HAPLN2	KRT81	ATP8A2	APOD	BEGAIN	DCST2
CNTNAP4	ITGA10	CPA4	KCNH3	TUBB1	ZNF185
CNDP1	MSX1	EMX2OS	LOC221946	LYNX1	PCAT6
AK5	ARPP21	NTRK3	AMPH	MEG3	TTYH1
KIF5A	LINGO1	TAGLN3	RNVU1-14	NKPD1	SLC10A1
PDZD4	FAM57B	PLEKHG6	NAT8L	SH3D21	DPH3P1
STMN4	FRRS1L	ADAM22	CPE	RUNDC3B	KIAA1161
LINC00461	OPCML	PSD2	LINC01426	RNF139-AS1	PPL
RBFOX1	SLC30A10	HOTS	STRA6	RAB26	ARHGAP32
GRIA2	SLC7A14	MIR200B	KIAA0319	CPD	DCST1
TMEM179	TMEM151A	RBFOX3	ADAM11	TUBA8	ZNF337-AS1
MBP	CELF4	CLVS2	NPTXR	EPS8	SLC45A4
KCNB1	CERS1	GDF1	TMIGD3	ARNT2	CA14
SLC24A2	PIANP	LOC100132813	LOC100996419	CD163	SCNN1A
GPR37L1	TBR1	BSN	DOCK5	CES4A	C17orf53
RPH3A	MUC3A	SAA1	SNORA12	ARHGEF33	ERBB3
STMN2	PRRT1	PCDH9	RAD9B	LOC100506281	C2orf82
LCN2	MAPT	ARHGAP40	KIAA1755	FAM134B	PLXDC1
SOX2-OT	L1CAM	GRM3	C1QL3	ANKS1B	DAGLA
IL20RB	FAM131B	SULT4A1	FKBP5	ACPT	ZNF131
HCN2	HEPACAM	LINC01088	PDZD3	RNF157	CDH1
ELAVL3	3-Sep	S100B	TFCP2L1	PLLP	WBSCR27
TUBB4A	PHACTR3	SERPINA5	CYP2W1	CEND1	PCDH1
CHD5	APLP1	UBL4B	GABRD	PYGB	ABALON
PNMA5	ZCCHC12	SPTSSA	KRT19	FAM86JP	BDH1
MOG	GP2	MAP1A	GPR179	WNT4	SAYSD1
SYN1	TRIM9	FAIM2	REEP2	LINC00672	NEDD4L
ATP6V1G2	CRYAB	IQSEC3	SPDYA	SLC25A18	MORC2-AS1
KCNJ9	TMEM132B	PDIA2	ELOVL2	FAM114A1	ADAM1A
ATCAY	HSD11B2	SCD5	VSIG4	CORO2B	SPICE1
SLC6A1	EFHD1	RIMS4	DHRS3	TTC6	TMEM144

SLC25A4	LPP-AS2	MYO5B	DAGLB	KCTD2	ARMC7
WNT3	NOS1AP	PABPC1L	TMEM265	UBFD1	CSNK1D
TANC2	CAPN2	ELP5	ZSCAN20	DCAF8	ZNF740
PA2G4P4	LOC100287497	POP1	MRGBP	DNAJC5	FAM127B
FBXL19-AS1	CMTM4	TACO1	SCFD2	BLCAP	TM9SF4
EPB41L1	USP31	NFS1	PNPO	COG1	EMC1
TMEM186	UMPS	MSH5	SS18L1	ATP6V1G2-	GYS1
STK39	PLEKHA7	SEC14L2	KIAA0556	DDX39B	RPTOR
ITFG1	METTL23	ZNF8-ERVK3-	TRPV1	ZNF512B	CTNS
TBC1D8	TOP3B	1	ERAL1	ANKZF1	GTF3C2
GGNBP2	EPHA1	OSBPL2	PDPR	HEXIM1	PIP4K2B
C1orf210	TMC4	ZNF26	CLDN7	CAMK2G	COASY
ATP9A	ERLIN2	RAD52	ZHX3	ARHGAP35	JMJD6
LOC153684	CHKA	UCKL1	HDHD3	RPRD1B	AGBL5
ELMO3	ZNF8	TBC1D16	TTPAL	GID8	MAPK3
GPAT4	PICK1	PDXDC2P	CABLES2	LOC646471	FOXK2
BTBD18	COG7	PCYT1A	IQCE	RCOR3	CWC25
MTA3	TRAF4	BCL2L1	RECQL5	KIF1B	PINK1-AS
THRA	STK36	TRIO	ARRDC1-AS1	TPD52L2	ATXN7L3
NCDN	MSH5-	AKAP1	YTHDF1	LRTOMT	
KIF9	SAPCD1	MMACHC	ARL16	CANT1	
SPTBN2	GTF2IRD1	PCYT2	MORC2	NOP2	
C9orf131	CLCN2	RBM4	DDX39B	SRSF2	

B.3.2 Genes with a significant loss of expression in the metastasis compared to the primary

Genes are listed in order from greatest change in expression to least.

TCF21	MFAP4	MS4A2	COX4I2	LOC100507616	IGF1
TBX5	FMO2	FGF7	CSF3	P2RX1	MYOCD
CCL21	BHLHE22	MYH11	MOXD1	ADAMTS1	NR5A1
ADH1B	COL6A6	SPIB	DNAH9	FCRLA	MEOX2
CILP	PLA2G2A	PRB1	DCC	COL14A1	LMCD1
TBX5-AS1	CCL19	LOC100507175	SFTPC	TLR10	LINC00968
LOC100131315	EMILIN1	PODN	CHRDL2	GAS1	SSC5D
DPT	INHBA-AS1	LOC285043	NPR1	SPON1	ROBO2
AGBL1	FCRL1	FNDC1	CNN1	SELE	PLA2G2D
TMEM119	LTF	MMP11	CCDC141	INHBA	PDPN
TBX4	TPSAB1	BLK	HDC	MYH8	SLIT2
DES	SVEP1	GPR25	MIR5685	ANKRD36BP2	ADAMTS12
WT1	CHIT1	SCGB1A1	CD300E	MYH13	PIM2
GREM1	OR7E47P	FOXF1	POU2AF1	CRISPLD2	MEOX1
TNFRSF13B	MS4A1	WISP2	PDGFRA	C11orf88	PDLIM3
TPSB2	ADAMTS8	COL6A5	GJA5	CD1E	IL1RL1
RSPO1	FENDRR	NR4A3	CCL14	KCNJ15	CD79A
CHRDL1	HTRA3	HPSE2	PLA2G5	TIGIT	MIR1281
C11orf96	RARRES2	MIR4538	EGFL6	PLA2G4E	WT1-AS
AOC3	INMT	FAM30A	HMCN1	FAM216B	IL6
LRRC15	CYGB	TNFRSF8	MMP23B	IRF4	ABI3BP
MMP2	GFRA1	TNFRSF9	ADIPOQ	MIR663A	KLHL6

CLDN18	SPN	SCO2	ISM1	IL21R-AS1	IGFBP7
SLC5A8	LHFPL3-AS2	SLAMF7	C16orf54	VKORC1L1	TIE1
SIGLEC6	MEIS1-AS2	SKOR2	SULF1	KCNA3	ANKRD44-IT1
MIR4539	HSD11B1	CDC42EP5	FOXF2	PIK3CD-AS1	LINC00612
LINC01429	FCMR	MCOLN2	SPHK1	LOC101927322	PDE3B
ROCK1P1	PAPPA2	LOC389199	CD48	IL2RB	MATK
CPA3	PTMS	CD28	CNR2	LOC101928988	PTGDR
LRRC18	WNT2	LINC01266	TBX3	ANTXR1	LINC00926
MIR4537	LINC00528	SOCS3	C10orf55	CCR5	IL12RB1
TDRD1	DNAH12	MMP19	REM1	ERG	FAM110D
FAM26F	COMP	MYLK	PLEKHO1	DERL3	ADARB1
WFDC1	RBMXL2	COL6A3	PODNL1	RAMP3	LBH
MRGPRF	RGS16	LY9	COL12A1	LYZ	WDFY4
LUM	CCDC80	TBX2	H2AFX	C7	MAP4K1
PRDM6	FAM162B	TMEM158	CD83	KLRG1	TIMP3
PTGER3	C8orf34-AS1	SH2D1A	IL16	COL6A2	PTRF
LAG3	FCRL4	LINC00861	CD19	TMC8	COL10A1
CLEC17A	LOC613266	DUSP2	HMCN2	GATA6	ARHGAP9
RASAL3	DTHD1	DNAH8	SMTNL2	LINC01108	HLX
FCRL3	SIRPG-AS1	ICOS	RUNX3	MIR3945HG	GVINP1
NOTUM	LOC101927666	CD3E	TAL1	MSC	ADAMTS4
KLK4	LAX1	ST6GALNAC5	TSPAN32	LILRB2	THEMIS2
IGLL5	SLC15A3	TNFSF11	TWIST2	CD3G	ITGAX
TNFRSF13C	CD6	SGCD	AGER	TNXB	BCL11B
LMOD1	ELN	SFRP2	GPX3	KCNJ5	PALLD
FAM129C	GRAP2	ASB2	DACT1	TNFRSF4	NCF1
MEIS1	MIR3142HG	JCHAIN	PGR	CD300LB	ZAP70
SRPX	LINC00982	ABCA6	SOCS2	HTRA4	TESPA1
ZNF80	LOC649133	CLIC5	MIR3606	MSRB3	PLXNC1
GSTM5	SOX18	LOC102723766	FCN1	FAM186A	RASIP1
CXCL14	ANGPTL2	AEBP1	XCR1	ADAM19	FGD5
LOC101928882	SNX20	FOXP3	ARHGEF15	CD274	PRCD
GABRA1	TYMP	RND1	LTB	SP140	GAS6
COL15A1	KCNJ2	C11orf21	XYLT1	CD2	IKZF1
RUVBL1-AS1	CPXM2	ZBP1	RNF19B	METRNL	COL1A2
ANTXRLP1	C15orf53	CASS4	ZC3H12D	PTPRVP	IKZF3
VGLL3	FCER2	PLA2G4E-AS1	TAGAP	WISP1	ZNRF1
XIRP1	GLI2	CD1C	TFPI2	RASSF1-AS1	FAM196B
SLIT3	FCN3	HS3ST3A1	GNLY	TMEM200A	BMP2
FAM180A	NCR3	ITGAL	COL5A1	TRG-AS1	SNORD17
MXRA5	MIR4532	SELP	TGFB1	ALPK2	FHL1
LINC01279	CTLA4	GLI1	IL24	CCR7	PIK3CG
IL21R	GPR68	LINC01934	RGCC	LILRB1	APOBR
CSF2RB	PDCD1	LHFPL3	LOC101927793	ASGR1	TCF4-AS1
A2M	NUGGC	LOC728095	TMEM220-AS1	LIMS2	ITGA11
FABP4	CYR61	PTGIR	ZNF521	POU2F2	
G0S2	PRND	GXYLT2	MEX3D	PLEK	SENCR
LRRK2	ZNF831	IGSF10	PTGIS	PYHIN1	MEI1
SFTPD	IL33	CLEC3B	CPED1	SERPING1	COL3A1
CD22	PLA2G7	DACH1	OSM	STX11	OLFML1

GRASP	KCNE1	IER5	MRC2	MAN1A1	DOK3
GNG11	PCED1B-AS1	IRF8	CLEC14A	PLEKHO2	SH3BP1
PTPN7	CEACAM21	GLIPR2	ENTPD1-AS1	JUND	VOPP1
RRN3P2	EGR3	IL4I1	ZNF331	ACP5	RIN3
FLI1	FGF2	SMAD7	ABHD17A	NLRC3	HLA-E
WIPF1	NFAM1	LILRB4	FAM78A	ARHGAP30	SAMHD1
MAGI2-AS3	ODF3B	AKNA	PTPRCAP	BCL6B	CSK
ARHGAP6	IL10RA	MMP14	NFATC2	PKIG	DENND1C
FAT4	ADAMTSL2	ARID5A	SOCS1	RELT	SPSB1
GIMAP8	PECAM1	PCDH17	JARID2-AS1	CD37	FMNL3
FMNL1	NRROS	RNF122	TRAF3IP3	RCSD1	PTMA
SOD2	S1PR2	LOC105370526	MYL9	IER2	RNF166
ARHGAP25	TNFRSF1B	JAK3	C9orf66	PSTPIP1	SH2B3
MAF	SCIMP	EGR2	EGFL7	SYNE3	
COL5A2	SEMA7A	IL15RA	LOC101927954	ETS1	
BACH2	DPYSL3	CDH5	WAS	IFNAR2	
UBE2R2	CLEC11A	ACVRL1	SH2D2A	PPP1R18	
RASL12	LSP1	PRDM1	SCARF1	CHSY1	

B.3.3 Ontology enrichment results for genes significantly upregulated in metastasis compared to primary

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	old Enrichme	Bonferroni	Benjamini	FDR
GOTERM_BP_DIRECT	GO:0007268~chemical synaptic transmission	17	4.27135678	1.22E-05	GRIA2, SLC12A5, DOC2A, SLC1A2, SLC6A1, SYN2, GRIN2B, SYN1, GRM1, GRIN1, GRM3, KIF5A, PLP1, AMPH, MBP, BSN, GABRD	315	240	16792	3.77597884	0.01697607	0.01093316	0.01093316
GOTERM_BP_DIRECT	GO:0007417~central nervous system development	12	3.01507538	1.56E-05	ARNT2, MOG, PTPRZ1, NCAN, ADAM22, CNTN2, MBP, TAGLN3, S100B, HPCAL4, TIMP4, HAPLN2	315	120	16792	5.33079365	0.02162916	0.01093316	0.01093316
GOTERM_BP_DIRECT	GO:0007628~adult walking behavior	5	1.25628141	0.00253399	ZIC1, CNTN2, CEND1, CTNS, MAPT	315	31	16792	8.59805428	0.9716248	0.86128166	0.86128166
GOTERM_BP_DIRECT	GO:0006811~ion transport	9	2.26130653	0.00267948	SLC24A2, PDZD3, SLC12A5, PLLP, SLC1A2, LCN2, CTNS, SLC17A7, TMC4	315	127	16792	3.77772778	0.97687946	0.86128166	0.86128166
GOTERM_BP_DIRECT	GO:0007411~axon guidance	10	2.51256281	0.00306724	MEG3, KIF5C, KIF5A, CNTN2, L1CAM, WNT3, ARHGAP35, CDK5R1, SPTBN2, MAPK3	315	159	16792	3.35270041	0.98660704	0.86128166	0.86128166
GOTERM_BP_DIRECT	GO:0007612~learning	6	1.50753769	0.00418874	JPH4, SLC24A2, SLC12A5, CNTN2, SLC6A1, STRA6	315	57	16792	5.61136174	0.99724224	0.94185777	0.94185777
GOTERM_BP_DIRECT	GO:0009636~response to toxic substance	7	1.75879397	0.00517677	PDZD3, CHKA, CDH1, PON1, MBP, SLC6A1, MAPK3	315	85	16792	4.39006536	0.99931566	0.94185777	0.94185777
GOTERM_BP_DIRECT	GO:0000381~regulation of alternative mRNA splicing, via spliceosome	5	1.25628141	0.00536671	RBM4, RBFOX1, RBFOX3, CELF4, SRSF2	315	38	16792	7.01420217	0.99947659	0.94185777	0.94185777
GOTERM_BP_DIRECT	GO:0006810~transport	15	3.76884422	0.00606308	SLC12A5, SLC10A1, ATP1B2, TRPV1, SLC6A1, GRIN2B, CLCN2, SEC14L2, ATCAY, CLVS2, PMP2, TTPAL, CNIH2, SLC25A4, GABRD	315	348	16792	2.29775588	0.99980421	0.94584107	0.94584107
GOTERM_BP_DIRECT	GO:0051966~regulation of synaptic transmission, glutamatergic	4	1.00502513	0.00869274	GRM3, GRM8, ATP1A2, GRM1	315	23	16792	9.27094548	0.99999525	1	1
GOTERM_BP_DIRECT	GO:0061180~mammary gland epithelium development	3	0.75376884	0.00905931	MSX1, WNT3, WNT4	315	8	16792	19.9904762	0.99999718	1	1
GOTERM_BP_DIRECT	GO:0007420~brain development	10	2.51256281	0.00962536	ARNT2, NDRG4, ZIC2, STK36, ZIC1, CTNS, SLC17A7, APOD, TBR1, CDK5R1	315	190	16792	2.80568087	0.99999873	1	1
GOTERM_BP_DIRECT	GO:0035235~ionotropic glutamate receptor signaling pathway	4	1.00502513	0.00979908	GRIA2, GRIN2B, CDK5R1, GRIN1	315	24	16792	8.88465608	0.99999901	1	1
GOTERM_BP_DIRECT	GO:0031175~neuron projection development	7	1.75879397	0.01119981	GNAO1, ATCAY, CDH1, EFHD1, STMN4, NCDN, CDK5R1	315	100	16792	3.73155556	0.99999986	1	1
GOTERM_BP_DIRECT	GO:1990573~potassium ion import across plasma membrane	3	0.75376884	0.01150472	ATP1A2, HCN2, KCNJ3	315	9	16792	17.7693122	0.99999991	1	1
GOTERM_BP_DIRECT	GO:0042391~regulation of membrane potential	6	1.50753769	0.01317061	KCNH3, RIMS4, NEDD4L, CNIH2, HCN2, GRIN1	315	75	16792	4.26463492	0.99999999	1	1
GOTERM_BP_DIRECT	GO:0010107~potassium ion import	4	1.00502513	0.01501498	KCNJ9, ATP1A2, ATP1B2, KCNJ3	315	28	16792	7.6154195	1	1	1
GOTERM_BP_DIRECT	GO:0019233~sensory perception of pain	5	1.25628141	0.01606534	FAM134B, SCN3B, GRM1, GRIN1, MAPK3	315	52	16792	5.12576313	1	1	1
GOTERM_BP_DIRECT	GO:0034220~ion transmembrane transport	10	2.51256281	0.0175365	SLC24A2, GRIA2, ATP6V1G2, TTYH1, SCNN1A, NEDD4L, ATP1A2, ATP1B2, GRIN2B, CLCN2	315	210	16792	2.53847317	1	1	1
GOTERM_BP_DIRECT	GO:0043547~positive regulation of GTPase activity	19	4.77386935	0.02025076	ARHGEF33, DOCK5, TRIO, PLEKHG6, IQSEC3, GRIN2B, ARHGAP35, GRIN1, GNAO1, ARHGAP32, ERBB3, PSD2, ARHGAP40, AMPH, CAMK2G, MCF2L2, NGEF, SPTBN2, WNT4	315	565	16792	1.79265627	1	1	1
GOTERM_BP_DIRECT	GO:0031669~cellular response to nutrient levels	3	0.75376884	0.02032719	RPTOR, KCNB1, LCN2	315	12	16792	13.3269841	1	1	1
GOTERM_BP_DIRECT	GO:0048013~ephrin receptor signaling pathway	6	1.50753769	0.02255126	NTRK3, EPHA1, GRIN2B, NGEF, CDK5R1, GRIN1	315	86	16792	3.71915836	1	1	1
GOTERM_BP_DIRECT	GO:0086012~membrane depolarization during cardiac muscle cell action potential	3	0.75376884	0.02373002	ATP1A2, SCN3B, HCN2	315	13	16792	12.3018315	1	1	1
GOTERM_BP_DIRECT	GO:0007416~synapse assembly	5	1.25628141	0.02720611	KIRREL3, PCLO, CDH1, BSN, SPTBN2	315	61	16792	4.36950299	1	1	1
GOTERM_BP_DIRECT	GO:0007611~learning or memory	4	1.00502513	0.02934782	PTPRZ1, THRA, S100B, GRIN2B	315	36	16792	5.92310406	1	1	1
GOTERM_BP_DIRECT	GO:0048168~regulation of neuronal synaptic plasticity	3	0.75376884	0.03519222	CNTN2, NCDN, S100B	315	16	16792	9.9952381	1	1	1
GOTERM_BP_DIRECT	GO:0007409~axonogenesis	6	1.50753769	0.03680393	LINGO1, BRSK2, ATP8A2, PTPRZ1, CNTN2, S100B	315	98	16792	3.26375121	1	1	1
GOTERM_BP_DIRECT	GO:0007413~axonal fasciculation	3	0.75376884	0.04378994	CNTN2, ARHGAP35, CDK5R1	315	18	16792	8.88465608	1	1	1
GOTERM_BP_DIRECT	GO:0030705~cytoskeleton-dependent intracellular transport	3	0.75376884	0.04378994	KIF5C, KIF5A, KIF1B	315	18	16792	8.88465608	1	1	1
GOTERM_BP_DIRECT	GO:0001764~neuron migration	6	1.50753769	0.04719585	KIRREL3, NTRK3, CNTN2, MAPT, KIAA0319, CDK5R1	315	105	16792	3.0461678	1	1	1
GOTERM_BP_DIRECT	GO:0035725~sodium ion transmembrane transport	5	1.25628141	0.04780808	SLC24A2, SCNN1A, SLC17A7, SCN3B, HCN2	315	73	16792	3.65122853	1	1	1
GOTERM_BP_DIRECT	GO:0006883~cellular sodium ion homeostasis	3	0.75376884	0.04835048	NEDD4L, ATP1A2, ATP1B2	315	19	16792	8.41704261	1	1	1
GOTERM_BP_DIRECT	GO:0051493~regulation of cytoskeleton organization	3	0.75376884	0.04835048	CAPN2, STMN4, MAPK3	315	19	16792	8.41704261	1	1	1
GOTERM_BP_DIRECT	GO:0008542~visual learning	4	1.00502513	0.05165459	NDRG4, CTNS, ATP1A2, GRIN1	315	45	16792	4.73848325	1	1	1
GOTERM_BP_DIRECT	GO:0061205~paramesonephric duct development	2	0.50251256	0.0550589	WNT4, STRA6	315	3	16792	35.5386243	1	1	1
GOTERM_BP_DIRECT	GO:0034765~regulation of ion transmembrane transport	6	1.50753769	0.05735563	KCNH3, KCNB1, KCNJ9, NEDD4L, HCN2, KCNJ3	315	111	16792	2.88151008	1	1	1
GOTERM_BP_DIRECT	GO:0035249~synaptic transmission, glutamatergic	3	0.75376884	0.05795532	SLC17A7, CNIH2, GRIN1	315	21	16792	7.6154195	1	1	1
GOTERM_BP_DIRECT	GO:0016079~synaptic vesicle exocytosis	3	0.75376884	0.0629845	TRIM9, PCLO, DNAJC5	315	22	16792	7.26926407	1	1	1
GOTERM_BP_DIRECT	GO:0035023~regulation of Rho protein signal transduction	5	1.25628141	0.06524609	ARHGEF33, TRIO, PLEKHG6, MCF2L2, NGEF	315	81	16792	3.29061336	1	1	1
GOTERM_BP_DIRECT	GO:0006814~sodium ion transport	5	1.25628141	0.06524609	SLC10A1, NEDD4L, ATP1A2, ATP1B2, SCN3B	315	81	16792	3.29061336	1	1	1
GOTERM_BP_DIRECT	GO:0006813~potassium ion transport	5	1.25628141	0.06763034	SLC24A2, KCNH3, ATP1A2, ATP1B2, KCNJ3	315	82	16792	3.25048393	1	1	1

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	old Enrichme	Bonferroni	Benjamini	FDR
GOTERM_BP_DIRECT	GO:1901798~positive regulation of signal transduction by p53 class mediator	2	0.50251256	0.07273187	HEXIM1, CHD5	315	4	16792	26.6539683	1	1	1
GOTERM_BP_DIRECT	GO:0051938~L-glutamate import	2	0.50251256	0.07273187	SLC1A2, SLC17A7	315	4	16792	26.6539683	1	1	1
GOTERM_BP_DIRECT	GO:0008344~adult locomotory behavior	4	1.00502513	0.07325529	EPS8, ADAM22, ATP1A2, GRIN1	315	52	16792	4.1006105	1	1	1
GOTERM_BP_DIRECT	GO:0034097~response to cytokine	4	1.00502513	0.07325529	GNAO1, TIMP4, PTGES, BCL2L1	315	52	16792	4.1006105	1	1	1
GOTERM_BP_DIRECT	GO:0019369~arachidonic acid metabolic process	3	0.75376884	0.07889299	DAGLA, DAGLB, MAPK3	315	25	16792	6.39695238	1	1	1
GOTERM_BP_DIRECT	GO:0046513~ceramide biosynthetic process	3	0.75376884	0.07889299	FAM57B, SPTSSA, CERS1	315	25	16792	6.39695238	1	1	1
GOTERM_BP_DIRECT	GO:0006656~phosphatidylcholine biosynthetic process	3	0.75376884	0.07889299	PCYT1A, CHKA, GPAT4	315	25	16792	6.39695238	1	1	1
GOTERM_BP_DIRECT	GO:0018105~peptidyl-serine phosphorylation	6	1.50753769	0.08554561	BRSK2, MAST1, STK39, CSNK1D, CDK5R1, MAPK3	315	125	16792	2.55878095	1	1	1
GOTERM_BP_DIRECT	GO:0032387~negative regulation of intracellular transport	2	0.50251256	0.09007534	MAPT, CRYAB	315	5	16792	21.3231746	1	1	1
GOTERM_BP_DIRECT	GO:0007155~cell adhesion	14	3.51758794	0.09147417	MOG, HEPACAM, NCAN, APLP1, ADAM22, ATP1B2, L1CAM, TM9SF4, HAPLN2, MAG, ITGA10, CNTN2, CNTNAP4, OPCML	315	459	16792	1.62595013	1	1	1
GOTERM_BP_DIRECT	GO:0014047~glutamate secretion	3	0.75376884	0.09589222	SLC1A2, SLC17A7, TRPV1	315	28	16792	5.71156463	1	1	1
GOTERM_BP_DIRECT	GO:0007616~long-term memory	3	0.75376884	0.09589222	CTNS, SLC17A7, GRIN1	315	28	16792	5.71156463	1	1	1
GOTERM_BP_DIRECT	GO:1902476~chloride transmembrane transport	5	1.25628141	0.09672807	SLC12A5, TTYH1, SLC17A7, GABRD, CLCN2	315	93	16792	2.86601809	1	1	1

B.3.4 Ontology enrichment results for genes significantly downregulated in metastasis compared to primary

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	old Enrichme	Bonferroni	Benjamini	FDR
GOTERM_BP_DIRECT	GO:0006954~inflammatory response	40	7.50469043	9.91E-14	CCL14, SEMA7A, NRROS, GPR68, IL24, PTGER3, ITGAL, AGER, PIK3CG, PSTPIP1, CNR2, TNFRSF8, CCR7, CCL19, CCR5, RELT, TNFRSF4, PTGDR, KLRG1, AOC3, PLA2G2D, PTGIR, TGFB1, XCR1, CCL21, TNFRSF9, RARRES2, NFAM1, SPHK1, TNFRSF1B, LYZ, SELE, SELP, ZAP70, BMP2, IL6, NCR3, THEMIS2, TLR10, MS4A2	427	379	16792	4.15045139	2.23E-10	1.61E-10	1.55E-10
GOTERM_BP_DIRECT	GO:0006955~immune response	42	7.87992495	1.42E-13	CCL14, CD274, CSF3, SEMA7A, NRROS, IL24, WAS, CD1E, SAMHD1, ETS1, CXCL14, JCHAIN, SPN, IL1RL1, C7, CNR2, CTLA4, TNFSF11, TNFRSF8, CCR7, CCL19, CCR5, ICOS, RELT, TNFRSF4, CCL21, TNFRSF9, OSM, IL16, LILRB2, TNFRSF1B, LAX1, HLA-E, CHIT1, ZAP70, IL6, NCR3, IRF8, TLR10, LTB, MS4A2, CD22	427	421	16792	3.92321171	3.21E-10	1.61E-10	1.55E-10
GOTERM_BP_DIRECT	GO:0007155~cell adhesion	42	7.87992495	2.44E-12	ROBO2, SRPX, COL15A1, SPON1, TNXB, COL12A1, DPT, ITGAL, WISP1, CYR61, PCDH17, WISP2, COMP, CDH5, PSTPIP1, HMCN2, ITGAX, SLAMF7, IGFBP7, EMILIN1, SVEP1, PLXNC1, SCARF1, AOC3, EGFL7, EGFL6, SELE, SELP, CD2, MFAP4, THEMIS2, COL5A1, COL6A2, ITGA11, CASS4, PECAM1, COL6A3, COL6A6, COL6A5, GAS6, CD22, SIGLEC6	427	459	16792	3.59841423	5.49E-09	1.83E-09	1.77E-09
GOTERM_BP_DIRECT	GO:0030198~extracellular matrix organization	26	4.87804878	3.12E-11	HPSE2, COL14A1, ELN, ITGAL, FGF2, CYR61, COMP, ABI3BP, ADAMTSL2, ITGAX, COL10A1, EMILIN1, EGFL6, FOXF2, LUM, FOXF1, COL3A1, CCDC80, COL1A2, COL5A1, CRISPLD2, COL6A2, COL5A2, ITGA11, PECAM1, COL6A3	427	196	16792	5.21665153	7.04E-08	1.76E-08	1.70E-08
GOTERM_BP_DIRECT	GO:0030574~collagen catabolic process	16	3.00187617	4.65E-11	COL15A1, MMP2, COL12A1, MRC2, MMP11, COL3A1, MMP14, COL1A2, COL5A1, COL6A2, COL5A2, MMP19, COL6A3, COL10A1, COL6A6, COL6A5	427	64	16792	9.83138173	1.05E-07	2.10E-08	2.03E-08
GOTERM_BP_DIRECT	GO:0002250~adaptive immune response	22	4.12757974	1.63E-10	TNFRSF13B, SH2D1A, FCRL4, RNF19B, LILRB1, CD1E, TNFRSF13C, LILRB2, PRDM1, CD1C, LAX1, LILRB4, PIK3CG, JCHAIN, HLA-E, CD79A, ZAP70, CD6, CTLA4, SLAMF7, CSK, JAK3	427	148	16792	5.84568644	3.67E-07	6.12E-08	5.93E-08
GOTERM_BP_DIRECT	GO:0030199~collagen fibril organization	12	2.25140713	2.18E-09	GREM1, MMP11, COL3A1, SFRP2, TNXB, COL1A2, COL5A1, LUM, COL14A1, COL12A1, COL5A2, DPT	427	39	16792	12.1001621	4.91E-06	7.01E-07	6.79E-07
GOTERM_BP_DIRECT	GO:0007165~signal transduction	60	11.2570356	2.34E-07	ARHGAP9, ACVRL1, CD83, PDE3B, LSP1, ITGAL, ARHGAP6, FGF2, ANTXR1, CXCL14, WISP1, WISP2, SPN, MRC2, FGF7, PSTPIP1, TAGAP, TNFRSF8, SCARF1, GRASP, IL15RA, SPHK1, NFAM1, GREM1, PECAM1, VOPP1, RIN3, CD48, PGR, ERG, LTB, ZNF831, GAS6, DTHD1, CD274, COL15A1, SH2D2A, CSF2RB, RASAL3, RASIP1, IL1RL1, SH3BP1, PDPN, IL12RB1, SH2B3, BCL11B, TIE1, LILRB1, IGF1, LILRB2, ARHGAP25, GNG11, LILRB4, ARHGAP30, P2RX1, SCGB1A1, IL2RB, PDCD1, PKIG, ASB2	427	1161	16792	2.03232697	5.26E-04	6.58E-05	6.37E-05
GOTERM_BP_DIRECT	GO:0008285~negative regulation of cell proliferation	30	5.62851782	3.41E-07	ACVRL1, IL24, DPT, ADARB1, ETS1, FGF2, TSPAN32, CDH5, ADAMTS1, TNFRSF8, IGFBP7, CSK, SLIT3, PIM2, ADAMTS8, MYOCD, TGFB1, BCL11B, TNFRSF9, OSM, INHBA, PODN, SOD2, TBX5, FOXP3, BMP2, IL6, SFRP2, RGCC, WT1	427	396	16792	2.97920659	7.69E-04	8.55E-05	8.28E-05
GOTERM_BP_DIRECT	GO:0031295~T cell costimulation	13	2.43902439	5.82E-07	CD274, CCL21, CD3G, TNFRSF13C, CD3E, SPN, GRAP2, CD28, CTLA4, CSK, PDCD1, CCL19, ICOS	427	78	16792	6.55425449	0.00131114	1.31E-04	1.27E-04
GOTERM_BP_DIRECT	GO:0042110~T cell activation	10	1.87617261	2.32E-06	CD2, ZAP70, IRF4, WAS, CD48, CD3G, NLRC3, CD3E, FOXP3, PIK3CG	427	47	16792	8.36713339	0.00521231	4.75E-04	4.60E-04
GOTERM_BP_DIRECT	GO:0042130~negative regulation of T cell proliferation	9	1.68855535	3.25E-06	SPN, PLA2G2D, TGFB1, SFTPD, SCGB1A1, CTLA4, LILRB1, LILRB2, FOXP3	427	37	16792	9.56566871	0.00730334	6.11E-04	5.92E-04
GOTERM_BP_DIRECT	GO:0001525~angiogenesis	20	3.75234522	4.32E-06	ACVRL1, COL15A1, EGFL7, TIE1, SH2D2A, MMP2, PDE3B, MEOX2, TBX4, RASIP1, PIK3CG, TYMP, FMNL3, MMP14, MEIS1, TAL1, SOX18, GJA5, PECAM1, MMP19	427	223	16792	3.52695309	0.00969977	7.12E-04	6.89E-04

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	old Enrichme	Bonferroni	Benjamini	FDR
GOTERM_BP_DIRECT	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	50	9.38086304	4.42E-06	ACVRL1, CSF3, TCF21, IKZF1, GLI1, MEOX2, IKZF3, ETS1, FGF2, MEOX1, CYR61, GLI2, SOX18, TNFSF11, MYOCD, FOXP3, NR5A1, GREM1, SFRP2, MAF, IRF4, TAL1, PGR, AKNA, GATA6, WNT2, IL33, EGR2, TGFB1, JUND, BCL11B, FOXF2, LUM, FOXF1, OSM, NFATC2, LILRB1, IGF1, INHBA, POU2F2, TBX5, FLI1, SMAD7, BMP2, IL6, MEIS1, RGCC, NR4A3, WT1, CD28	427	981	16792	2.00435917	0.00991459	7.12E-04	6.89E-04
GOTERM_BP_DIRECT	GO:0030336~negative regulation of cell migration	13	2.43902439	4.97E-06	ACVRL1, TIE1, IL24, ADIPOQ, ADARB1, PODN, SULF1, TBX5, SMAD7, DACH1, SFRP2, DPYSL3, SLIT2	427	95	16792	5.3813879	0.01113169	7.46E-04	7.23E-04
GOTERM_BP_DIRECT	GO:0007267~cell-cell signaling	21	3.93996248	8.00E-06	PTGIR, CCL21, SH2D1A, INHBA, LILRB2, TBX5, CXCL14, CYR61, WISP1, TSPAN32, WISP2, NR5A1, GREM1, BMP2, SFRP2, GRAP2, PGR, LTB, CCR5, WNT2, SIGLEC6	427	254	16792	3.25132309	0.01786868	0.00112689	0.00109139
GOTERM_BP_DIRECT	GO:0042102~positive regulation of T cell proliferation	10	1.87617261	1.88E-05	SPN, CD274, IL6, CD6, CD28, LILRB2, TNFRSF13C, CCL19, CD3E, JAK3	427	60	16792	6.55425449	0.04147781	0.00238989	0.00231461
GOTERM_BP_DIRECT	GO:0032496~response to lipopolysaccharide	16	3.00187617	1.91E-05	PTGIR, JUND, IL10RA, TNFRSF9, CSF2RB, SOD2, SELE, SELP, CNR2, CD6, SCGB1A1, TNFRSF8, ACP5, CCR7, RELT, TNFRSF4	427	164	16792	3.83663677	0.04210624	0.00238989	0.00231461
GOTERM_BP_DIRECT	GO:0030324~lung development	11	2.06378987	2.13E-05	HSD11B1, MMP14, CRISPLD2, FOXF1, PDPN, TBX5, GLI1, WNT2, FGF2, TBX4, GLI2	427	76	16792	5.69185258	0.04681164	0.00252328	0.00244379
GOTERM_BP_DIRECT	GO:0006968~cellular defense response	10	1.87617261	2.47E-05	SPN, FCMR, NCF1, GNLY, CD19, SH2D1A, LILRB2, LSP1, CCR5, KLRG1	427	62	16792	6.34282692	0.05407189	0.0027794	0.00269185
GOTERM_BP_DIRECT	GO:0050776~regulation of immune response	16	3.00187617	4.98E-05	SH2D1A, SFTPD, LILRB1, CD3G, LILRB2, ITGAL, CD1C, CD3E, HLA-E, COL3A1, NCR3, COL1A2, CD19, CD300LB, CD300E, SLAMF7	427	178	16792	3.53487883	0.10626183	0.00526372	0.00509792
GOTERM_BP_DIRECT	GO:0051897~positive regulation of protein kinase B signaling	11	2.06378987	5.14E-05	CSF3, IL6, TGFB1, CCL21, CD28, TNFSF11, CCR7, CCL19, GAS6, FGF2, PIK3CG	427	84	16792	5.14977138	0.10935103	0.00526372	0.00509792
GOTERM_BP_DIRECT	GO:0050853~B cell receptor signaling pathway	9	1.68855535	6.01E-05	BLK, IGLL5, CD79A, ZAP70, KLHL6, CD19, NFAM1, CTLA4, NFATC2	427	54	16792	6.55425449	0.12672297	0.0056681	0.00548956
GOTERM_BP_DIRECT	GO:0032689~negative regulation of interferon-gamma production	7	1.31332083	6.04E-05	IL33, IL1RL1, CD274, SCGB1A1, LILRB1, GAS6, FOXP3	427	28	16792	9.83138173	0.12719095	0.0056681	0.00548956
GOTERM_BP_DIRECT	GO:0006959~humoral immune response	9	1.68855535	8.93E-05	IL6, CD83, POU2AF1, SH2D1A, CD28, PDCD1, POU2F2, MS4A1, LTF	427	57	16792	6.20929373	0.18240448	0.00805514	0.00780141
GOTERM_BP_DIRECT	GO:0045087~innate immune response	26	4.87804878	1.09E-04	BLK, NRROS, NCF1, ADARB1, PRDM1, LY9, AGER, PIK3CG, JCHAIN, PSTPIP1, CD300LB, CSK, JAK3, KLRG1, ZBP1, SH2D1A, SFTPD, MATK, SSCSD, HLA-E, IGLL5, ZAP70, CD6, SERPING1, CD300E, TLR10	427	430	16792	2.37782256	0.21839566	0.00947666	0.00917815
GOTERM_BP_DIRECT	GO:0045060~negative thymic T cell selection	5	0.9380863	1.17E-04	SPN, ZAP70, CD28, CCR7, CD3E	427	11	16792	17.8752395	0.23180905	0.00976672	0.00945908
GOTERM_BP_DIRECT	GO:0019221~cytokine-mediated signaling pathway	13	2.43902439	1.27E-04	IL15RA, LRRC15, CSF3, IL10RA, PODN, SOCS2, SOCS3, IL6, PODNL1, SOCS1, IL2RB, TNFSF11, IL12RB1	427	131	16792	3.90253321	0.2488294	0.01006276	0.00974579
GOTERM_BP_DIRECT	GO:0045669~positive regulation of osteoblast differentiation	9	1.68855535	1.29E-04	IL6, BMP2, SFRP2, JUND, TMEM119, IGF1, FGF2, CYR61, LTF	427	60	16792	5.89882904	0.25311114	0.01006276	0.00974579
GOTERM_BP_DIRECT	GO:0070374~positive regulation of ERK1 and ERK2 cascade	15	2.81425891	1.54E-04	CCL14, PDGFRA, SEMA7A, TGFB1, CCL21, PLA2G2A, PLA2G5, SCIMP, FGF2, BMP2, IL6, GLIPR2, CCR7, CCL19, GAS6	427	175	16792	3.37075945	0.29286866	0.01155041	0.01118657
GOTERM_BP_DIRECT	GO:0050728~negative regulation of inflammatory response	10	1.87617261	1.71E-04	SOCS3, PTGIS, CNR2, FOXF1, ADIPOQ, ACP5, WFDC1, METRNL, TNFRSF1B, ETS1	427	79	16792	4.9779148	0.32012585	0.01244563	0.0120536
GOTERM_BP_DIRECT	GO:0001558~regulation of cell growth	10	1.87617261	1.89E-04	SOCS2, HTRA3, HTRA4, TMC8, WFDC1, IGFBP7, CYR61, WISP1, PAPPA2, WISP2	427	80	16792	4.91569087	0.34634906	0.01325503	0.0128375
GOTERM_BP_DIRECT	GO:0001501~skeletal system development	13	2.43902439	1.94E-04	COL12A1, AEBP1, IGF1, TBX3, GLI2, ADAMTS4, COMP, COL3A1, BMP2, COL1A2, GJA5, COL5A2, COL10A1	427	137	16792	3.73161934	0.3543246	0.01325503	0.0128375
GOTERM_BP_DIRECT	GO:0007166~cell surface receptor signaling pathway	19	3.56472795	2.32E-04	IFNAR2, CD274, LAG3, TNFRSF13B, NPR1, GFRA1, CD3G, INHBA, LILRB2, CD3E, AGER, TSPAN32, SPN, CD2, CD19, CD28, CD37, CCR5, KLRG1	427	274	16792	2.7269526	0.40690376	0.01536288	0.01487896
GOTERM_BP_DIRECT	GO:0030308~negative regulation of cell growth	12	2.25140713	2.57E-04	GREM1, ZC3H12D, ACVRL1, TGFB1, SFRP2, NPR1, WT1, FHL1, WFDC1, SLIT3, INHBA, SLIT2	427	121	16792	3.90005226	0.44031171	0.01658002	0.01605775
GOTERM_BP_DIRECT	GO:0002576~platelet degranulation	11	2.06378987	2.86E-04	SELP, CLEC3B, TGFB1, RARRES2, PLEK, PECAM1, SERPING1, TIMP3, IGF1, A2M, GAS6	427	103	16792	4.19981356	0.47571077	0.01793387	0.01736896
GOTERM_BP_DIRECT	GO:0001568~blood vessel development	7	1.31332083	3.57E-04	EGFL7, COL1A2, COL5A1, FOXF1, SPHK1, ANTXR1, TBX3	427	38	16792	7.24417601	0.55277748	0.02174474	0.02105979
GOTERM_BP_DIRECT	GO:0001570~vasculogenesis	8	1.50093809	5.11E-04	MYOCD, EGFL7, TGFB1, WT1, FOXF1, SOX18, TIE1, RASIP1	427	56	16792	5.61793242	0.68388235	0.0302986	0.0293442

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	old Enrichme	Bonferroni	Benjamini	FDR
GOTERM_BP_DIRECT	GO:0009887~organ morphogenesis	10	1.87617261	5.41E-04	COMP, BMP2, ELN, FHL1, ITGAX, TNFSF11, SLIT3, FGF2, FLI1, TBX3	427	92	16792	4.2745138	0.70473855	0.03127087	0.03028585
GOTERM_BP_DIRECT	GO:0022617~extracellular matrix disassembly	9	1.68855535	6.71E-04	ADAMTS4, MMP11, MMP14, ELN, MMP2, KLK4, MMP19, A2M, TPSAB1	427	76	16792	4.65697029	0.7795849	0.03706085	0.03589345
GOTERM_BP_DIRECT	GO:0006366~transcription from RNA polymerase II promoter	27	5.06566604	6.74E-04	AKNA, GATA6, TCF21, MEOX2, ETS1, IKZF3, MEOX1, BACH2, GLI2, MSC, MYOCD, EGR2, JUND, BCL11B, FOXF2, FOXF1, NFATC2, POU2F2, RUNX3, FLI1, NR5A1, MEIS1, MAF, WT1, IRF4, POU2AF1, ZNF831	427	513	16792	2.06976458	0.78128972	0.03706085	0.03589345
GOTERM_BP_DIRECT	GO:0045893~positive regulation of transcription, DNA-templated	27	5.06566604	7.18E-04	ACVRL1, GATA6, GLI1, ETS1, FGF2, GLI2, FGF7, LBH, SOX18, PIM2, MYOCD, EGR2, TGFB1, FOXF2, FOXF1, NFATC2, IGF1, INHBA, TBX5, RUNX3, FOXP3, TBX3, BMP2, IL6, WT1, IRF4, TAL1	427	515	16792	2.06172665	0.80192274	0.03853611	0.03732224
GOTERM_BP_DIRECT	GO:0002407~dendritic cell chemotaxis	5	0.9380863	7.48E-04	CCL21, CCR7, CCL19, CCR5, PIK3CG	427	17	16792	11.5663315	0.8147148	0.03919137	0.03795685
GOTERM_BP_DIRECT	GO:0001503~ossification	9	1.68855535	9.44E-04	IGSF10, CLEC3B, COL5A2, TMEM119, TNFSF11, CHRDL2, RUNX3, CHRDL1, LTF	427	80	16792	4.42412178	0.88111425	0.04837697	0.04685312
GOTERM_BP_DIRECT	GO:0043552~positive regulation of phosphatidylinositol 3-kinase activity	6	1.12570356	0.0010117	PDGFRA, TGFB1, CCL21, CCR7, CCL19, FGF2	427	31	16792	7.61139231	0.8978711	0.05067478	0.04907855
GOTERM_BP_DIRECT	GO:0043066~negative regulation of apoptotic process	24	4.50281426	0.00139072	BLK, FCMR, EGR3, BCL11B, SPHK1, GATA6, TWIST2, IGF1, SOD2, CYR61, TBX3, GLI2, SOCS2, GREM1, COMP, SOCS3, IL6, WT1, IL2RB, PIM2, PDCD1, GAS6, LIMS2, LTF	427	455	16792	2.07431351	0.95658113	0.0681452	0.06599866
GOTERM_BP_DIRECT	GO:0032753~positive regulation of interleukin-4 production	5	0.9380863	0.00143306	IL33, CD28, CD3E, FOXP3, HLA-E	427	20	16792	9.83138173	0.96053902	0.06872609	0.06656125
GOTERM_BP_DIRECT	GO:0007157~heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules	7	1.31332083	0.00159089	SELP, NCR3, CD6, FAT4, TIGIT, ITGAL, SELE	427	50	16792	5.50557377	0.97236629	0.07470549	0.0723523
GOTERM_BP_DIRECT	GO:0048565~digestive tract development	6	1.12570356	0.00177885	MYOCD, COL3A1, TGFB1, FOXF1, CLDN18, FAT4	427	35	16792	6.7415189	0.98192255	0.08182693	0.07924942
GOTERM_BP_DIRECT	GO:0061364~apoptotic process involved in luteolysis	3	0.56285178	0.00189394	ROBO2, SLIT3, SLIT2	427	3	16792	39.3255269	0.98606005	0.08370488	0.08106821
GOTERM_BP_DIRECT	GO:0010560~positive regulation of glycoprotein biosynthetic process	3	0.56285178	0.00189394	CCL21, IGF1, CCL19	427	3	16792	39.3255269	0.98606005	0.08370488	0.08106821
GOTERM_BP_DIRECT	GO:0060045~positive regulation of cardiac muscle cell proliferation	5	0.9380863	0.00207868	GATA6, TBX5, WNT2, FGF2, TBX2	427	22	16792	8.93761976	0.9908153	0.09002324	0.08718754
GOTERM_BP_DIRECT	GO:0097190~apoptotic signaling pathway	8	1.50093809	0.00211678	SPN, PTGIS, CD28, TNFRSF8, CD3E, TNFRSF1B, RELT, TNFRSF4	427	71	16792	4.43104529	0.99157278	0.09002324	0.08718754
GOTERM_BP_DIRECT	GO:0045786~negative regulation of cell cycle	6	1.12570356	0.00229153	BMP2, TGFB1, LILRB1, INHBA, RUNX3, ETS1	427	37	16792	6.37711248	0.99432132	0.09088858	0.08802563
GOTERM_BP_DIRECT	GO:0051024~positive regulation of immunoglobulin secretion	4	0.75046904	0.00229842	IL33, IL6, TNFRSF4, HLA-E	427	11	16792	14.3001916	0.99440908	0.09088858	0.08802563
GOTERM_BP_DIRECT	GO:0050868~negative regulation of T cell activation	4	0.75046904	0.00229842	LAG3, TIGIT, LAX1, JAK3	427	11	16792	14.3001916	0.99440908	0.09088858	0.08802563
GOTERM_BP_DIRECT	GO:0033690~positive regulation of osteoblast proliferation	4	0.75046904	0.00229842	BMP2, TMEM119, CYR61, LTF	427	11	16792	14.3001916	0.99440908	0.09088858	0.08802563
GOTERM_BP_DIRECT	GO:0050729~positive regulation of inflammatory response	8	1.50093809	0.00248389	CCL14, IL33, IL1RL1, FABP4, PLA2G2A, TLR10, ETS1, PLA2G7	427	73	16792	4.30964679	0.99632298	0.09652911	0.09348849
GOTERM_BP_DIRECT	GO:0001657~ureteric bud development	6	1.12570356	0.00258465	ROBO2, TGFB1, WT1, TCF21, SLIT2, SMAD7	427	38	16792	6.20929373	0.99707173	0.09709667	0.09403817
GOTERM_BP_DIRECT	GO:0010811~positive regulation of cell-substrate adhesion	6	1.12570356	0.00258465	EGFL6, CCDC80, FOXF1, ABI3BP, EMILIN1, CYR61	427	38	16792	6.20929373	0.99707173	0.09709667	0.09403817
GOTERM_BP_DIRECT	GO:0030335~positive regulation of cell migration	13	2.43902439	0.00262841	PDGFRA, CD274, LRRC15, SEMA7A, TGFB1, FOXF1, SPHK1, IGF1, GLI1, CYR61, MYLK, MMP14, BMP2	427	184	16792	2.77843397	0.99734745	0.09712189	0.0940626
GOTERM_BP_DIRECT	GO:0000122~negative regulation of transcription from RNA polymerase II promoter	32	6.00375235	0.00297484	GATA6, TCF21, IKZF1, AEBP1, PRDM1, BACH2, GLI2, DACH1, SOX18, MSC, LMCD1, DACT1, MYOCD, TGFB1, JUND, SKOR2, FOXF1, NFATC2, RUNX3, FOXP3, TBX3, SMAD7, TBX2, BMP2, NR4A3, MAF, WT1, TAL1, BCL6B, SCGB1A1, IRF8, PKIG	427	720	16792	1.7478012	0.99878775	0.10428181	0.10099698
GOTERM_BP_DIRECT	GO:0050921~positive regulation of chemotaxis	4	0.75046904	0.00300724	TGFB1, CCL21, RARRES2, CCL19	427	12	16792	13.108509	0.99887338	0.10428181	0.10099698
GOTERM_BP_DIRECT	GO:0032693~negative regulation of interleukin-10 production	4	0.75046904	0.00300724	CD274, AGER, JAK3, FOXP3	427	12	16792	13.108509	0.99887338	0.10428181	0.10099698
GOTERM_BP_DIRECT	GO:0050777~negative regulation of immune response	4	0.75046904	0.00300724	COL3A1, CTLA4, LILRB2, FOXP3	427	12	16792	13.108509	0.99887338	0.10428181	0.10099698
GOTERM_BP_DIRECT	GO:0044267~cellular protein metabolic process	10	1.87617261	0.00313227	SFTPC, MMP2, SFTPD, GATA6, CSF2RB, IGF1, LYZ, LMCD1, PAPPA2, LTF	427	118	16792	3.33267177	0.9991508	0.10697168	0.10360212
GOTERM_BP_DIRECT	GO:0046426~negative regulation of JAK-STAT cascade	6	1.12570356	0.00325095	SOCS2, LRRC15, SOCS3, PODNL1, SOCS1, PODN	427	40	16792	5.89882904	0.99935067	0.10936775	0.10592271
GOTERM_BP_DIRECT	GO:0001816~cytokine production	5	0.9380863	0.00338551	FABP4, MAF, NFATC2, FOXP3, PIK3CG	427	25	16792	7.86510539	0.99952103	0.11221955	0.10868468
GOTERM_BP_DIRECT	GO:0070527~platelet aggregation	6	1.12570356	0.00362674	PDGFRA, PLEK, GAS6, MYL9, TSPAN32, PIK3CG	427	41	16792	5.75495516	0.99972245	0.11603433	0.1123793
GOTERM_BP_DIRECT	GO:0051414~response to cortisol	3	0.56285178	0.00372425	IGFBP7, SLIT3, SLIT2	427	4	16792	29.4941452	0.9997774	0.11603433	0.1123793
GOTERM_BP_DIRECT	GO:0070232~regulation of T cell apoptotic process	3	0.56285178	0.00372425	CD274, JAK3, GIMAP8	427	4	16792	29.4941452	0.9997774	0.11603433	0.1123793

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	old Enrichme	Bonferroni	Benjamini	FDR
GOTERM_BP_DIRECT	GO:2001205~negative regulation of osteoclast development	3	0.56285178	0.00372425	CLDN18, LILRB1, LTF	427	4	16792	29.4941452	0.9997774	0.11603433	0.1123793
GOTERM_BP_DIRECT	GO:0007275~multicellular organism development	25	4.69043152	0.00378314	CSF3, TWIST2, MEOX2, MEOX1, DACH1, LBH, HLX, TNFRSF8, RELT, WNT2, TNFRSF4, EGFL6, TNFRSF9, OSM, SSC5D, TNFRSF1B, TBX4, TDRD1, IGSF10, MMP11, SFRP2, ANGPTL2, PDCD1, ERG, ZNF831	427	521	16792	1.88702145	0.99980517	0.11603433	0.1123793
GOTERM_BP_DIRECT	GO:0001817~regulation of cytokine production	4	0.75046904	0.00383638	CD83, SFTPD, CSK, LTF	427	13	16792	12.1001621	0.99982728	0.11603433	0.1123793
GOTERM_BP_DIRECT	GO:0006935~chemotaxis	10	1.87617261	0.00391243	SPN, XCR1, RARRES2, CCR7, LSP1, CCR5, FGF2, CXCL14, CYR61, TYMP	427	122	16792	3.22340385	0.99985458	0.11603433	0.1123793
GOTERM_BP_DIRECT	GO:0050900~leukocyte migration	10	1.87617261	0.00391243	CD2, SELP, SPN, COL1A2, PECAM1, ITGAX, CD48, ITGAL, GAS6, SELE	427	122	16792	3.22340385	0.99985458	0.11603433	0.1123793
GOTERM_BP_DIRECT	GO:0001658~branching involved in ureteric bud morphogenesis	6	1.12570356	0.00403269	GREM1, BMP2, WT1, TCF21, FAT4, FGF2	427	42	16792	5.61793242	0.99988923	0.1170371	0.11335048
GOTERM_BP_DIRECT	GO:0008284~positive regulation of cell proliferation	23	4.315197	0.00405009	CCL14, PDGFRA, CSF3, MYOCD, TGFB1, HPSE2, IL24, CLEC11A, OSM, MATK, IGF1, GLI1, ETS1, FGF2, TBX3, TBX2, GREM1, FGF7, IL6, SFRP2, HLX, S1PR2, WNT2	427	466	16792	1.94095948	0.99989351	0.1170371	0.11335048
GOTERM_BP_DIRECT	GO:0009967~positive regulation of signal transduction	7	1.31332083	0.00440305	SOCS2, SH2D2A, GRAP2, SH2D1A, ADIPOQ, SH2B3, ARHGAP6	427	61	16792	4.51276539	0.9999521	0.12562629	0.12166912
GOTERM_BP_DIRECT	GO:0060349~bone morphogenesis	5	0.9380863	0.00451325	SFRP2, CHSY1, ACP5, PAPPA2, LTF	427	27	16792	7.28250499	0.99996268	0.12574716	0.12178618
GOTERM_BP_DIRECT	GO:0016477~cell migration	12	2.25140713	0.00451886	FMNL3, PSTPIP1, TGFB1, COL5A1, PALLD, MATK, NFATC2, CSK, ERG, ADAMTS12, GAS6, JAK3	427	172	16792	2.74364141	0.99996315	0.12574716	0.12178618
GOTERM_BP_DIRECT	GO:0050731~positive regulation of peptidyl-tyrosine phosphorylation	8	1.50093809	0.00477486	CSF3, IL6, FGF7, TGFB1, ADIPOQ, OSM, IGF1, CD3E	427	82	16792	3.83663677	0.99997936	0.13012261	0.1260238
GOTERM_BP_DIRECT	GO:0043537~negative regulation of blood vessel endothelial cell migration	4	0.75046904	0.00479156	ACVRL1, TGFB1, RGCC, FGF2	427	14	16792	11.2358648	0.99998013	0.13012261	0.1260238
GOTERM_BP_DIRECT	GO:0030514~negative regulation of BMP signaling pathway	6	1.12570356	0.00544439	GREM1, SKOR2, SFRP2, HTRA3, CHRDL1, SMAD7	427	45	16792	5.24340359	0.99999547	0.14437229	0.13982463
GOTERM_BP_DIRECT	GO:0032870~cellular response to hormone stimulus	6	1.12570356	0.00544439	SOCS2, ROBO2, JUND, IGFBP7, SLIT3, SLIT2	427	45	16792	5.24340359	0.99999547	0.14437229	0.13982463
GOTERM_BP_DIRECT	GO:0002042~cell migration involved in sprouting angiogenesis	4	0.75046904	0.00587785	GREM1, EGR3, SLIT2, FGF2	427	15	16792	10.4868072	0.9999983	0.1511672	0.1464055
GOTERM_BP_DIRECT	GO:0045165~cell fate commitment	6	1.12570356	0.00598382	BMP2, TAL1, GATA6, GAS1, PRDM1, WNT2	427	46	16792	5.12941656	0.99999867	0.1511672	0.1464055
GOTERM_BP_DIRECT	GO:0072540~T-helper 17 cell lineage commitment	3	0.56285178	0.00610302	IL6, IRF4, LY9	427	5	16792	23.5953162	0.99999898	0.1511672	0.1464055
GOTERM_BP_DIRECT	GO:0045085~negative regulation of interleukin-2 biosynthetic process	3	0.56285178	0.00610302	LAG3, SFTPD, FOXP3	427	5	16792	23.5953162	0.99999898	0.1511672	0.1464055
GOTERM_BP_DIRECT	GO:0045077~negative regulation of interferon-gamma biosynthetic process	3	0.56285178	0.00610302	LILRB1, INHBA, FOXP3	427	5	16792	23.5953162	0.99999898	0.1511672	0.1464055
GOTERM_BP_DIRECT	GO:0032713~negative regulation of interleukin-4 production	3	0.56285178	0.00610302	CD83, SCGB1A1, FOXP3	427	5	16792	23.5953162	0.99999898	0.1511672	0.1464055
GOTERM_BP_DIRECT	GO:0006952~defense response	7	1.31332083	0.00697279	CD83, SP140, WAS, CD48, INHBA, PTPRCAP, TPSAB1	427	67	16792	4.10863714	0.99999986	0.17083348	0.1654523
GOTERM_BP_DIRECT	GO:0036149~phosphatidylinositol acyl-chain remodeling	4	0.75046904	0.00709965	PLA2G2D, PLA2G4E, PLA2G2A, PLA2G5	427	16	16792	9.83138173	0.99999989	0.17207115	0.16665099
GOTERM_BP_DIRECT	GO:0007259~JAK-STAT cascade	5	0.9380863	0.00837281	SOCS2, IFNAR2, IL15RA, SOCS3, SOCS1	427	32	16792	6.14461358	0.99999999	0.20076927	0.19444512
GOTERM_BP_DIRECT	GO:2000107~negative regulation of leukocyte apoptotic process	3	0.56285178	0.00900139	CCL21, CCR7, CCL19	427	6	16792	19.6627635	1	0.21134508	0.2046878
GOTERM_BP_DIRECT	GO:0001768~establishment of T cell polarity	3	0.56285178	0.00900139	CCL21, CCR7, CCL19	427	6	16792	19.6627635	1	0.21134508	0.2046878
GOTERM_BP_DIRECT	GO:0043547~positive regulation of GTPase activity	25	4.69043152	0.00984896	ARHGAP9, CCL14, DENND1C, RGS16, LRRK2, CSF2RB, RASAL3, FGF2, ARHGAP6, FGF7, FGD5, SH3BP1, TAGAP, CCL19, A2M, JAK3, PDGFRA, PTGIR, ARHGEF15, GFRA1, ARHGAP25, ARHGAP30, CDC42EP5, IL2RB, RIN3	427	565	16792	1.74006756	1	0.22606738	0.21894636
GOTERM_BP_DIRECT	GO:0048557~embryonic digestive tract morphogenesis	4	0.75046904	0.00996452	PDGFRA, HLX, FOXF1, TCF21	427	18	16792	8.73900598	1	0.22606738	0.21894636
GOTERM_BP_DIRECT	GO:0036150~phosphatidylserine acyl-chain remodeling	4	0.75046904	0.00996452	PLA2G2D, PLA2G4E, PLA2G2A, PLA2G5	427	18	16792	8.73900598	1	0.22606738	0.21894636
GOTERM_BP_DIRECT	GO:0034097~response to cytokine	6	1.12570356	0.01002961	CD274, COL3A1, JUND, SCGB1A1, ACP5, PLA2G5	427	52	16792	4.5375608	1	0.22606738	0.21894636
GOTERM_BP_DIRECT	GO:0007585~respiratory gaseous exchange	5	0.9380863	0.01037856	DACH1, ELN, SFTPC, SFTPD, CSF2RB	427	34	16792	5.78316573	1	0.22776968	0.22059503
GOTERM_BP_DIRECT	GO:0048286~lung alveolus development	5	0.9380863	0.01037856	MYOCD, FOXF1, PDPN, SFTPD, TCF21	427	34	16792	5.78316573	1	0.22776968	0.22059503
GOTERM_BP_DIRECT	GO:0033209~tumor necrosis factor-mediated signaling pathway	9	1.68855535	0.01040829	TNFRSF13B, TNFRSF9, TNFSF11, TNFRSF8, LTB, TNFRSF13C, TNFRSF1B, RELT, TNFRSF4	427	118	16792	2.9994046	1	0.22776968	0.22059503
GOTERM_BP_DIRECT	GO:0042832~defense response to protozoan	4	0.75046904	0.01161355	IL6, IRF4, IRF8, TSPAN32	427	19	16792	8.2790583	1	0.25147069	0.24354948
GOTERM_BP_DIRECT	GO:0048146~positive regulation of fibroblast proliferation	6	1.12570356	0.01171447	PDGFRA, TGFB1, SPHK1, IGF1, GAS6, WNT2	427	54	16792	4.36950299	1	0.25147069	0.24354948
GOTERM_BP_DIRECT	GO:0048251~elastic fiber assembly	3	0.56285178	0.01239157	MFAP4, TNXB, MYH11	427	7	16792	16.8537973	1	0.2586167	0.25047039
GOTERM_BP_DIRECT	GO:0034616~response to laminar fluid shear stress	3	0.56285178	0.01239157	TGFB1, ETS1, SMAD7	427	7	16792	16.8537973	1	0.2586167	0.25047039
GOTERM_BP_DIRECT	GO:0002606~positive regulation of dendritic cell antigen processing and presentation	3	0.56285178	0.01239157	CCL21, CCR7, CCL19	427	7	16792	16.8537973	1	0.2586167	0.25047039

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	old Enrichme	Bonferroni	Benjamini	FDR
GOTERM_BP_DIRECT	GO:0001666~response to hypoxia	11	2.06378987	0.01251973	ACVRL1, MYOCD, MMP14, BMP2, TGFB1, MMP2, ADIPOQ, SOD2, ETS1, AGER, CYGB	427	172	16792	2.51500463	1	0.25889429	0.25073924
GOTERM_BP_DIRECT	GO:0060021~palate development	7	1.31332083	0.01264383	FOXF2, TCF21, INHBA, MSC, MEOX2, TBX3, TBX2	427	76	16792	3.62208801	1	0.25908363	0.25092261
GOTERM_BP_DIRECT	GO:0007229~integrin-mediated signaling pathway	8	1.50093809	0.01299497	COL3A1, SEMA7A, ADAMTS1, ITGA11, PLEK, ITGAX, ITGAL, TSPAN32	427	99	16792	3.17782036	1	0.2615238	0.25328591
GOTERM_BP_DIRECT	GO:0006469~negative regulation of protein kinase activity	8	1.50093809	0.01299497	SOCS2, LRRC15, SOCS3, IL6, FABP4, PODNL1, SOCS1, PODN	427	99	16792	3.17782036	1	0.2615238	0.25328591
GOTERM_BP_DIRECT	GO:0030154~cell differentiation	21	3.93996248	0.01459142	BLK, COL15A1, EGFL6, SH2D2A, FHL1, TWIST2, MATK, INHBA, CHRDL2, CHRDL1, FLI1, TYMP, SPIB, NR5A1, FCRLA, IGSF10, HLX, MMP19, CSK, ERG, SH2B3	427	462	16792	1.78752395	1	0.29105359	0.28188553
GOTERM_BP_DIRECT	GO:0032720~negative regulation of tumor necrosis factor production	5	0.9380863	0.01526415	ADIPOQ, ACP5, NLRC3, GAS6, FOXP3	427	38	16792	5.17441144	1	0.30180167	0.29229505
GOTERM_BP_DIRECT	GO:0001934~positive regulation of protein phosphorylation	9	1.68855535	0.01570053	SPN, BMP2, TGFB1, LRRK2, RARRES2, ADIPOQ, RSPO1, GAS6, CYR61	427	127	16792	2.78684837	1	0.30773033	0.29803696
GOTERM_BP_DIRECT	GO:0001763~morphogenesis of a branching structure	3	0.56285178	0.01624688	FOXF1, TCF21, PRDM1	427	8	16792	14.7470726	1	0.30773495	0.29804143
GOTERM_BP_DIRECT	GO:0034695~response to prostaglandin E	3	0.56285178	0.01624688	CCL21, CCR7, CCL19	427	8	16792	14.7470726	1	0.30773495	0.29804143
GOTERM_BP_DIRECT	GO:0010518~positive regulation of phospholipase activity	3	0.56285178	0.01624688	PLA2G5, ARHGAP6, CYR61	427	8	16792	14.7470726	1	0.30773495	0.29804143
GOTERM_BP_DIRECT	GO:0001955~blood vessel maturation	3	0.56285178	0.01624688	ACVRL1, CDH5, MMP2	427	8	16792	14.7470726	1	0.30773495	0.29804143
GOTERM_BP_DIRECT	GO:0001649~osteoblast differentiation	8	1.50093809	0.01667085	MRC2, SEMA7A, BMP2, ITGA11, TMEM119, GLI1, CYR61, GLI2	427	104	16792	3.02504053	1	0.31189918	0.30207449
GOTERM_BP_DIRECT	GO:0043406~positive regulation of MAP kinase activity	6	1.12570356	0.01674348	TGFB1, LRRK2, TNFSF11, CSK, FGF2, PIK3CG	427	59	16792	3.99920613	1	0.31189918	0.30207449
GOTERM_BP_DIRECT	GO:0043410~positive regulation of MAPK cascade	7	1.31332083	0.01690563	IL6, BMP2, OSM, IGF1, TNFRSF1B, RELT, TNFRSF4	427	81	16792	3.39850233	1	0.31233841	0.30249989
GOTERM_BP_DIRECT	GO:0071345~cellular response to cytokine stimulus	4	0.75046904	0.01745235	CSF3, FOXF1, DPYSL3, CCR7	427	22	16792	7.15009581	1	0.31981788	0.30974376
GOTERM_BP_DIRECT	GO:0038083~peptidyl-tyrosine autophosphorylation	5	0.9380863	0.01816849	BLK, ZAP70, MATK, CSK, JAK3	427	40	16792	4.91569087	1	0.33025623	0.3198533
GOTERM_BP_DIRECT	GO:0007507~heart development	11	2.06378987	0.01857069	ADAM19, COL3A1, BMP2, PDLIM3, TGFB1, WT1, FOXF1, XIRP1, SOD2, TBX5, GLI2	427	183	16792	2.36382949	1	0.33486673	0.32431857
GOTERM_BP_DIRECT	GO:0048754~branching morphogenesis of an epithelial tube	4	0.75046904	0.01970024	MMP14, SLIT2, RASIP1, GLI2	427	23	16792	6.83922208	1	0.35241532	0.34131439
GOTERM_BP_DIRECT	GO:0046641~positive regulation of alpha-beta T cell proliferation	3	0.56285178	0.02054163	ZAP70, CD28, CD3E	427	9	16792	13.108509	1	0.36172529	0.3503311
GOTERM_BP_DIRECT	GO:0060539~diaphragm development	3	0.56285178	0.02054163	WT1, TCF21, MSC	427	9	16792	13.108509	1	0.36172529	0.3503311
GOTERM_BP_DIRECT	GO:0097191~extrinsic apoptotic signaling pathway	5	0.9380863	0.02139403	IL33, TGFB1, G0S2, INHBA, TNFRSF1B	427	42	16792	4.68161035	1	0.36257252	0.35115165
GOTERM_BP_DIRECT	GO:0022008~neurogenesis	5	0.9380863	0.02139403	BCL11B, PRDM6, FAT4, BHLHE22, WNT2	427	42	16792	4.68161035	1	0.36257252	0.35115165
GOTERM_BP_DIRECT	GO:0045740~positive regulation of DNA replication	5	0.9380863	0.02139403	PDGFRA, IL6, IGF1, GLI1, GLI2	427	42	16792	4.68161035	1	0.36257252	0.35115165
GOTERM_BP_DIRECT	GO:0002548~monocyte chemotaxis	5	0.9380863	0.02139403	CCL14, IL6, CCL21, TNFSF11, CCL19	427	42	16792	4.68161035	1	0.36257252	0.35115165
GOTERM_BP_DIRECT	GO:0045599~negative regulation of fat cell differentiation	5	0.9380863	0.02139403	MMP11, IL6, TGFB1, ADIPOQ, SOD2	427	42	16792	4.68161035	1	0.36257252	0.35115165
GOTERM_BP_DIRECT	GO:0036152~phosphatidylethanolamine acyl-chain remodeling	4	0.75046904	0.0221001	PLA2G2D, PLA2G4E, PLA2G2A, PLA2G5	427	24	16792	6.55425449	1	0.36627666	0.35473911
GOTERM_BP_DIRECT	GO:0009954~proximal/distal pattern formation	4	0.75046904	0.0221001	GREM1, CHSY1, GLI1, GLI2	427	24	16792	6.55425449	1	0.36627666	0.35473911
GOTERM_BP_DIRECT	GO:0048870~cell motility	4	0.75046904	0.0221001	MMP14, PDPN, ETS1, IER2	427	24	16792	6.55425449	1	0.36627666	0.35473911
GOTERM_BP_DIRECT	GO:0030512~negative regulation of transforming growth factor beta receptor signaling pathway	6	1.12570356	0.02303193	TGFB1, SKOR2, HTRA3, ADAMTSL2, HTRA4, SMAD7	427	64	16792	3.68676815	1	0.37893405	0.3669978
GOTERM_BP_DIRECT	GO:0032735~positive regulation of interleukin-12 production	4	0.75046904	0.02465209	IRF8, CCR7, CCL19, AGER	427	25	16792	6.29208431	1	0.38983651	0.37755683
GOTERM_BP_DIRECT	GO:0071222~cellular response to lipopolysaccharide	8	1.50093809	0.02503925	CSF3, IL6, IL24, IRF8, LILRB1, LILRB2, CCR5, TNFRSF1B	427	113	16792	2.7841081	1	0.38983651	0.37755683
GOTERM_BP_DIRECT	GO:2000727~positive regulation of cardiac muscle cell differentiation	3	0.56285178	0.02525117	GREM1, MYOCD, TGFB1	427	10	16792	11.7976581	1	0.38983651	0.37755683
GOTERM_BP_DIRECT	GO:0001771~immunological synapse formation	3	0.56285178	0.02525117	CD6, CCL21, CCL19	427	10	16792	11.7976581	1	0.38983651	0.37755683
GOTERM_BP_DIRECT	GO:0002675~positive regulation of acute inflammatory response	3	0.56285178	0.02525117	IL6, OSM, PIK3CG	427	10	16792	11.7976581	1	0.38983651	0.37755683
GOTERM_BP_DIRECT	GO:0046888~negative regulation of hormone secretion	3	0.56285178	0.02525117	IL6, ADIPOQ, OSM	427	10	16792	11.7976581	1	0.38983651	0.37755683
GOTERM_BP_DIRECT	GO:0050765~negative regulation of phagocytosis	3	0.56285178	0.02525117	TGFB1, ADIPOQ, CSK	427	10	16792	11.7976581	1	0.38983651	0.37755683
GOTERM_BP_DIRECT	GO:0003197~endocardial cushion development	3	0.56285178	0.02525117	FOXF1, ERG, TBX5	427	10	16792	11.7976581	1	0.38983651	0.37755683
GOTERM_BP_DIRECT	GO:0033089~positive regulation of T cell differentiation in thymus	3	0.56285178	0.02525117	EGR3, TESPA1, GLI2	427	10	16792	11.7976581	1	0.38983651	0.37755683
GOTERM_BP_DIRECT	GO:0007517~muscle organ development	7	1.31332083	0.02562231	EGR3, SGCD, FHL1, ITGA11, COL6A3, AEBP1, IGF1	427	89	16792	3.09301897	1	0.39287539	0.38049999
GOTERM_BP_DIRECT	GO:0008360~regulation of cell shape	9	1.68855535	0.0264433	FMNL3, FMNL1, IL6, FGD5, WIPF1, CDC42EP5, PDPN, PLEKHO1, SYNE3	427	140	16792	2.52806959	1	0.40272436	0.39003873

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	old Enrichme	Bonferroni	Benjamini	FDR
GOTERM_BP_DIRECT	GO:0008015~blood circulation	5	0.9380863	0.02685165	ACVRL1, ELN, SERPING1, MEOX2, FLI1	427	45	16792	4.36950299	1	0.40349074	0.39078096
GOTERM_BP_DIRECT	GO:0030838~positive regulation of actin filament polymerization	5	0.9380863	0.02685165	CSF3, CCL21, CDC42EP5, LMOD1, CCR7	427	45	16792	4.36950299	1	0.40349074	0.39078096
GOTERM_BP_DIRECT	GO:0045766~positive regulation of angiogenesis	8	1.50093809	0.02722948	GREM1, ACVRL1, SFRP2, PTGIS, SPHK1, GATA6, ETS1, FGF2	427	115	16792	2.73568883	1	0.40645857	0.39365531
GOTERM_BP_DIRECT	GO:0003151~outflow tract morphogenesis	5	0.9380863	0.02883906	SFRP2, SOX18, GJA5, TBX3, TBX2	427	46	16792	4.2745138	1	0.42745175	0.41398721
GOTERM_BP_DIRECT	GO:0045892~negative regulation of transcription, DNA-templated	21	3.93996248	0.03016263	TGFB1, RAMP3, FOXF2, PRDM6, ADIPOQ, GATA6, TWIST2, ARID5A, IKZF1, FOXP3, TBX3, TBX2, GREM1, BMP2, DACH1, FABP4, SFRP2, WT1, SOX18, GAS6, TNFRSF4	427	499	16792	1.6549821	1	0.42745175	0.41398721
GOTERM_BP_DIRECT	GO:0055007~cardiac muscle cell differentiation	4	0.75046904	0.03021136	MYOCD, BMP2, GATA6, TBX5	427	27	16792	5.82600399	1	0.42745175	0.41398721
GOTERM_BP_DIRECT	GO:0036151~phosphatidylcholine acyl-chain remodeling	4	0.75046904	0.03021136	PLA2G2D, PLA2G4E, PLA2G2A, PLA2G5	427	27	16792	5.82600399	1	0.42745175	0.41398721
GOTERM_BP_DIRECT	GO:0035987~endodermal cell differentiation	4	0.75046904	0.03021136	MMP14, MMP2, COL12A1, INHBA	427	27	16792	5.82600399	1	0.42745175	0.41398721
GOTERM_BP_DIRECT	GO:0003181~atrioventricular valve morphogenesis	3	0.56285178	0.03035178	BMP2, TBX5, CYR61	427	11	16792	10.7251437	1	0.42745175	0.41398721
GOTERM_BP_DIRECT	GO:0001946~lymphangiogenesis	3	0.56285178	0.03035178	ACVRL1, SOX18, PDPN	427	11	16792	10.7251437	1	0.42745175	0.41398721
GOTERM_BP_DIRECT	GO:0001938~positive regulation of endothelial cell proliferation	6	1.12570356	0.03067567	ACVRL1, BMP2, EGFL7, EGR3, WNT2, FGF2	427	69	16792	3.41961104	1	0.42745175	0.41398721
GOTERM_BP_DIRECT	GO:0006508~proteolysis	21	3.93996248	0.03073563	FCN1, CPA3, FCN3, TPSB2, CPXM2, HTRA3, MMP2, HTRA4, KLK4, ADAMTS4, MMP11, MMP14, AGBL1, ADAMTS1, MMP23B, ADAMTSL2, MMP19, TPSAB1, ADAMTS8, PAPPA2, LTF	427	500	16792	1.65167213	1	0.42745175	0.41398721
GOTERM_BP_DIRECT	GO:0071230~cellular response to amino acid stimulus	5	0.9380863	0.03091155	COL3A1, SOCS1, COL1A2, MMP2, COL5A2	427	47	16792	4.18356669	1	0.42745175	0.41398721
GOTERM_BP_DIRECT	GO:0045860~positive regulation of protein kinase activity	5	0.9380863	0.03091155	CCL21, CCR7, CCL19, GAS6, CYR61	427	47	16792	4.18356669	1	0.42745175	0.41398721
GOTERM_BP_DIRECT	GO:0051781~positive regulation of cell division	5	0.9380863	0.03091155	FGF7, TGFB1, TAL1, OSM, FGF2	427	47	16792	4.18356669	1	0.42745175	0.41398721
GOTERM_BP_DIRECT	GO:0033138~positive regulation of peptidyl-serine phosphorylation	6	1.12570356	0.03237375	CSF3, IL6, TGFB1, SFRP2, OSM, GAS6	427	70	16792	3.37075945	1	0.4449416	0.43092613
GOTERM_BP_DIRECT	GO:0032715~negative regulation of interleukin-6 production	4	0.75046904	0.03321732	CSK, NLRC3, GAS6, FOXP3	427	28	16792	5.61793242	1	0.45103521	0.4368278
GOTERM_BP_DIRECT	GO:0003281~ventricular septum development	4	0.75046904	0.03321732	GJA5, PRDM1, TBX5, CYR61	427	28	16792	5.61793242	1	0.45103521	0.4368278
GOTERM_BP_DIRECT	GO:0071347~cellular response to interleukin-1	6	1.12570356	0.03412939	CCL14, IL6, PTGIS, CCL21, CCL19, ADAMTS12	427	71	16792	3.32328397	1	0.45874898	0.44429859
GOTERM_BP_DIRECT	GO:0050852~T cell receptor signaling pathway	9	1.68855535	0.03514504	ZAP70, THEMIS2, GRAP2, CD28, WAS, CSK, CD3G, CD3E, FOXP3	427	148	16792	2.39141718	1	0.45874898	0.44429859
GOTERM_BP_DIRECT	GO:0071560~cellular response to transforming growth factor beta stimulus	5	0.9380863	0.03531342	ACVRL1, CLEC3B, TGFB1, WNT2, SMAD7	427	49	16792	4.01280887	1	0.45874898	0.44429859
GOTERM_BP_DIRECT	GO:0050710~negative regulation of cytokine secretion	3	0.56285178	0.03582068	IL6, RGCC, FOXP3	427	12	16792	9.83138173	1	0.45874898	0.44429859
GOTERM_BP_DIRECT	GO:0035385~Roundabout signaling pathway	3	0.56285178	0.03582068	ROBO2, SLIT3, SLIT2	427	12	16792	9.83138173	1	0.45874898	0.44429859
GOTERM_BP_DIRECT	GO:0045086~positive regulation of interleukin-2 biosynthetic process	3	0.56285178	0.03582068	IRF4, CD28, CD3E	427	12	16792	9.83138173	1	0.45874898	0.44429859
GOTERM_BP_DIRECT	GO:0043206~extracellular fibril organization	3	0.56285178	0.03582068	MFAP4, COL3A1, COL5A1	427	12	16792	9.83138173	1	0.45874898	0.44429859
GOTERM_BP_DIRECT	GO:2000279~negative regulation of DNA biosynthetic process	3	0.56285178	0.03582068	ACVRL1, DACH1, ADIPOQ	427	12	16792	9.83138173	1	0.45874898	0.44429859
GOTERM_BP_DIRECT	GO:0032695~negative regulation of interleukin-12 production	3	0.56285178	0.03582068	ACP5, TIGIT, JAK3	427	12	16792	9.83138173	1	0.45874898	0.44429859
GOTERM_BP_DIRECT	GO:0045779~negative regulation of bone resorption	3	0.56285178	0.03582068	IL6, CLDN18, CSK	427	12	16792	9.83138173	1	0.45874898	0.44429859
GOTERM_BP_DIRECT	GO:0042113~B cell activation	4	0.75046904	0.03637283	CD79A, ZAP70, LAX1, IKZF3	427	29	16792	5.42421061	1	0.46318851	0.44859828
GOTERM_BP_DIRECT	GO:0071773~cellular response to BMP stimulus	4	0.75046904	0.03967656	ACVRL1, BMP2, GATA6, ADAMTS12	427	30	16792	5.24340359	1	0.50242117	0.48659512
GOTERM_BP_DIRECT	GO:0019722~calcium-mediated signaling	5	0.9380863	0.04005957	LRRK2, SPHK1, TNFSF11, CCR5, SELE	427	51	16792	3.85544382	1	0.50443727	0.48854772
GOTERM_BP_DIRECT	GO:0031274~positive regulation of pseudopodium assembly	3	0.56285178	0.04163601	CCL21, CDC42EP5, CCR7	427	13	16792	9.0751216	1	0.51004115	0.49397508
GOTERM_BP_DIRECT	GO:0051145~smooth muscle cell differentiation	3	0.56285178	0.04163601	MYOCD, FOXF1, GATA6	427	13	16792	9.0751216	1	0.51004115	0.49397508
GOTERM_BP_DIRECT	GO:0033630~positive regulation of cell adhesion mediated by integrin	3	0.56285178	0.04163601	ZAP70, SFRP2, CCL21	427	13	16792	9.0751216	1	0.51004115	0.49397508
GOTERM_BP_DIRECT	GO:0060389~pathway-restricted SMAD protein phosphorylation	3	0.56285178	0.04163601	BMP2, TGFB1, SMAD7	427	13	16792	9.0751216	1	0.51004115	0.49397508
GOTERM_BP_DIRECT	GO:0060413~atrial septum morphogenesis	3	0.56285178	0.04163601	GJA5, TBX5, CYR61	427	13	16792	9.0751216	1	0.51004115	0.49397508
GOTERM_BP_DIRECT	GO:0030513~positive regulation of BMP signaling pathway	4	0.75046904	0.04312694	ACVRL1, GATA6, SULF1, CYR61	427	31	16792	5.07426154	1	0.52544933	0.50889791
GOTERM_BP_DIRECT	GO:0016337~single organismal cell-cell adhesion	7	1.31332083	0.04358483	CD2, CDH5, FOXF1, COL14A1, PDPN, ICOS, LIMS2	427	101	16792	2.72553157	1	0.5281732	0.51153598
GOTERM_BP_DIRECT	GO:0007498~mesoderm development	4	0.75046904	0.0467222	TIE1, MATK, IKZF1, IKZF3	427	32	16792	4.91569087	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:0050873~brown fat cell differentiation	4	0.75046904	0.0467222	FABP4, RARRES2, ADIPOQ, METRNL	427	32	16792	4.91569087	1	0.55351524	0.53607976

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	old Enrichme	Bonferroni	Benjamini	FDR
GOTERM_BP_DIRECT	GO:0030889~negative regulation of B cell proliferation	3	0.56285178	0.04777675	TNFRSF13B, CTLA4, PRDM1	427	14	16792	8.42689863	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:0031529~ruffle organization	3	0.56285178	0.04777675	CCL21, PLEK, CCR7	427	14	16792	8.42689863	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:0070208~protein heterotrimerization	3	0.56285178	0.04777675	COL1A2, COL6A2, ADIPOQ	427	14	16792	8.42689863	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:0045579~positive regulation of B cell differentiation	3	0.56285178	0.04777675	ZAP70, MMP14, PRDM1	427	14	16792	8.42689863	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:2000147~positive regulation of cell motility	3	0.56285178	0.04777675	CCL21, CCR7, CCL19	427	14	16792	8.42689863	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:0006644~phospholipid metabolic process	5	0.9380863	0.04782574	PLA2G2D, PLA2G4E, PLA2G2A, GATA6, PLA2G5	427	54	16792	3.64125249	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:0001701~in utero embryonic development	10	1.87617261	0.04936226	ACVRL1, BMP2, FOXF1, SOX18, SCO2, TIE1, RARRES2, GATA6, TBX3, GLI2	427	187	16792	2.10296935	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:0060032~notochord regression	2	0.37523452	0.05009632	GLI1, GLI2	427	2	16792	39.3255269	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:0045066~regulatory T cell differentiation	2	0.37523452	0.05009632	TGFB1, CD28	427	2	16792	39.3255269	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:0035051~cardiocyte differentiation	2	0.37523452	0.05009632	MYOCD, BMP2	427	2	16792	39.3255269	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:0001869~negative regulation of complement activation, lectin pathway	2	0.37523452	0.05009632	SERPING1, A2M	427	2	16792	39.3255269	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:0021965~spinal cord ventral commissure morphogenesis	2	0.37523452	0.05009632	DCC, GLI2	427	2	16792	39.3255269	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:0072610~interleukin-12 secretion	2	0.37523452	0.05009632	CCR7, CCL19	427	2	16792	39.3255269	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:0002669~positive regulation of T cell anergy	2	0.37523452	0.05009632	CD3E, FOXP3	427	2	16792	39.3255269	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:2001183~negative regulation of interleukin-12 secretion	2	0.37523452	0.05009632	TNFRSF9, LILRB1	427	2	16792	39.3255269	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:1903225~negative regulation of endodermal cell differentiation	2	0.37523452	0.05009632	COL5A1, COL5A2	427	2	16792	39.3255269	1	0.55351524	0.53607976
GOTERM_BP_DIRECT	GO:0010718~positive regulation of epithelial to mesenchymal transition	4	0.75046904	0.05046035	BMP2, TGFB1, RGCC, GLIPR2	427	33	16792	4.76673054	1	0.55481767	0.53734116
GOTERM_BP_DIRECT	GO:0042060~wound healing	6	1.12570356	0.0525733	PDGFRA, COL3A1, IL6, IL24, PECAM1, FGF2	427	80	16792	2.94941452	1	0.57524373	0.5571238
GOTERM_BP_DIRECT	GO:0048015~phosphatidylinositol-mediated signaling	7	1.31332083	0.0529274	PDGFRA, FGF7, CD19, CD28, IGF1, FGF2, PIK3CG	427	106	16792	2.59696876	1	0.57632062	0.55816678
GOTERM_BP_DIRECT	GO:0035050~embryonic heart tube development	3	0.56285178	0.05422274	SOX18, GJA5, TBX2	427	15	16792	7.86510539	1	0.58047688	0.56219212
GOTERM_BP_DIRECT	GO:0002063~chondrocyte development	3	0.56285178	0.05422274	SFRP2, CHSY1, SULF1	427	15	16792	7.86510539	1	0.58047688	0.56219212
GOTERM_BP_DIRECT	GO:0032270~positive regulation of cellular protein metabolic process	3	0.56285178	0.05422274	TGFB1, ADIPOQ, INHBA	427	15	16792	7.86510539	1	0.58047688	0.56219212
GOTERM_BP_DIRECT	GO:0043588~skin development	4	0.75046904	0.05433923	COL3A1, COL5A1, COL5A2, LTB	427	34	16792	4.62653258	1	0.58047688	0.56219212
GOTERM_BP_DIRECT	GO:0007204~positive regulation of cytosolic calcium ion concentration	8	1.50093809	0.05469394	PDGFRA, PTGIR, XCR1, PTGER3, CCR7, CCR5, PIK3CG, PTGDR	427	134	16792	2.34779265	1	0.58128904	0.56297869
GOTERM_BP_DIRECT	GO:0006936~muscle contraction	7	1.31332083	0.05493104	DES, MYH8, LMOD1, MYH13, MYH11, MYL9, MYLK	427	107	16792	2.57269802	1	0.58128904	0.56297869
GOTERM_BP_DIRECT	GO:0086091~regulation of heart rate by cardiac conduction	4	0.75046904	0.05835651	KCNJ5, KCNE1, GJA5, KCNJ2	427	35	16792	4.49434594	1	0.61465224	0.59529097
GOTERM_BP_DIRECT	GO:0048566~embryonic digestive tract development	3	0.56285178	0.06095461	FOXF2, RARRES2, GLI2	427	16	16792	7.3735363	1	0.63023711	0.61038492
GOTERM_BP_DIRECT	GO:0051895~negative regulation of focal adhesion assembly	3	0.56285178	0.06095461	ACVRL1, MMP14, ARHGAP6	427	16	16792	7.3735363	1	0.63023711	0.61038492
GOTERM_BP_DIRECT	GO:0035855~megakaryocyte development	3	0.56285178	0.06095461	MEIS1, TAL1, FLI1	427	16	16792	7.3735363	1	0.63023711	0.61038492
GOTERM_BP_DIRECT	GO:0003203~endocardial cushion morphogenesis	3	0.56285178	0.06095461	ACVRL1, BMP2, TBX2	427	16	16792	7.3735363	1	0.63023711	0.61038492
GOTERM_BP_DIRECT	GO:0009612~response to mechanical stimulus	5	0.9380863	0.06248245	COL3A1, MMP14, JUND, ETS1, GLI2	427	59	16792	3.33267177	1	0.6375425	0.6174602
GOTERM_BP_DIRECT	GO:0030177~positive regulation of Wnt signaling pathway	4	0.75046904	0.06250971	BMP2, RSPO1, SULF1, DACT1	427	36	16792	4.36950299	1	0.6375425	0.6174602
GOTERM_BP_DIRECT	GO:0048469~cell maturation	4	0.75046904	0.06250971	CCL21, SOX18, CCL19, POU2F2	427	36	16792	4.36950299	1	0.6375425	0.6174602
GOTERM_BP_DIRECT	GO:0050830~defense response to Gram-positive bacterium	6	1.12570356	0.06490254	IL6, PLA2G2A, ACP5, SSC5D, LYZ, HLA-E	427	85	16792	2.77591955	1	0.65896539	0.63820828
GOTERM_BP_DIRECT	GO:0050679~positive regulation of epithelial cell proliferation	5	0.9380863	0.06566713	IL6, FGF7, TGFB1, NR4A3, IGF1	427	60	16792	3.27712724	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0007162~negative regulation of cell adhesion	4	0.75046904	0.06679621	SPN, ACVRL1, PDE3B, RND1	427	37	16792	4.25140832	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0001822~kidney development	6	1.12570356	0.0675468	ROBO2, WT1, ADAMTS1, TCF21, SULF1, GLI2	427	86	16792	2.74364141	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0071310~cellular response to organic substance	3	0.56285178	0.06795379	CLEC3B, EGR2, TIMP3	427	17	16792	6.93979887	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0032331~negative regulation of chondrocyte differentiation	3	0.56285178	0.06795379	GREM1, ADAMTS12, GLI2	427	17	16792	6.93979887	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0042517~positive regulation of tyrosine phosphorylation of Stat3 protein	4	0.75046904	0.07121324	SOCS3, IL6, IL24, OSM	427	38	16792	4.13952915	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0016525~negative regulation of angiogenesis	5	0.9380863	0.07228494	RGCC, NPR1, TIE1, PDE3B, SULF1	427	62	16792	3.17141346	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0030168~platelet activation	7	1.31332083	0.07259444	COL3A1, IL6, COL1A2, P2RX1, PDPN, GAS6, PIK3CG	427	115	16792	2.39372773	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0071731~response to nitric oxide	2	0.37523452	0.07419751	CCR7, CCL19	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0002774~Fc receptor mediated inhibitory signaling pathway	2	0.37523452	0.07419751	LILRB1, LILRB2	427	3	16792	26.217018	1	0.66103237	0.64021015

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	old Enrichme	Bonferroni	Benjamini	FDR
GOTERM_BP_DIRECT	GO:0060128~corticotropin hormone secreting cell differentiation	2	0.37523452	0.07419751	BMP2, FGF2	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0007418~ventral midline development	2	0.37523452	0.07419751	GLI1, GLI2	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0021834~chemorepulsion involved in embryonic olfactory bulb interneuron precursor migration	2	0.37523452	0.07419751	SLIT3, SLIT2	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0002408~myeloid dendritic cell chemotaxis	2	0.37523452	0.07419751	CCR7, CCL19	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:2000669~negative regulation of dendritic cell apoptotic process	2	0.37523452	0.07419751	LILRB1, GAS6	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0021773~striatal medium spiny neuron differentiation	2	0.37523452	0.07419751	BCL11B, INHBA	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0002513~tolerance induction to self antigen	2	0.37523452	0.07419751	TGFB1, FOXP3	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:2000195~negative regulation of female gonad development	2	0.37523452	0.07419751	NR5A1, WT1	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0071847~TNFSF11-mediated signaling pathway	2	0.37523452	0.07419751	CLDN18, TNFSF11	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0060414~aorta smooth muscle tissue morphogenesis	2	0.37523452	0.07419751	COL3A1, MYLK	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0014707~branchiomeric skeletal muscle development	2	0.37523452	0.07419751	TCF21, MSC	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0072277~metanephric glomerular capillary formation	2	0.37523452	0.07419751	PDGFRA, TCF21	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0048592~eye morphogenesis	2	0.37523452	0.07419751	COL5A1, COL5A2	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0042512~negative regulation of tyrosine phosphorylation of Stat1 protein	2	0.37523452	0.07419751	SOCS3, SOCS1	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0010572~positive regulation of platelet activation	2	0.37523452	0.07419751	SELP, PLEK	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0060923~cardiac muscle cell fate commitment	2	0.37523452	0.07419751	WT1, TBX3	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0034021~response to silicon dioxide	2	0.37523452	0.07419751	SCGB1A1, SOD2	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0002767~immune response-inhibiting cell surface receptor signaling pathway	2	0.37523452	0.07419751	LILRB1, LILRB2	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:2001180~negative regulation of interleukin-10 secretion	2	0.37523452	0.07419751	TNFRSF9, LILRB1	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0002361~CD4-positive, CD25-positive, alpha-beta regulatory T cell differentiation	2	0.37523452	0.07419751	PLA2G2D, FOXP3	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0060596~mammary placode formation	2	0.37523452	0.07419751	TBX3, TBX2	427	3	16792	26.217018	1	0.66103237	0.64021015
GOTERM_BP_DIRECT	GO:0001782~B cell homeostasis	3	0.56285178	0.07520244	TNFRSF13B, TNFRSF13C, FOXP3	427	18	16792	6.55425449	1	0.66185432	0.6410062
GOTERM_BP_DIRECT	GO:0036148~phosphatidylglycerol acyl-chain remodeling	3	0.56285178	0.07520244	PLA2G2D, PLA2G2A, PLA2G5	427	18	16792	6.55425449	1	0.66185432	0.6410062
GOTERM_BP_DIRECT	GO:0045954~positive regulation of natural killer cell mediated cytotoxicity	3	0.56285178	0.07520244	NCR3, LAG3, SH2D1A	427	18	16792	6.55425449	1	0.66185432	0.6410062
GOTERM_BP_DIRECT	GO:0050727~regulation of inflammatory response	5	0.9380863	0.07571657	SEMA7A, FABP4, SCGB1A1, ADAMTS12, SELE	427	63	16792	3.12107357	1	0.66185432	0.6410062
GOTERM_BP_DIRECT	GO:0002062~chondrocyte differentiation	4	0.75046904	0.07575795	BMP2, TGFB1, RUNX3, GLI2	427	39	16792	4.03338738	1	0.66185432	0.6410062
GOTERM_BP_DIRECT	GO:0032436~positive regulation of proteasomal ubiquitin-dependent protein catabolic process	5	0.9380863	0.07922887	IL33, LRRK2, RNF19B, RNF166, SMAD7	427	64	16792	3.07230679	1	0.6868534	0.66521782
GOTERM_BP_DIRECT	GO:0060337~type I interferon signaling pathway	5	0.9380863	0.07922887	IFNAR2, IRF4, IRF8, SAMHD1, HLA-E	427	64	16792	3.07230679	1	0.6868534	0.66521782
GOTERM_BP_DIRECT	GO:0030326~embryonic limb morphogenesis	4	0.75046904	0.08042736	GREM1, GJA5, TBX5, TBX4	427	40	16792	3.93255269	1	0.69457194	0.67269323
GOTERM_BP_DIRECT	GO:0008589~regulation of smoothened signaling pathway	3	0.56285178	0.08268349	GAS1, GLI1, GLI2	427	19	16792	6.20929373	1	0.70444693	0.68225717
GOTERM_BP_DIRECT	GO:0008217~regulation of blood pressure	5	0.9380863	0.08282096	ACVRL1, COL1A2, NPR1, P2RX1, SOD2	427	65	16792	3.02504053	1	0.70444693	0.68225717
GOTERM_BP_DIRECT	GO:0060326~cell chemotaxis	5	0.9380863	0.08282096	CCL14, PDGFRA, CCL21, CCL19, CXCL14	427	65	16792	3.02504053	1	0.70444693	0.68225717
GOTERM_BP_DIRECT	GO:0014068~positive regulation of phosphatidylinositol 3-kinase signaling	5	0.9380863	0.08282096	SELP, PDGFRA, CSF3, CD28, IGF1	427	65	16792	3.02504053	1	0.70444693	0.68225717
GOTERM_BP_DIRECT	GO:0008283~cell proliferation	15	2.81425891	0.08412997	MAP4K1, IFNAR2, IL15RA, ELN, SH2D2A, OSM, MATK, IGF1, CYR61, GLI2, DACH1, PDPN, PIM2, ERG, GAS6	427	366	16792	1.61170192	1	0.71289083	0.69043508
GOTERM_BP_DIRECT	GO:0007179~transforming growth factor beta receptor signaling pathway	6	1.12570356	0.08464419	ACVRL1, CDH5, COL3A1, TGFB1, COL1A2, SMAD7	427	92	16792	2.56470828	1	0.71406071	0.69156812
GOTERM_BP_DIRECT	GO:0051209~release of sequestered calcium ion into cytosol	4	0.75046904	0.08521843	CCL21, CCR7, CCL19, FGF2	427	41	16792	3.83663677	1	0.71406071	0.69156812
GOTERM_BP_DIRECT	GO:0048839~inner ear development	4	0.75046904	0.08521843	BMP2, TGFB1, MAF, CXCL14	427	41	16792	3.83663677	1	0.71406071	0.69156812
GOTERM_BP_DIRECT	GO:0007566~embryo implantation	4	0.75046904	0.09012799	PTGIS, MMP2, SCGB1A1, IGFBP7	427	42	16792	3.74528828	1	0.74698507	0.72345538
GOTERM_BP_DIRECT	GO:0030101~natural killer cell activation	3	0.56285178	0.09038053	CD2, IL21R, SLAMF7	427	20	16792	5.89882904	1	0.74698507	0.72345538

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	old Enrichme	Bonferroni	Benjamini	FDR
GOTERM_BP_DIRECT	GO:0006801~superoxide metabolic process	3	0.56285178	0.09038053	NRROS, NCF1, SOD2	427	20	16792	5.89882904	1	0.74698507	0.72345538
GOTERM_BP_DIRECT	GO:0008584~male gonad development	6	1.12570356	0.09080475	NR5A1, MMP14, SFRP2, WT1, GATA6, INHBA	427	94	16792	2.51014002	1	0.74698507	0.72345538
GOTERM_BP_DIRECT	GO:0046854~phosphatidylinositol phosphorylation	6	1.12570356	0.09080475	PDGFRA, FGF7, CD19, CD28, FGF2, PIK3CG	427	94	16792	2.51014002	1	0.74698507	0.72345538
GOTERM_BP_DIRECT	GO:0018108~peptidyl-tyrosine phosphorylation	8	1.50093809	0.09506801	PDGFRA, ZAP70, FGF7, TIE1, CSF2RB, IL12RB1, JAK3, FGF2	427	153	16792	2.0562367	1	0.76916072	0.7449325
GOTERM_BP_DIRECT	GO:0003257~positive regulation of transcription from RNA polymerase II promoter involved in myocardial precursor cell differentiation	2	0.37523452	0.0976886	GREM1, MYOCD	427	4	16792	19.6627635	1	0.76916072	0.7449325
GOTERM_BP_DIRECT	GO:0002826~negative regulation of T-helper 1 type immune response	2	0.37523452	0.0976886	IL33, IL1RL1	427	4	16792	19.6627635	1	0.76916072	0.7449325
GOTERM_BP_DIRECT	GO:0060129~thyroid-stimulating hormone-secreting cell differentiation	2	0.37523452	0.0976886	BMP2, FGF2	427	4	16792	19.6627635	1	0.76916072	0.7449325
GOTERM_BP_DIRECT	GO:0042482~positive regulation of odontogenesis	2	0.37523452	0.0976886	BMP2, TGFB1	427	4	16792	19.6627635	1	0.76916072	0.7449325
GOTERM_BP_DIRECT	GO:0021938~smoothened signaling pathway involved in regulation of cerebellar granule cell precursor cell proliferation	2	0.37523452	0.0976886	GLI1, GLI2	427	4	16792	19.6627635	1	0.76916072	0.7449325
GOTERM_BP_DIRECT	GO:0061056~sclerotome development	2	0.37523452	0.0976886	SFRP2, MEOX1	427	4	16792	19.6627635	1	0.76916072	0.7449325
GOTERM_BP_DIRECT	GO:1901203~positive regulation of extracellular matrix assembly	2	0.37523452	0.0976886	TGFB1, RGCC	427	4	16792	19.6627635	1	0.76916072	0.7449325
GOTERM_BP_DIRECT	GO:0060841~venous blood vessel development	2	0.37523452	0.0976886	ACVRL1, FOXF1	427	4	16792	19.6627635	1	0.76916072	0.7449325
GOTERM_BP_DIRECT	GO:1903142~positive regulation of establishment of endothelial barrier	2	0.37523452	0.0976886	CDH5, S1PR2	427	4	16792	19.6627635	1	0.76916072	0.7449325
GOTERM_BP_DIRECT	GO:0060244~negative regulation of cell proliferation involved in contact inhibition	2	0.37523452	0.0976886	SRPX, DACH1	427	4	16792	19.6627635	1	0.76916072	0.7449325
GOTERM_BP_DIRECT	GO:2001206~positive regulation of osteoclast development	2	0.37523452	0.0976886	GPR68, TNFSF11	427	4	16792	19.6627635	1	0.76916072	0.7449325
GOTERM_BP_DIRECT	GO:0048844~artery morphogenesis	3	0.56285178	0.09827786	GJA5, PRDM1, SMAD7	427	21	16792	5.61793242	1	0.76916072	0.7449325
GOTERM_BP_DIRECT	GO:0042346~positive regulation of NF-kappaB import into nucleus	3	0.56285178	0.09827786	GREM1, SPHK1, CCL19	427	21	16792	5.61793242	1	0.76916072	0.7449325

B.4 Single sample gene set enrichment analysis

B.4.1 ssGSEA gene sets

Lung (GTEx; 48 unique genes)

A2M	EPAS1	HLA-DRA	IGKJ4	RPL10A	S100A9	SLPI
AGER	FCN3	HLA-DRB1	LGALS1	RPL11	SAT1	TAGLN2
AQP1	FN1	HLA-E	MFAP4	RPL27	SERPING1	TGM2
BGN	GPX3	IFI30	MGP	RPLP0	SFTPA1	TIMP3
CD74	GSTP1	IFITM2	MYL9	RPS8	SFTPA2	TXNIP
CTSD	HLA-A	IFITM3	NAPSA	S100A11	SFTPB	ZFP36
DUSP1	HLA-C	IGHA1	RGCC	S100A6	SFTPC	

Brain (GTEx; 169 unique genes)

- Top 100 unique genes, mitochondrial excluded

ACAP3	CADM3	DNM1	ITM2C	PAQR6	RTN1	TLE2
AES	CALM1	EEF1A2	KIF5A	PCP4	RTN3	TMA7
ALDH2	CALM2	EIF4A2	LDHB	PCSK1N	RUNDC3A	TMEM178A
ALDOA	CALM3	ENO2	LENG8	PDIA2	S100B	TMEM59L
ALDOC	CALY	FAM107A	LINC00599	PEA15	SCD	TMEM66
APLP1	CAMK2A	FAM127A	MARCKSL1	PEBP1	SCG2	TP53INP2
APLP2	CAMK2N1	FBXL16	MBP	PFDN5	SEPW1	TPI1
APOD	CBLN1	FKBP1A	MOBP	PHYHIP	SEPT4	TSPAN7
APOE	CBLN3	FKBP8	MRFAP1	PKD1	SERINC1	TSPYL2
APP	CDR1	GABRD	MT3	PKM	SLC17A7	TUBA1A
ATN1	CEND1	GAP43	MTND2P28	PLEKHB1	SLC1A2	TUBA1B
ATP1A1	CHCHD2	GDI1	NCDN	PLP1	SLC22A17	TUBB4A
ATP1A2	CHGB	GFAP	NCS1	PPP1R1B	SNAP25	UBB
ATP1B1	CHN1	GLUL	NDRG2	PRDX1	SNCB	UCHL1
ATP1B2	CKB	GNAS	NDRG4	PRDX5	SNCG	VAMP2
ATP5A1	CLU	GNG7	NDUFA1	PRNP	SNRNP70	VSNL1
ATP5B	COX6A1	H3F3B	NDUFS5	PRRT2	SPP1	YWHAE
ATP6V0C	COX8A	HNRNPA2B	NEFL	PTGDS	SST	YWHAG
ATP6V1G2	CPE	1	NEFM	PTMS	STMN2	YWHAH
AVP	CPLX2	HPCA	NGFRAP1	PVALB	STMN3	ZBTB18
BASP1	CRYAB	HSP90AA1	NGRN	QDPR	STXBP1	
BEX1	CST3	HSPA8	NNAT	RAB3A	SYNDIG1L	
BSG	CTXN1	HTRA1	NPDC1	RHOB	SYP	
C7orf41	DDX17	IDS	NRGN	RPL34	TAGLN3	
CA11	DKK3	ITM2B	NSMF	RPL38	TF	

Allograft rejection (GTEx;"HALLMARK_ALLOGRAFT_REJECTION")

AARS1	BCL10	CCL2	CCR5	CD40	CD8B	CXCL9
ABCE1	BCL3	CCL22	CD1D	CD40LG	CD96	CXCR3
ABI1	BRCA1	CCL4	CD2	CD47	CDKN2A	DARS1
ACHE	C2	CCL5	CD247	CD7	CFP	DEGS1
ACVR2A	CAPG	CCL7	CD28	CD74	CRTAM	DYRK3
AKT1	CARTPT	CCND2	CD3D	CD79A	CSF1	EGFR
APBB1	CCL11	CCND3	CD3E	CD80	CSK	EIF3A
B2M	CCL13	CCR1	CD3G	CD86	CTSS	EIF3D
BCAT1	CCL19	CCR2	CD4	CD8A	CXCL13	EIF3J

EIF4G3	GZMB	IKBKB	IL7	MAP3K7	RARS1	TGFB1
EIF5A	HCLS1	IL10	IL9	MAP4K1	RIPK2	TGFB2
ELANE	HDAC9	IL11	INHBA	MBL2	RPL39	THY1
ELF4	HIF1A	IL12A	INHBB	MMP9	RPL3L	TIMP1
EREG	HLA-A	IL12B	IRF4	MRPL3	RPL9	TLR1
ETS1	HLA-DMA	IL12RB1	IRF7	MTIF2	RPS19	TLR2
F2	HLA-DMB	IL13	IRF8	NCF4	RPS3A	TLR3
F2R	HLA-DOA	IL15	ITGAL	NCK1	RPS9	TLR6
FAS	HLA-DOB	IL16	ITGB2	NCR1	SIT1	TNF
FASLG	HLA-DQA1	IL18	ITK	NLRP3	SOCS1	TPD52
FCGR2B	HLA-DRA	IL18RAP	JAK2	NME1	SOCS5	TRAF2
FGR	HLA-E	IL1B	KLRD1	NOS2	SPI1	TRAT1
FLNA	HLA-G	IL2	KRT1	NPM1	SRGN	UBE2D1
FYB1	ICAM1	IL27RA	LCK	PF4	ST8SIA4	UBE2N
GALNT1	ICOSLG	IL2RA	LCP2	PRF1	STAB1	WARS1
GBP2	IFNAR2	IL2RB	LIF	PRKCB	STAT1	WAS
GCNT1	IFNG	IL2RG	LTB	PRKCG	STAT4	ZAP70
GLMN	IFNGR1	IL4	LY75	PSMB10	TAP1	
GPR65	IFNGR2	IL4R	LY86	PTPN6	TAP2	
GZMA	IGSF6	IL6	LYN	PTPRC	TAPBP	

Astrocyte (PMID:29892006)

APC	CLU	ETNPPL	GLUD1	MACF1	SLC1A2
AQP4	CPE	FAM171B	GPM6A	NTRK2	SLC1A3
ATP1B2	DDAH1	GJA1	GPR98	PTPRZ1	SYNE1

Oligodendrocyte (PMID:29892006)

CLDND1	ERMN	QKI	SLC44A1	TMEM144	
CNP	MBP	SCD	SOX2-OT	TUBA1A	
CRYAB	PLP1	SEPT7	TF	ZEB2	

SOX2 gene signature(PMID:29374236)

ABCC5	CASP4	EIF4A2	KCNMB2	NCRNA0008	PRKCI	TBL1XR1
ACAP2	CEP70	EPCAM	KCNMB3	6	RAB6B	TMEM20
ACPL2	CHST7	FAM20B	KLHL24	NMNAT3	RASL11A	TRA2B
ACTL6A	CLCN2	FEZF1	LOC1001306	NTRK2	RIMKLA	TRIM59
ADRA2B	COCH	GK5	91	NTS	RSRC1	TSEN15
B4GALT4	CYP26A1	GLT25D2	LSG1	OPA1	SENP2	ULBP1
BEX2	DCUN1D1	GPC3	MCCC1	OR7A5	SGEF	VWA5B2
C3orf21	DLG1	GPR160	MED12L	PCCB	SLC25A36	WNK2
C3orf58	DLX6AS	IFT80	MRAP2	PHC3	SOX2OT	ZNF639
C6orf168	DMRT 3	IQCB1	MSL2	PIR	STRBP	ZXDC
CACNA1B	DVL3	KCNG3	MYNN	PNCK	TBCCD1	

B.4.2 ssGSEA scores and statistics

ssGSEA scores

Lung GTEx ssGSEA score		Brain GTEx ssGSEA score		Allograft rejection ssGSEA scores		Astrocyte ssGSEA scores		Oligodentrocyte ssGSEA scores		SOX2 ssGSEA scores	
Tumors	Mets	Tumors	Mets	Tumors	Mets	Tumors	Mets	Tumors	Mets	Tumors	Mets
12961.7389	12678.3365	9228.72008	9525.8257	7238.47513	7776.69088	4795.29956	3204.32608	5909.27913	6599.03874	4084.89126	3715.01411
12804.5422	12071.2259	9705.35559	11740.1236	7438.98584	5275.39933	4662.92952	6938.38511	6383.95447	10013.092	4738.89191	4658.7728
12358.4309	12552.0401	9223.55434	9697.6268	5164.68143	6069.49652	2796.60839	5862.93364	5673.93453	5577.52741	2879.5509	4299.97083
12386.2212	9841.63986	9213.84241	11522.2425	7173.85904	1625.49431	3130.14825	9695.95062	8559.24553	10270.0931	3749.64967	4658.139
12365.1561	11890.8392	9484.67564	9893.69787	6304.69463	3805.08108	4084.00609	5411.73567	5884.41574	9752.90899	5284.93456	6595.84646
7745.55305	12260.3042	6963.98316	9585.87607	4385.93647	4927.35916	474.560391	5715.05154	2419.01487	6749.57144	1657.72632	4242.83152
12388.313	11348.5812	9849.88264	10233.5892	7187.96019	1446.90263	2381.09339	3099.21726	6555.94741	6397.56006	4905.30411	7867.27217
12730.1238	12035.854	9155.0152	9360.62624	7251.00447	3965.37656	3572.42732	3594.22732	5077.23232	4510.84015	2989.16321	5273.17508
11529.2331	10134.8318	9025.81875	9117.26394	4701.59508	7043.919	2958.29914	5393.33433	4881.73308	7286.4716	2913.52309	3887.60839
12685.0084	11571.2137	8811.31425	8279.22293	7052.08264	3853.61332	5927.01258	4085.40604	6047.12151	5363.58587	3994.55193	6774.81888
12946.1062	11423.8126	9467.60888	8793.7206	7414.50746	4004.96418	5582.46527	2746.79184	6392.8102	6701.38476	8766.44232	8574.41052
8308.22723	11621.0666	7418.33574	9111.15713	7091.82106	5222.86064	5550.98425	3587.47838	3224.68796	2515.49359	4514.061	2442.3059
12896.8121	12334.6834	9371.51725	10639.3568	6618.60581	3651.8291	4785.18926	8501.21713	5189.02778	9898.93598	2200.09628	2977.7509
12307.7965	12513.438	9279.05134	9667.42716	6405.5179	7304.22048	2605.27179	2391.3451	3553.02089	5625.29873	1514.48209	2547.56135
8033.00219	12394.5813	6920.45668	12091.0017	4198.28659	6214.83108	2717.77405	9294.22782	2366.88528	10185.7277	4448.68885	4719.73681
11310.245	11664.9284	9124.08495	10199.991	6053.69187	5448.35112	1706.20619	4276.77038	3266.34331	7704.74957	3770.93353	6995.96168
12911.387	12034.0989	9559.14792	10156.2411	8030.79955	4366.8395	6899.98872	6002.5188	6939.6437	6014.61246	4073.87013	5370.57495
12589.396	11666.8312	8456.60692	10998.4546	7009.97075	2626.45619	5216.61045	6909.38444	4334.60763	6902.38075	4103.7726	4828.04787
12302.2711	12408.6915	10402.9806	10505.6813	4933.62111	5248.77557	6688.09386	4963.07512	5769.87989	7067.45521	5451.92487	5325.59525
11955.855	11682.0783	8831.2731	8923.64728	6115.12478	2402.9687	2976.78735	3914.61429	4351.72274	5694.11823	3487.75266	4235.89126
5946.96608	12418.4731	6276.11411	10075.7076	6547.57376	6574.17267	4262.63207	5645.80204	2472.28069	5410.12117	3978.22308	3719.16093
12354.4838	11788.1564	9067.68588	9691.22603	6575.84248	5007.29143	1603.39023	1764.28409	2785.14987	8413.49096	2410.10222	2860.82715
12626.717	11972.8389	10587.5963	10900.631	4682.34025	2791.60272	2596.51353	2347.87484	3979.587	6815.15308	2427.99157	3372.67665
12650.1849	12557.0921	9192.42037	10235.5333	6388.06521	5538.10903	2880.15937	3458.46608	2983.07865	6781.23039	488.090023	1488.22289
12922.7349	11828.5542	9777.52832	10818.3363	6605.75292	3917.88206	4622.31285	9877.86478	7370.96582	6600.29652	4775.35061	5409.28031
12939.9479	12196.5078	8977.58953	9012.87037	6955.99089	4339.93953	5824.5155	4512.33278	6517.19138	6964.92353	5099.6754	7075.40377
12980.7107	12187.0764	10240.8834	9948.94379	6739.06614	6189.62441	4259.2863	6594.11321	5537.79256	9041.31492	3387.03884	4909.95321
12903.5088	12030.7713	9477.13641	9042.98241	6424.41992	5238.25334	5721.81539	3658.79336	6916.25576	8010.53695	5362.90525	6993.75008
12071.7733	11309.831	9226.96429	9040.29467	5914.36289	3642.14387	6351.33727	3609.90609	6774.59333	1955.9768	3192.94235	4786.44083
12948.9027	12095.6474	9182.68287	9300.14823	4738.13809	3782.30999	4979.58318	5081.70905	7014.97044	8657.79118	6454.92873	6710.28757

ssGSEA statistics

Lung GTEx Paired t test		Brain GTEx Paired t test		Allograft rejection Paired t test		Astrocytre Paired t test		Olygodendrocyte Paired t test		SOX2 Paired t test	
P value	0.9525	P value	0.001	P value	0.0001	P value	0.0504	P value	0.0004	P value	<0.0001
P value summary	ns	P value summary	***	P value summary	***	P value summary	ns	P value summary	***	P value summary	****
Significantly different (P < 0.05)?	No	Significantly different (P < 0.05)?	Yes	Significantly different (P < 0.05)?	Yes	Significantly different (P < 0.05)?	No	Significantly different (P < 0.05)?	Yes	Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed	One- or two-tailed P value?	Two-tailed	One- or two-tailed P value?	Two-tailed	One- or two-tailed P value?	Two-tailed	One- or two-tailed P value?	Two-tailed	One- or two-tailed P value?	Two-tailed
t, df	t=0.06008, df=29	t, df	t=3.666, df=29	t, df	t=4.429, df=29	t, df	t=2.041, df=29	t, df	t=3.980, df=29	t, df	t=4.887, df=29
Number of pairs	30	Number of pairs	30	Number of pairs	30	Number of pairs	30	Number of pairs	30	Number of pairs	30

B.4.3 Metastasis purity vs ssGSEA brain score analysis

Data	
ssGSEA Brain score (unique GTEX 170 genes)	Metastasis Purity
7869.11681	0.43
11706.8383	0.45
8582.09313	0.52
8012.35057	0.72
9319.00169	0.85
7881.3228	0.6
8921.81985	0.28
6795.65459	0.78
7374.63797	0.54
8062.08111	0.4
7355.72808	0.64
9097.93529	0.3
8846.87147	0.84
10414.9034	0.84
9251.39157	0.24
10126.3643	0.55
7517.66847	0.81
8756.10382	0.56
7489.05478	0.68

Results (Linear regression)	
Best-fit values	
Slope	-0.00003015
Y-intercept	0.8398
X-intercept	27851
1/slope	-33163
Std. Error	
Slope	0.00003824
Y-intercept	0.3319
95% Confidence Intervals	
Slope	-0.0001108 to 5.052e-005
Y-intercept	0.1395 to 1.540
X-intercept	13738 to +infinity
Goodness of Fit	
R squared	0.03529
Sy.x	0.1989
Is slope significantly non-zero?	
F	0.622
DFn, DFd	1, 17
P value	0.4412
Deviation from zero?	Not Significant
Equation	Y = -3.015e-005*X + 0.8398
Data	
Number of X values	19
Maximum number of Y replicates	1
Total number of values	19
Number of missing values	0

B.4 Enriched transcription factor binding motifs within recurrent DMV hyperDMRs

B.4.1 All significantly enriched motifs

AME (Analysis of Motif Enrichment): Version 5.3.3 compiled on Feb 21 2021 at 14:51:06
The format of this file is described at https://meme-suite.org/meme/doc/ame-output-format.html.
ame --verbose 1 --oc . --scoring avg --method fisher --hit-lo-fraction 0.25 --evalue-report-threshold 10.0 --control --shuffle-- --kmer 2 p_hypo_b_hyper_autosomal_DMRs_in_at_least_2_samples_sorted_overlapping_DMVs.fasta db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme

Rank	Motif DB	Motif ID	Consensus	p-value	adj p-value	Call (from Yin, et al)
1	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MAZ_HUMAN.H11MO.0.A	GGGGGAGGGGGDGRGRRRGRRG	1.04E-49	1.04E-46	NA
2	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PATZ1_HUMAN.H11MO.0.C	GGGGMGGGGGMKGGRRVGGVRG	1.14E-43	1.24E-40	NA
3	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZBT17_HUMAN.H11MO.0.A	SRRGGWGGGGGAGGGGMRR	7.09E-43	7.62E-40	NA
4	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	EGR2_HUMAN.H11MO.0.A	GRGRRKGWGKGGGHGGRG	1.13E-38	9.40E-36	Little effect
5	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	KLF15_HUMAN.H11MO.0.A	RGGGMGGRGVDGGGGGRGG	1.26E-38	9.53E-36	MethylPlus
6	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	WT1_HUMAN.H11MO.0.C	RGGVGGGGGAGGRGGVGGRG	4.69E-38	3.74E-35	NA
7	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	VEZF1_HUMAN.H11MO.0.C	GGRRRRGRRGGAGGGGGRGRRR	1.28E-37	1.42E-34	NA
8	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN467_HUMAN.H11MO.0.C	GGGGAGGGGRRGGRGRRGGRRR	3.39E-36	2.41E-33	NA
9	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF3_HUMAN.H11MO.0.B	RAAARGGAAAVDGAAASDGA	5.86E-34	4.25E-31	Inconclusive
10	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN341_HUMAN.H11MO.0.C	GRGGRRRRGGRGRRGGAASAGC	1.78E-33	1.88E-30	NA
11	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXC10_HUMAN.H11MO.0.D	WWWWWWWNNWTTTWATKRC	7.85E-33	4.25E-30	NA
12	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN263_HUMAN.H11MO.0.A	GGGAGGAGGRRGRGGRGGRR	2.02E-31	1.49E-28	NA
13	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TAF1_HUMAN.H11MO.0.A	RARRWGGCGGMGGMGR	7.03E-30	7.24E-27	NA
14	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	KLF6_HUMAN.H11MO.0.A	GSRRGGGHGGGGMHGGGRV	3.47E-29	3.52E-26	MethylPlus
15	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO3F3_HUMAN.H11MO.0.D	WWWAWTTATGCWWATTW	1.22E-28	6.69E-26	NA
16	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	EGR2_HUMAN.H11MO.1.A	GRDGHGKGGGHGGRGG	4.86E-28	3.55E-25	Little effect
17	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	VEZF1_HUMAN.H11MO.1.C	RRRRGGAGGGGR	3.20E-27	4.54E-24	NA
18	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN394_HUMAN.H11MO.0.C	NRARWRGAAWNGAMWGRAAK	6.90E-27	5.79E-24	NA
19	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BC11A_HUMAN.H11MO.0.A	RAAAGAGGAAGTGARAV	5.79E-27	5.96E-24	NA
20	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	STAT1_HUMAN.H11MO.1.A	RRRAAAADGAAASTRAAAV	1.42E-26	9.36E-24	NA
21	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	Z354A_HUMAN.H11MO.0.C	RTTTARWYCATTTACATTTAATGT	5.71E-26	1.95E-23	NA
22	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TBX15_HUMAN.H11MO.0.D	GGGGGGGGGGGGGTGGGDR	3.50E-26	2.52E-23	NA
23	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SP4_HUMAN.H11MO.0.A	SRGVARGRGGCGGRGCHDRR	2.87E-26	2.83E-23	NA
24	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ETS2_HUMAN.H11MO.0.B	RDVRGGAARVRRR	6.12E-26	8.23E-23	MethylMinus
25	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXG1_HUMAN.H11MO.0.D	DTTGTTTAYDTWTDTTT	5.33E-25	2.34E-22	MethylPlus
26	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	STAT2_HUMAN.H11MO.0.A	RRGRAAAHGAAACTGAAAV	6.88E-25	2.80E-22	NA
27	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	CPEB1_HUMAN.H11MO.0.D	TTTTTATTYTTTTTTTT	2.87E-24	6.81E-22	NA
28	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ANDR_HUMAN.H11MO.0.A	TGTTCTTKKYTGTTTRYW	2.13E-24	1.05E-21	NA
29	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PRDM6_HUMAN.H11MO.0.C	RRRARGAAAAAAA	1.25E-24	1.17E-21	NA
30	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD12_HUMAN.H11MO.0.D	DDTTTTAHKRN	4.86E-24	3.74E-21	MethylPlus
31	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LMX1B_HUMAN.H11MO.0.D	TAATTAAHWRMYTWWWWW	2.62E-23	6.53E-21	MethylPlus
32	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PURA_HUMAN.H11MO.0.D	GGWRKKGGDGGGCVKGG	9.88E-24	8.62E-21	NA
33	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	E2F7_HUMAN.H11MO.0.B	GDGGCGGGAARDR	1.04E-23	1.12E-20	MethylMinus
34	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MNX1_HUMAN.H11MO.0.D	MBATTTDHWTHWCYHYTHH	1.83E-23	1.50E-20	MethylPlus
35	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD11_HUMAN.H11MO.0.D	DWTTTATKRHN	2.73E-23	2.02E-20	MethylPlus
36	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	EGR1_HUMAN.H11MO.0.A	VDRDGCGKGGGYGGRRR	3.44E-23	2.80E-20	Little effect
37	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GSX2_HUMAN.H11MO.0.D	YTAATTRSHWWTTAAT	8.96E-23	3.73E-20	MethylPlus
38	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	DLX1_HUMAN.H11MO.0.D	WWWWWTATGSTAATTA	6.49E-22	2.91E-19	NA
39	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PATZ1_HUMAN.H11MO.1.C	AGGGAGGGG	2.80E-22	3.57E-19	NA
40	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NFAC1_HUMAN.H11MO.0.B	AWGGAAARWVWGAMW	7.26E-22	7.02E-19	MethylPlus
41	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXP1_HUMAN.H11MO.0.A	TBTGTTTMY	2.74E-21	2.28E-18	NA
42	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GSX1_HUMAN.H11MO.0.D	TAATTAGNDDTTAAT	4.32E-21	2.30E-18	MethylPlus
43	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD10_HUMAN.H11MO.0.D	TGCTTTAATT	4.44E-21	2.39E-18	MethylPlus
44	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NKX61_HUMAN.H11MO.0.B	GAYWAATKRSHWWTTAATK	5.74E-21	2.62E-18	NA
45	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FUBP1_HUMAN.H11MO.0.D	TYGTNTTTTTTT	4.60E-21	2.87E-18	NA
46	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN281_HUMAN.H11MO.0.A	RGGWGGGGGAGGGGV	5.56E-21	3.34E-18	NA
47	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ETV5_HUMAN.H11MO.0.C	RRRSAGGAARDGRV	5.31E-21	6.41E-18	MethylMinus
48	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SP2_HUMAN.H11MO.0.A	GGSSVGGGGGCGGGGCCDGSGS	8.45E-21	7.31E-18	MethylPlus
49	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXA11_HUMAN.H11MO.0.D	TWAADTTTTAYGR	1.42E-20	7.91E-18	MethylPlus
50	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MEF2D_HUMAN.H11MO.0.A	DKCTATTTTTAG	4.40E-20	1.42E-17	NA
51	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXD3_HUMAN.H11MO.0.D	KBTGTTTNCWTWRBH	7.27E-20	6.45E-17	NA
52	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SOX2_HUMAN.H11MO.1.A	BCCWTTGTYATGYWRA	1.21E-19	7.70E-17	NA
53	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MEF2C_HUMAN.H11MO.0.A	DKCTATTTTTRGM	1.70E-19	7.72E-17	NA
54	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	KLF16_HUMAN.H11MO.0.D	GGGGGYGKGKBGGGGGGGG	1.27E-19	9.40E-17	MethylPlus
55	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SOX10_HUMAN.H11MO.1.A	BBYCTTTGTTYYY	9.21E-20	1.03E-16	NA
56	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SOX2_HUMAN.H11MO.0.A	BBCCTTTGTYYYB	1.52E-19	1.28E-16	NA
57	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN713_HUMAN.H11MO.0.D	RWAGAARAADGVSANGMAA	2.02E-19	1.42E-16	MethylPlus
58	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MEF2B_HUMAN.H11MO.0.A	DKKKCTATTTTTRG	3.22E-19	1.82E-16	NA
59	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ARI3A_HUMAN.H11MO.0.D	RATTAAWBRRWATCAAAWTWWA	1.05E-18	2.03E-16	NA
60	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FLI1_HUMAN.H11MO.0.A	SRRGGMAGGAAGGRRRGR	3.69E-19	2.84E-16	MethylMinus
61	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF1_HUMAN.H11MO.0.A	RAAANWGAAASTGAAASYRR	8.90E-19	2.92E-16	NA
62	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN350_HUMAN.H11MO.0.C	CAGYYYTTTTATDWCCHR	7.00E-19	6.02E-16	NA
63	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MEF2A_HUMAN.H11MO.0.A	DKCTATTTTTRGM	2.14E-18	9.90E-16	NA
64	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXF1_HUMAN.H11MO.0.D	WTGTTTATTTW	7.30E-18	2.18E-15	NA
65	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF7_HUMAN.H11MO.0.C	GAAASYGAAA	2.93E-18	2.26E-15	MethylMinus
66	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SRY_HUMAN.H11MO.0.B	YWTTGTTTH	3.91E-18	3.07E-15	NA
67	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	E2F6_HUMAN.H11MO.0.A	DGGGCGGGARRVR	2.97E-18	3.08E-15	NA
68	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GATA3_HUMAN.H11MO.0.A	RVAGATAAVAD	6.40E-18	4.65E-15	MethylPlus
69	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SPIC_HUMAN.H11MO.0.D	KVDRRWGGGKDRG	6.79E-18	9.07E-15	NA
70	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PIT1_HUMAN.H11MO.0.C	YTMATGAATATWY	2.27E-17	1.06E-14	MethylPlus
71	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	KLF5_HUMAN.H11MO.0.A	WGGGTGKGGCDGGG	2.11E-17	1.40E-14	Little effect
72	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NANOG_HUMAN.H11MO.0.A	BYYWTTGWNATGCAAAT	4.40E-17	2.32E-14	MethylPlus
73	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZFP28_HUMAN.H11MO.0.C	TTCTATTTCTTCTTGWGTCA	2.53E-16	4.90E-14	NA
74	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	CDC5L_HUMAN.H11MO.0.D	GBGWKTTAAYRTAWY	1.96E-16	5.13E-14	NA
75	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MSX1_HUMAN.H11MO.0.D	TAATTGSWWKTTAATTGS	4.45E-16	8.64E-14	MethylPlus
76	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXM1_HUMAN.H11MO.0.A	TGTTTRCTYWKB	2.04E-16	1.33E-13	NA
77	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BPTF_HUMAN.H11MO.0.D	KKTBTTGTKKKS	2.19E-16	1.70E-13	NA
78	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	DLX6_HUMAN.H11MO.0.D	WWWWWDNNNTAATTA	6.29E-16	2.45E-13	MethylPlus
79	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HNF1A_HUMAN.H11MO.0.C	DGTTAATKATTAACH	8.31E-16	2.80E-13	NA
80	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GBX2_HUMAN.H11MO.0.D	BTAATTRSYBHHWWWWW	8.57E-16	3.01E-13	MethylPlus
81	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HMGA1_HUMAN.H11MO.0.D	NWATTTT	4.25E-16	3.89E-13	NA
82	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BATF3_HUMAN.H11MO.0.B	BRSTTTCADTWTGASWM	8.79E-16	4.34E-13	Inconclusive
83	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXK1_HUMAN.H11MO.0.A	TGTTTMCHTT	7.16E-16	5.48E-13	NA
84	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SP3_HUMAN.H11MO.0.B	SGVVGGGGGCGGGGCBRGSS	5.70E-16	5.78E-13	Little effect
85	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	E2F3_HUMAN.H11MO.0.A	DDGGCGGGAAA	6.69E-16	6.85E-13	Little effect
86	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXO1_HUMAN.H11MO.0.A	KBBYTGTTTHCN	8.07E-16	7.15E-13	NA
87	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	VSX1_HUMAN.H11MO.0.D	TAATTAGCHWWWWW	2.89E-15	8.68E-13	MethylPlus
88	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXP3_HUMAN.H11MO.0.D	AATKTGWTT	2.48E-15	1.40E-12	MethylMinus
89	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZFX_HUMAN.H11MO.1.A	DGVRGVVSAGGCCKVGG	1.61E-15	1.93E-12	NA
90	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TBP_HUMAN.H11MO.0.A	MTATAAAARS	3.15E-15	1.94E-12	NA

Rank	Motif DB	Motif ID	Consensus	p-value	adj p-value	Call (from Yin, et al)
91	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MEOX1_HUMAN.H11MO.0.D	GGTAATTAGBGKHAWW	1.60E-14	2.40E-12	Little effect
92	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	VENTX_HUMAN.H11MO.0.D	YVYTAATYGSTTHHHWWTT	1.14E-14	4.09E-12	MethylMinus
93	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXC12_HUMAN.H11MO.0.D	DTTTTAYKACYT	1.37E-14	6.31E-12	MethylPlus
94	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZFP82_HUMAN.H11MO.0.C	TTCCADTTCAYTAATTYTCTCTTC	1.60E-14	6.68E-12	NA
95	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SP4_HUMAN.H11MO.1.A	GRRRGGGGMGGGGCBK	9.31E-15	7.12E-12	NA
96	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN148_HUMAN.H11MO.0.D	KGRGKGGGGGAGGGG	9.94E-15	8.11E-12	NA
97	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PBX2_HUMAN.H11MO.0.C	KRAWKTGATTGATKS	2.05E-14	8.65E-12	NA
98	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRX3_HUMAN.H11MO.0.D	WNWCATGTWMT	2.04E-14	9.75E-12	MethylPlus
99	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TBX1_HUMAN.H11MO.0.D	GGGYGGGYAGGGGTGGGMDD	2.95E-14	1.14E-11	NA
100	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LYL1_HUMAN.H11MO.0.A	CCAKCTGYTYYCTB	1.04E-14	1.19E-11	NA
101	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZBTB6_HUMAN.H11MO.0.C	GGWGMTGGAGCCD	2.70E-14	1.31E-11	NA
102	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO5F1_HUMAN.H11MO.0.A	CYWTTSWYATGCAAAT	4.70E-14	1.73E-11	MethylPlus
103	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXB13_HUMAN.H11MO.0.A	TTTTATDGSH	3.25E-14	1.80E-11	MethylPlus
104	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LEF1_HUMAN.H11MO.0.A	BCCTTTSWWNTBYH	1.60E-14	1.84E-11	MethylMinus
105	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXB3_HUMAN.H11MO.0.D	YTAATTASYTTTTW	4.53E-14	1.97E-11	NA
106	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BSH_HUMAN.H11MO.0.D	TAATKVGTTAATBR	5.63E-14	2.30E-11	MethylMinus
107	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FEV_HUMAN.H11MO.0.B	GCVGGAAGYG	1.75E-14	2.63E-11	MethylMinus
108	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SOX4_HUMAN.H11MO.0.B	BYCTTTGTYYYB	3.07E-14	2.68E-11	NA
109	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LMX1A_HUMAN.H11MO.0.D	TAATTAATTWKKTWATTWW	2.00E-13	4.21E-11	MethylPlus
110	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXJ3_HUMAN.H11MO.0.A	TTGTTTATKKTTW	2.86E-13	4.54E-11	MethylPlus
111	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TEF_HUMAN.H11MO.0.D	TGTTTATGTAAMTK	1.71E-13	4.98E-11	MethylMinus
112	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF8_HUMAN.H11MO.0.B	RAAARRGGAASTGAAASTDR	1.98E-13	7.02E-11	Inconclusive
113	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXA9_HUMAN.H11MO.0.B	RTGATTTATKR	1.82E-13	7.09E-11	MethylPlus
114	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TBX21_HUMAN.H11MO.0.A	GDGDRGGTGTGRV	8.25E-14	7.18E-11	NA
115	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PAX4_HUMAN.H11MO.0.D	TWAYTAATTMRH	1.58E-13	7.51E-11	Little effect
116	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN282_HUMAN.H11MO.0.D	MGKRTTRTGGGARA	1.60E-13	8.69E-11	NA
117	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXJ3_HUMAN.H11MO.1.B	TKKTTWWTGTTTA	6.89E-13	1.01E-10	MethylPlus
118	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SALL4_HUMAN.H11MO.0.B	GSDGGGWGGG	9.53E-14	1.07E-10	NA
119	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZSC22_HUMAN.H11MO.0.C	RGKCWGAGGGMGGAGG	1.88E-13	1.16E-10	NA
120	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF2_HUMAN.H11MO.0.A	RAAAVHGAAAGTGAAASTRV	5.01E-13	1.26E-10	MethylPlus
121	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	WT1_HUMAN.H11MO.1.B	GWGGGAGGAGS	1.26E-13	1.29E-10	NA
122	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO4F1_HUMAN.H11MO.0.D	CATTAATTATTCAT	9.64E-13	1.54E-10	NA
123	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO3F1_HUMAN.H11MO.0.C	VATKSWWATGCAAW	4.69E-13	2.20E-10	MethylPlus
124	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PROP1_HUMAN.H11MO.0.D	TAATTRRATTA	8.02E-13	2.76E-10	MethylPlus
125	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MEIS1_HUMAN.H11MO.0.A	RTGATTKATKRB	1.15E-12	4.02E-10	NA
126	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HAND1_HUMAN.H11MO.1.D	KSYCTGG	3.02E-13	4.13E-10	NA
127	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXJ2_HUMAN.H11MO.0.C	TGTTTRTTTW	1.63E-12	4.79E-10	NA
128	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXA10_HUMAN.H11MO.0.C	KATRRTTTATGW	1.71E-12	5.67E-10	MethylPlus
129	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NKX62_HUMAN.H11MO.0.D	NDTTAATKASN	1.63E-12	6.98E-10	NA
130	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXC11_HUMAN.H11MO.0.D	DTTTWAYGACY	1.53E-12	7.47E-10	MethylPlus
131	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF4_HUMAN.H11MO.0.A	RAAWGRGGAASTGAAASY	2.74E-12	1.14E-09	Little effect and MethylPlus
132	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ISX_HUMAN.H11MO.0.D	YTAATTRGHTWWWWWW	8.62E-12	1.72E-09	MethylPlus
133	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PROX1_HUMAN.H11MO.0.D	KGGKARGGCGKSKKGGG	2.08E-12	1.82E-09	MethylMinus
134	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZIM3_HUMAN.H11MO.0.C	KTRGGTTTCTGTTKYYTKY	1.26E-11	1.84E-09	NA
135	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXA13_HUMAN.H11MO.0.C	GDTTTTATTGG	5.89E-12	2.44E-09	MethylPlus and MethylMinus
136	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO3F4_HUMAN.H11MO.0.D	TMATTTGCATAATTWAW	1.48E-11	2.70E-09	MethylPlus
137	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MAZ_HUMAN.H11MO.1.A	RRGGGAGGGGS	3.56E-12	2.95E-09	NA
138	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NFAT5_HUMAN.H11MO.0.D	RRRGGAARRDN	2.82E-12	3.68E-09	MethylPlus
139	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SOX13_HUMAN.H11MO.0.D	YATTGTTY	6.36E-12	4.48E-09	NA
140	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LHX2_HUMAN.H11MO.0.A	KBYTAATTRGYK	9.94E-12	4.82E-09	NA
141	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LHX9_HUMAN.H11MO.0.D	YTAATTAGYNHWWWWWWD	2.41E-11	4.84E-09	MethylPlus
142	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN418_HUMAN.H11MO.0.C	TGCTTYTRGCYTCTKVY	9.77E-12	5.21E-09	NA
143	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HNF1B_HUMAN.H11MO.0.A	DGTTAATRATTAACH	2.72E-11	5.64E-09	NA
144	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXA1_HUMAN.H11MO.0.A	TGTTTACWYWGB	2.11E-11	6.24E-09	NA
145	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PRGR_HUMAN.H11MO.0.A	RGAACATTYTGTTCTB	1.78E-11	6.68E-09	NA
146	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXP2_HUMAN.H11MO.0.C	BTGTTTACH	2.10E-11	1.05E-08	NA
147	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO2F1_HUMAN.H11MO.0.C	AWKBRWATGSAAWYAW	2.99E-11	1.30E-08	NA
148	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO4F3_HUMAN.H11MO.0.D	TCATTAATTATKCAT	1.36E-10	1.41E-08	NA
149	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PRDM1_HUMAN.H11MO.0.A	RAAAGTGAAAGTGR	2.64E-11	1.60E-08	NA
150	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GATA2_HUMAN.H11MO.1.A	DSAGATAAGRR	3.33E-11	1.72E-08	NA
151	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SUH_HUMAN.H11MO.0.A	VSYSTGGGAAV	1.84E-11	1.80E-08	NA
152	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXA2_HUMAN.H11MO.0.A	TGTTTACWYWGB	7.63E-11	1.98E-08	NA
153	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	P5F1B_HUMAN.H11MO.0.D	DMATTTGCATATTYAW	1.30E-10	2.27E-08	POU5F1=MethylPlus
154	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN770_HUMAN.H11MO.0.C	GGGAGGCYGAGGBRGGAGGATC	4.90E-11	2.29E-08	NA
155	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXO3_HUMAN.H11MO.0.B	BYTGTTTACH	5.74E-11	2.35E-08	NA
156	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NOTO_HUMAN.H11MO.0.D	RTTAATTAGNT	7.60E-11	2.45E-08	NA
157	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZFHX3_HUMAN.H11MO.0.D	ATTAWTAATTA	1.62E-10	3.05E-08	NA
158	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXB1_HUMAN.H11MO.0.D	RTMAATRTTTACATAAB	2.78E-10	3.42E-08	NA
159	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GBX1_HUMAN.H11MO.0.D	BTAATTRRYNHWWWWWW	1.55E-10	3.53E-08	NA
160	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MIXL1_HUMAN.H11MO.0.D	GBTAATTARAT	9.80E-11	3.70E-08	MethylPlus
161	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MEOX2_HUMAN.H11MO.0.D	TTAATKAYBWT	1.11E-10	3.73E-08	Little effect
162	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXA3_HUMAN.H11MO.0.B	CHTGTTTACWTWG	1.18E-10	3.99E-08	NA
163	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	DMBX1_HUMAN.H11MO.0.D	YMMTWWWAAYCYAATTAABD	1.66E-10	4.15E-08	NA
164	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ALX1_HUMAN.H11MO.0.B	RTAATTRVATTA	1.35E-10	4.15E-08	NA
165	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TFDP1_HUMAN.H11MO.0.C	RRRRGGCGGGAARN	5.66E-11	4.34E-08	NA
166	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	EMX1_HUMAN.H11MO.0.D	TAATTAGYTAAW	1.96E-10	4.37E-08	NA
167	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SIX2_HUMAN.H11MO.0.A	KGAAAYHTGAKMC	1.00E-10	5.11E-08	NA
168	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXC13_HUMAN.H11MO.0.D	DDTTTTACKAG	1.08E-10	5.51E-08	MethylPlus
169	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ERG_HUMAN.H11MO.0.A	RRRSAGGAAGYGG	7.74E-11	5.81E-08	MethylMinus
170	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SOX3_HUMAN.H11MO.0.B	BCCWTTGTBYB	1.03E-10	7.08E-08	NA
171	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	EMX2_HUMAN.H11MO.0.D	TAATTAGYDWAW	3.66E-10	8.35E-08	NA
172	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	RAX2_HUMAN.H11MO.0.D	YTAATTARATWA	2.97E-10	8.67E-08	RAX=MethylPlus
173	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ESX1_HUMAN.H11MO.0.D	RYTAATTRGYNHWWWTW	4.99E-10	9.09E-08	NA
174	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LHX4_HUMAN.H11MO.0.D	DNRYTAATTRN	2.05E-10	9.17E-08	MethylPlus
175	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PDX1_HUMAN.H11MO.1.A	ATTAATKAB	2.26E-10	1.02E-07	MethylPlus
176	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BARH2_HUMAN.H11MO.0.D	VYTAAWYGNTSHWWWWHR	3.86E-10	1.11E-07	BARHL2=MethylMinus
177	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD4_HUMAN.H11MO.0.D	TTAATTGW	2.28E-10	1.11E-07	MethylPlus
178	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXA2_HUMAN.H11MO.0.D	NGTAATTARBD	3.51E-10	1.19E-07	MethylPlus
179	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZBT14_HUMAN.H11MO.0.C	GGAGCGCGC	3.78E-10	1.82E-07	MethylMinus
180	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ARI5B_HUMAN.H11MO.0.C	DBYKDGTATTSKR	5.03E-10	1.94E-07	NA

Rank	Motif DB	Motif ID	Consensus	p-value	adj p-value	Call (from Yin, et al)
181	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SMAD3_HUMAN.H11MO.0.B	SYCTSHCWSCWS	1.51E-10	1.98E-07	NA
182	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	KLF3_HUMAN.H11MO.0.B	GGRGVDGGGCGGGGCYGGG	3.03E-10	2.10E-07	MethylPlus
183	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	UBIP1_HUMAN.H11MO.0.D	SCAGARA	2.42E-10	2.31E-07	MethylPlus
184	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FEZF1_HUMAN.H11MO.0.C	DYTGYYCTTTTH	3.40E-10	3.05E-07	NA
185	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	RXRA_HUMAN.H11MO.0.A	VRGDVRGNVARAGGTCARRG	3.42E-10	3.37E-07	NA
186	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PRRX2_HUMAN.H11MO.0.C	YTAATTR	1.09E-09	3.44E-07	MethylPlus
187	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NFAC1_HUMAN.H11MO.1.B	AATGGAAAAW	5.41E-10	3.89E-07	MethylPlus
188	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LHX3_HUMAN.H11MO.0.C	AAAATTAATTARY	3.22E-09	4.18E-07	NA
189	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NR2E3_HUMAN.H11MO.0.C	AAAGTCAAAAGTCA	1.12E-09	5.05E-07	NA
190	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD9_HUMAN.H11MO.0.D	ARWTTNATKD	1.32E-09	5.40E-07	MethylPlus
191	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	VAX1_HUMAN.H11MO.0.D	RSTAATTAGBSTWAWTWW	6.89E-09	5.51E-07	NA
192	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BATF_HUMAN.H11MO.0.A	RBTYTCRDWATGASTCAB	1.77E-09	5.67E-07	BATF3=inconclusive (different gene cards entry)
193	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	RX_HUMAN.H11MO.0.D	TTAATTRRYWA	2.95E-09	6.61E-07	MethylPlus
194	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZF64A_HUMAN.H11MO.0.D	GGGGCCCGGGAMYCDG	1.11E-09	7.08E-07	NA
195	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TWST1_HUMAN.H11MO.0.A	MCATCTGKWWYYMATTA	3.29E-09	7.21E-07	NA
196	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HMX2_HUMAN.H11MO.0.D	DTTAAKTGNTT	1.50E-09	7.24E-07	NA
197	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MSX2_HUMAN.H11MO.0.D	TAATTDK	1.47E-09	8.28E-07	MethylPlus
198	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ETV7_HUMAN.H11MO.0.D	VAGGAAR	7.44E-10	9.43E-07	MethylMinus
199	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO3F2_HUMAN.H11MO.0.A	YYCATGCATWTYCATTM	2.63E-09	9.64E-07	MethylPlus
200	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXC6_HUMAN.H11MO.0.D	ATGATTTATTACTTT	5.11E-08	1.33E-06	NA
201	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZNF41_HUMAN.H11MO.0.C	MACAAGGRARTMARVCACCATGAG	6.45E-09	1.38E-06	NA
202	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXC8_HUMAN.H11MO.0.D	YTTSATTRATDSSC	3.58E-09	1.39E-06	MethylPlus
203	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SOX5_HUMAN.H11MO.0.C	YATTGTTW	2.73E-09	1.51E-06	NA
204	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SMCA1_HUMAN.H11MO.0.C	CMARGAAGMATD	1.92E-09	1.59E-06	NA
205	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	DLX4_HUMAN.H11MO.0.D	TAATTASNBHWWWWWD	7.88E-09	1.66E-06	MethylPlus
206	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXA5_HUMAN.H11MO.0.D	YTRATTAATG	4.89E-09	1.70E-06	MethylPlus
207	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXF2_HUMAN.H11MO.0.D	TGTTTACHTWD	5.52E-09	1.74E-06	NA
208	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NFAC4_HUMAN.H11MO.0.C	DGGAAAAYTW	2.38E-09	1.97E-06	MethylPlus
209	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXC9_HUMAN.H11MO.0.C	WTTTATKGCH	4.18E-09	2.01E-06	MethylPlus
210	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	EVI1_HUMAN.H11MO.0.B	WAGAYAAGATAANAKA	1.67E-08	2.24E-06	NA
211	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BARH1_HUMAN.H11MO.0.D	TAAWYGBYBHTAA	9.17E-09	2.39E-06	MethylMinus
212	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SHOX_HUMAN.H11MO.0.D	YTAATTRRYTWWWWWWD	2.87E-08	3.59E-06	MethylPlus
213	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SOX17_HUMAN.H11MO.0.C	SCCATTGTKBY	5.46E-09	3.98E-06	NA
214	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	VAX2_HUMAN.H11MO.0.D	DWWARSBTAATTAGTG	1.79E-08	4.14E-06	NA
215	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO2F3_HUMAN.H11MO.0.D	DTYATGCWWAWTTR	1.19E-08	4.27E-06	MethylPlus
216	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MYNN_HUMAN.H11MO.0.D	RGRCTYTTATTTTGAAA	1.41E-08	4.47E-06	NA
217	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BARX1_HUMAN.H11MO.0.D	BWAATKGTTTYY	6.66E-09	4.60E-06	NA
218	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ONEC3_HUMAN.H11MO.0.D	TATYGATTTTTT	2.16E-08	4.66E-06	MethylMinus
219	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZBT49_HUMAN.H11MO.0.D	NBSNTAATTRV	1.06E-08	4.99E-06	NA
220	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	E2F4_HUMAN.H11MO.0.A	SRGGGCGGGAARD	6.85E-09	5.01E-06	MethylMinus
221	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	DLX3_HUMAN.H11MO.0.C	GMTAATTRSW	2.25E-08	5.13E-06	MethylPlus
222	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SHOX2_HUMAN.H11MO.0.D	YTAATTRRYYHWWWWWD	5.37E-08	5.96E-06	NA
223	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZNF85_HUMAN.H11MO.0.C	AWAGAARWCTGMWGKAATCT	3.07E-08	6.93E-06	NA
224	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO5F1_HUMAN.H11MO.1.A	WHATGCAVAT	1.41E-08	7.70E-06	MethylPlus
225	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SP2_HUMAN.H11MO.1.B	RGGRGGCGGGGC	1.30E-08	9.56E-06	MethylPlus
226	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXO4_HUMAN.H11MO.0.C	TTGTTTAYK	2.69E-08	9.88E-06	NA
227	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	UNC4_HUMAN.H11MO.0.D	TAATTTAATTA	7.14E-08	1.10E-05	MethylPlus
228	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN394_HUMAN.H11MO.1.D	GAAWGGAAW	1.19E-08	1.13E-05	NA
229	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	DLX5_HUMAN.H11MO.0.D	TAATTAYAGGK	2.20E-08	1.17E-05	MethylPlus
230	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ETV2_HUMAN.H11MO.0.B	RRARRCAGGAARYRGS	3.53E-08	1.39E-05	MethylMinus
231	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GCR_HUMAN.H11MO.0.A	RGAACABMVTGTNCY	5.89E-08	1.48E-05	MethylPlus
232	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN816_HUMAN.H11MO.1.C	RADGGGGACAKGHW	1.87E-08	1.52E-05	NA
233	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ETV4_HUMAN.H11MO.0.B	VRCAGGAARBR	1.54E-08	1.61E-05	MethylMinus
234	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	CDX1_HUMAN.H11MO.0.C	HTTTATGD	2.59E-08	1.95E-05	MethylPlus
235	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXD1_HUMAN.H11MO.0.D	CWWTGTTTACHTWVR	5.39E-08	1.97E-05	NA
236	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LHX8_HUMAN.H11MO.0.D	YRYTRATTRCMATTAATT	1.52E-07	2.03E-05	NA
237	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	EGR4_HUMAN.H11MO.0.D	GGSGGYRGGGM	4.48E-08	2.29E-05	MethylPlus
238	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FLI1_HUMAN.H11MO.1.A	RRRCAGGAAGBRG	3.27E-08	2.42E-05	MethylMinus
239	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	RREB1_HUMAN.H11MO.0.D	DKGKKKGKGGKTGKTTTGGGKT	1.22E-07	2.50E-05	NA
240	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN214_HUMAN.H11MO.0.C	WCMWTYYTRWGSMHYTTGATGA	7.42E-08	3.22E-05	NA
241	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IKZF1_HUMAN.H11MO.0.C	BTGGGARR	3.67E-08	3.90E-05	NA
242	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HME1_HUMAN.H11MO.0.D	VCTAATTRSYHAWWWWW	2.96E-07	4.00E-05	MethylPlus
243	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD8_HUMAN.H11MO.0.D	WAATWTGCAYN	6.04E-08	4.17E-05	MethylPlus
244	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	E2F1_HUMAN.H11MO.0.A	RRNKGGCGGGAARR	6.36E-08	4.53E-05	MethylMinus
245	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TF7L1_HUMAN.H11MO.0.B	DBBCTTTGAWSTB	6.52E-08	4.62E-05	MethylMinus
246	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HME2_HUMAN.H11MO.0.D	NNTRRTTARNB	6.96E-08	5.48E-05	MethylPlus
247	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ETV6_HUMAN.H11MO.0.D	RCAGGAARK	5.44E-08	5.81E-05	NA
248	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO6F2_HUMAN.H11MO.0.D	KWTAATKAGST	4.03E-07	6.53E-05	NA
249	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN770_HUMAN.H11MO.1.C	DGGAGGCYGRG	1.44E-07	7.64E-05	NA
250	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ALX3_HUMAN.H11MO.0.D	TAATTARATTA	3.86E-07	7.99E-05	MethylPlus
251	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MZF1_HUMAN.H11MO.0.B	RRDRGGGAWTKRR	9.41E-08	8.53E-05	NA
252	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	E2F4_HUMAN.H11MO.1.A	GGCGGGAADTTBNR	1.92E-07	8.92E-05	MethylMinus
253	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ELF3_HUMAN.H11MO.0.A	RAABVAGGAAGTRR	2.16E-07	9.17E-05	MethylMinus
254	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXO6_HUMAN.H11MO.0.D	WWSWTGTTTACRT	5.08E-07	9.56E-05	NA
255	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ELF5_HUMAN.H11MO.0.A	VGRASSAGGAAGTRV	1.59E-07	1.01E-04	MethylMinus
256	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HNF1B_HUMAN.H11MO.1.A	DGTTAATVATT	4.47E-07	1.02E-04	NA
257	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN816_HUMAN.H11MO.0.C	AAARAAGGGGGACATGYWGGG	4.47E-07	1.13E-04	NA
258	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SPI1_HUMAN.H11MO.0.A	RRAAAGAGGAAGTGRVD	2.93E-07	1.13E-04	NA
259	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	AIRE_HUMAN.H11MO.0.C	HTTGGWWWWAWTGGDTWH	2.50E-07	1.20E-04	NA
260	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD3_HUMAN.H11MO.0.D	NGKTAATTAGK	3.68E-07	1.25E-04	MethylPlus
261	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXC1_HUMAN.H11MO.0.C	CDWTGTTTACTTWRG	1.37E-06	1.49E-04	NA
262	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXD2_HUMAN.H11MO.0.D	WAWGTAAATRTTWWYTTAA	1.92E-06	1.50E-04	NA
263	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LBX2_HUMAN.H11MO.0.D	RYTAATTRGBY	9.26E-07	1.74E-04	MethylPlus
264	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	OSR2_HUMAN.H11MO.0.C	GBYSCWGCTNCTGYDG	2.74E-07	1.94E-04	Inconclusive
265	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TCF7_HUMAN.H11MO.0.A	RMMMACATCAAAGGVA	3.65E-07	2.44E-04	MethylMinus
266	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SRBP2_HUMAN.H11MO.0.B	VSGTGGGGWGABG	2.92E-07	2.48E-04	MethylMinus
267	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NFAC2_HUMAN.H11MO.0.B	TGGAAAAWY	3.64E-07	2.70E-04	MethylPlus
268	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PBX1_HUMAN.H11MO.1.C	WAWTTGATTGATGDB	6.87E-07	2.88E-04	MethylPlus
269	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PTF1A_HUMAN.H11MO.0.B	SCAGCTGBYBBYBYYCHC	6.12E-07	3.56E-04	Little effect
270	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	CUX2_HUMAN.H11MO.0.D	WRATCAATRAA	5.94E-07	3.56E-04	Inconclusive

Rank	Motif DB	Motif ID	Consensus	p-value	adj p-value	Call (from Yin, et al)
271	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HMX1_HUMAN.H11MO.0.D	DTTAATTGHTTTTTAAT	3.46E-06	4.11E-04	NA
272	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXQ1_HUMAN.H11MO.0.C	HATTGTTTATTT	6.43E-06	4.63E-04	NA
273	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	KLF1_HUMAN.H11MO.0.A	DGGGYGKGGCYGGG	9.00E-07	5.14E-04	NA
274	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GCR_HUMAN.H11MO.1.A	HRRGAACAGWVW	1.05E-06	5.24E-04	MethylPlus
275	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ANDR_HUMAN.H11MO.1.A	RGNACABKSTGTTCYY	2.03E-06	5.97E-04	NA
276	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BARX2_HUMAN.H11MO.0.D	TYRWTAATKR	1.85E-06	6.69E-04	NA
277	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PRGR_HUMAN.H11MO.1.A	RGRACAGAM	9.41E-07	6.75E-04	NA
278	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXB7_HUMAN.H11MO.0.C	TTGATTRATK	7.61E-06	6.85E-04	MethylPlus
279	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ETV3_HUMAN.H11MO.0.D	KWMRTTTGCACTTC	2.08E-06	7.13E-04	MethylMinus
280	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	XBP1_HUMAN.H11MO.0.D	GSKGACGTGGC	2.06E-06	7.24E-04	MethylMinus
281	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NR1D1_HUMAN.H11MO.0.B	RRRNMASTGGGTCAGDRDR	1.28E-06	9.45E-04	NA
282	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF9_HUMAN.H11MO.0.C	GAAAGCGAAAYT	5.66E-06	9.78E-04	MethylPlus and Little effect
283	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TBX3_HUMAN.H11MO.0.C	DRAGGTGBSAR	1.15E-06	1.05E-03	NA
284	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TAL1_HUMAN.H11MO.1.A	VARSAARCAGVTGKB	1.75E-06	1.13E-03	NA
285	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO2F2_HUMAN.H11MO.0.A	AYATGCAAATK	3.08E-06	1.19E-03	MethylPlus
286	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF5_HUMAN.H11MO.0.D	KAAAGGRAARCMAAAWSWGA	2.47E-05	1.26E-03	Little effect
287	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HIF1A_HUMAN.H11MO.0.C	GDACGTGM	3.69E-06	1.30E-03	NA
288	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BRAC_HUMAN.H11MO.0.A	YTVRCAKSDARGTGTGAADT	3.31E-06	1.44E-03	NA
289	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	VSX2_HUMAN.H11MO.0.D	YTAATTAGCYTV	3.10E-05	1.45E-03	MethylPlus
290	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SOX9_HUMAN.H11MO.1.B	YCWTTGKYY	1.87E-06	1.67E-03	NA
291	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PHX2A_HUMAN.H11MO.0.D	TAATYTAATTA	1.27E-05	1.69E-03	NA
292	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SIX1_HUMAN.H11MO.0.A	TGWAAYHTGAKMCY	6.05E-06	1.73E-03	Little effect
293	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ASCL2_HUMAN.H11MO.0.D	VCASCTGCT	3.57E-06	1.74E-03	MethylMinus
294	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SOX18_HUMAN.H11MO.0.D	SHWHVMWTTGTTBTYKTCC	8.64E-06	2.02E-03	NA
295	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	DRGX_HUMAN.H11MO.0.D	TAATYTAATTA	2.28E-05	2.03E-03	NA
296	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NOBOX_HUMAN.H11MO.0.C	HTAATTRSY	8.02E-06	2.27E-03	NA
297	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	CLOCK_HUMAN.H11MO.0.C	CWGHCACGTGMVVM	4.53E-06	2.44E-03	MethylMinus and MethylPlus
298	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ETS1_HUMAN.H11MO.0.A	RRRCMGGAAGTGG	4.86E-06	2.82E-03	NA
299	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PHX2B_HUMAN.H11MO.0.D	TAATYTAATTA	1.53E-05	2.84E-03	NA
300	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ANDR_HUMAN.H11MO.2.A	BTSTGTTCYYKB	5.10E-06	3.32E-03	NA
301	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO4F2_HUMAN.H11MO.0.D	YATTAATGAGYTK	2.97E-05	4.30E-03	NA
302	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	COT1_HUMAN.H11MO.0.C	GRGGKSARAGGTCAGGG	6.41E-06	4.54E-03	MethylPlus
303	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MYF6_HUMAN.H11MO.0.C	CASCTGC	7.29E-06	4.66E-03	MethylPlus
304	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SOX9_HUMAN.H11MO.0.B	AADGGNNBCWTTGTTY	1.55E-05	4.86E-03	NA
305	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZFX_HUMAN.H11MO.0.A	SCAGGCCBCR	5.68E-06	4.87E-03	NA
306	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRX2_HUMAN.H11MO.0.D	RCDKGKBDKGDCRKGTDVK	9.67E-06	4.89E-03	NA
307	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ASCL1_HUMAN.H11MO.0.A	CHVCASCTGCYBYY	1.44E-05	6.98E-03	NA
308	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NFAC3_HUMAN.H11MO.0.B	TGGAAAAHY	1.10E-05	7.43E-03	MethylPlus
309	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	THAP1_HUMAN.H11MO.0.C	SGCCGCCATSTYGGSYBCGGGC	1.29E-05	7.50E-03	NA
310	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BRAC_HUMAN.H11MO.1.B	WGGTGTGAAAD	1.44E-05	8.48E-03	NA
311	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BHE23_HUMAN.H11MO.0.D	RRSRTRTGKTDB	1.49E-05	8.60E-03	NA
312	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD13_HUMAN.H11MO.0.D	TTTATTAGRGA	1.50E-05	8.83E-03	MethylPlus
313	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXH1_HUMAN.H11MO.0.A	TGTGGATTS	3.27E-05	1.17E-02	NA
314	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GATA6_HUMAN.H11MO.0.A	DVAGATAAGAWAA	5.68E-05	1.17E-02	Little effect
315	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NKX61_HUMAN.H11MO.1.B	WAATGRV	1.73E-05	1.23E-02	NA

B.4.2 Significantly enriched motifs with methylation preference specified in Yin, et al.

Rank	Motif DB	Motif ID	Consensus	p-value	adj p-value	Call (from Yin, et al)
4	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	EGR2_HUMAN.H11MO.0.A	GRGRRKGWGKGGGHGGRG	1.13E-38	9.40E-36	Little effect
5	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	KLF15_HUMAN.H11MO.0.A	RGGGMGGRGVDGGGGGRGG	1.26E-38	9.53E-36	MethylPlus
9	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF3_HUMAN.H11MO.0.B	RAAARGGAAAVDGAAASDGA	5.86E-34	4.25E-31	Inconclusive
14	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	KLF6_HUMAN.H11MO.0.A	GSRRGGGHGGGGMHGGGRV	3.47E-29	3.52E-26	MethylPlus
16	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	EGR2_HUMAN.H11MO.1.A	GRDGHGKGGGHGGRGG	4.86E-28	3.55E-25	Little effect
24	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ETS2_HUMAN.H11MO.0.B	RDVRGGAARVRRR	6.12E-26	8.23E-23	MethylMinus
25	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXG1_HUMAN.H11MO.0.D	DTTGTTTAYDTWTDTTT	5.33E-25	2.34E-22	MethylPlus
30	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD12_HUMAN.H11MO.0.D	DDTTTTAHKRN	4.86E-24	3.74E-21	MethylPlus
31	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LMX1B_HUMAN.H11MO.0.D	TAATTAAHWRMYTWWWWW	2.62E-23	6.53E-21	MethylPlus
33	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	E2F7_HUMAN.H11MO.0.B	GDGGCGGGAARDR	1.04E-23	1.12E-20	MethylMinus
34	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MNX1_HUMAN.H11MO.0.D	MBATTTDHWTHWCYHYTHH	1.83E-23	1.50E-20	MethylPlus
35	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD11_HUMAN.H11MO.0.D	DWTTTATKRHN	2.73E-23	2.02E-20	MethylPlus
36	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	EGR1_HUMAN.H11MO.0.A	VDRDGCGKGGGYGGRRR	3.44E-23	2.80E-20	Little effect
37	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GSX2_HUMAN.H11MO.0.D	YTAATTRSHWWTTAAT	8.96E-23	3.73E-20	MethylPlus
40	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NFAC1_HUMAN.H11MO.0.B	AWGGAAARWVWGAMW	7.26E-22	7.02E-19	MethylPlus
42	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GSX1_HUMAN.H11MO.0.D	TAATTAGNDDTTAAT	4.32E-21	2.30E-18	MethylPlus
43	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD10_HUMAN.H11MO.0.D	TGCTTTAATT	4.44E-21	2.39E-18	MethylPlus
47	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ETV5_HUMAN.H11MO.0.C	RRRSAGGAARDGRV	5.31E-21	6.41E-18	MethylMinus
48	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SP2_HUMAN.H11MO.0.A	GGSSVGGGGGCGGGGCCDGSGS	8.45E-21	7.31E-18	MethylPlus
49	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXA11_HUMAN.H11MO.0.D	TWAADTTTTAYGR	1.42E-20	7.91E-18	MethylPlus
54	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	KLF16_HUMAN.H11MO.0.D	GGGGGYGKGKBGGGGGGG	1.27E-19	9.40E-17	MethylPlus
57	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZN713_HUMAN.H11MO.0.D	RWAGAARAADGVSANGMAA	2.02E-19	1.42E-16	MethylPlus
60	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FLI1_HUMAN.H11MO.0.A	SRRGGMAGGAAGGRRRGR	3.69E-19	2.84E-16	MethylMinus
65	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF7_HUMAN.H11MO.0.C	GAAASYGAAA	2.93E-18	2.26E-15	MethylMinus
68	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GATA3_HUMAN.H11MO.0.A	RVAGATAAVAD	6.40E-18	4.65E-15	MethylPlus
70	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PIT1_HUMAN.H11MO.0.C	YTMATGAATATWY	2.27E-17	1.06E-14	MethylPlus
71	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	KLF5_HUMAN.H11MO.0.A	WGGGTGKGGCDGGG	2.11E-17	1.40E-14	Little effect
72	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NANOG_HUMAN.H11MO.0.A	BYYWTTGWNATGCAAAT	4.40E-17	2.32E-14	MethylPlus
75	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MSX1_HUMAN.H11MO.0.D	TAATTGSWWKTTAATTGS	4.45E-16	8.64E-14	MethylPlus
78	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	DLX6_HUMAN.H11MO.0.D	WWWWWDNNNTAATTA	6.29E-16	2.45E-13	MethylPlus
80	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GBX2_HUMAN.H11MO.0.D	BTAATTRSYBHHWWWWW	8.57E-16	3.01E-13	MethylPlus
82	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BATF3_HUMAN.H11MO.0.B	BRSTTTCADTWTGASWM	8.79E-16	4.34E-13	Inconclusive
84	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SP3_HUMAN.H11MO.0.B	SGVVGGGGGCGGGGCBRGSS	5.70E-16	5.78E-13	Little effect
85	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	E2F3_HUMAN.H11MO.0.A	DDGGCGGGAAA	6.69E-16	6.85E-13	Little effect
87	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	VSX1_HUMAN.H11MO.0.D	TAATTAGCHWWWWW	2.89E-15	8.68E-13	MethylPlus
88	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXP3_HUMAN.H11MO.0.D	AATKTGWTT	2.48E-15	1.40E-12	MethylMinus
91	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MEOX1_HUMAN.H11MO.0.D	GGTAATTAGBGKHAWW	1.60E-14	2.40E-12	Little effect
92	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	VENTX_HUMAN.H11MO.0.D	YVYTAATYGSTTHHHWWTT	1.14E-14	4.09E-12	MethylMinus
93	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXC12_HUMAN.H11MO.0.D	DTTTTAYKACYT	1.37E-14	6.31E-12	MethylPlus
98	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRX3_HUMAN.H11MO.0.D	WNWCATGTWMT	2.04E-14	9.75E-12	MethylPlus
102	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	POU5F1_HUMAN.H11MO.0.A	CYWTTSWYATGCAAAT	4.70E-14	1.73E-11	MethylPlus
103	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXB13_HUMAN.H11MO.0.A	TTTTATDGSH	3.25E-14	1.80E-11	MethylPlus
104	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LEF1_HUMAN.H11MO.0.A	BCCTTTSWWNTBYH	1.60E-14	1.84E-11	MethylMinus
106	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BSH_HUMAN.H11MO.0.D	TAATKVGTTAATBR	5.63E-14	2.30E-11	MethylMinus
107	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FEV_HUMAN.H11MO.0.B	GCVGGAAGYG	1.75E-14	2.63E-11	MethylMinus
109	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LMX1A_HUMAN.H11MO.0.D	TAATTAATTWKKTWATTWW	2.00E-13	4.21E-11	MethylPlus
110	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXJ3_HUMAN.H11MO.0.A	TTGTTTATKKTTW	2.86E-13	4.54E-11	MethylPlus
111	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TEF_HUMAN.H11MO.0.D	TGTTTATGTAAMTK	1.71E-13	4.98E-11	MethylMinus
112	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF8_HUMAN.H11MO.0.B	RAAARRGGAASTGAAASTDR	1.98E-13	7.02E-11	Inconclusive
113	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXA9_HUMAN.H11MO.0.B	RTGATTTATKR	1.82E-13	7.09E-11	MethylPlus
115	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PAX4_HUMAN.H11MO.0.D	TWAYTAATTMRH	1.58E-13	7.51E-11	Little effect
117	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FOXJ3_HUMAN.H11MO.1.B	TKKTTWWTGTTTA	6.89E-13	1.01E-10	MethylPlus
120	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF2_HUMAN.H11MO.0.A	RAAAVHGAAAGTGAAASTRV	5.01E-13	1.26E-10	MethylPlus
123	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO3F1_HUMAN.H11MO.0.C	VATKSWWATGCAAW	4.69E-13	2.20E-10	MethylPlus
124	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PROP1_HUMAN.H11MO.0.D	TAATTRRATTA	8.02E-13	2.76E-10	MethylPlus
128	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXA10_HUMAN.H11MO.0.C	KATRRTTTATGW	1.71E-12	5.67E-10	MethylPlus
130	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXC11_HUMAN.H11MO.0.D	DTTTWAYGACY	1.53E-12	7.47E-10	MethylPlus
131	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF4_HUMAN.H11MO.0.A	RAAWGRGGAASTGAAASY	2.74E-12	1.14E-09	Little effect and MethylPlus
132	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ISX_HUMAN.H11MO.0.D	YTAATTRGHTWWWWWW	8.62E-12	1.72E-09	MethylPlus
133	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PROX1_HUMAN.H11MO.0.D	KGGKARGGCGKSKKGGG	2.08E-12	1.82E-09	MethylMinus
135	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXA13_HUMAN.H11MO.0.C	GDTTTTATTGG	5.89E-12	2.44E-09	MethylPlus and MethylMinus
136	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO3F4_HUMAN.H11MO.0.D	TMATTTGCATAATTWAW	1.48E-11	2.70E-09	MethylPlus
138	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NFAT5_HUMAN.H11MO.0.D	RRRGGAARRDN	2.82E-12	3.68E-09	MethylPlus
141	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LHX9_HUMAN.H11MO.0.D	YTAATTAGYNHWWWWWWD	2.41E-11	4.84E-09	MethylPlus
153	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	P5F1B_HUMAN.H11MO.0.D	DMATTTGCATATTYAW	1.30E-10	2.27E-08	POU5F1=MethylPlus
160	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MIXL1_HUMAN.H11MO.0.D	GBTAATTARAT	9.80E-11	3.70E-08	MethylPlus
161	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MEOX2_HUMAN.H11MO.0.D	TTAATKAYBWT	1.11E-10	3.73E-08	Little effect
168	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXC13_HUMAN.H11MO.0.D	DDTTTTACKAG	1.08E-10	5.51E-08	MethylPlus
169	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ERG_HUMAN.H11MO.0.A	RRRSAGGAAGYGG	7.74E-11	5.81E-08	MethylMinus
172	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	RAX2_HUMAN.H11MO.0.D	YTAATTARATWA	2.97E-10	8.67E-08	RAX=MethylPlus
174	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LHX4_HUMAN.H11MO.0.D	DNRYTAATTRN	2.05E-10	9.17E-08	MethylPlus
175	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PDX1_HUMAN.H11MO.1.A	ATTAATKAB	2.26E-10	1.02E-07	MethylPlus
176	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BARH2_HUMAN.H11MO.0.D	VYTAAWYGNTSHWWWWHR	3.86E-10	1.11E-07	BARHL2=MethylMinus
177	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD4_HUMAN.H11MO.0.D	TTAATTGW	2.28E-10	1.11E-07	MethylPlus
178	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXA2_HUMAN.H11MO.0.D	NGTAATTARBD	3.51E-10	1.19E-07	MethylPlus
179	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ZBT14_HUMAN.H11MO.0.C	GGAGCGCGC	3.78E-10	1.82E-07	MethylMinus
182	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	KLF3_HUMAN.H11MO.0.B	GGRGVDGGGCGGGGCYGGG	3.03E-10	2.10E-07	MethylPlus
183	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	UBIP1_HUMAN.H11MO.0.D	SCAGARA	2.42E-10	2.31E-07	MethylPlus
186	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PRRX2_HUMAN.H11MO.0.C	YTAATTR	1.09E-09	3.44E-07	MethylPlus
187	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NFAC1_HUMAN.H11MO.1.B	AATGGAAAAW	5.41E-10	3.89E-07	MethylPlus
190	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD9_HUMAN.H11MO.0.D	ARWTTNATKD	1.32E-09	5.40E-07	MethylPlus
192	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BATF_HUMAN.H11MO.0.A	RBTYTCRDWATGASTCAB	1.77E-09	5.67E-07	BATF3=inconclusive (different gene cards entry)
193	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	RX_HUMAN.H11MO.0.D	TTAATTRRYWA	2.95E-09	6.61E-07	MethylPlus
197	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MSX2_HUMAN.H11MO.0.D	TAATTDK	1.47E-09	8.28E-07	MethylPlus
198	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ETV7_HUMAN.H11MO.0.D	VAGGAAR	7.44E-10	9.43E-07	MethylMinus
199	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO3F2_HUMAN.H11MO.0.A	YYCATGCATWTYCATTM	2.63E-09	9.64E-07	MethylPlus
202	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXC8_HUMAN.H11MO.0.D	YTTSATTRATDSSC	3.58E-09	1.39E-06	MethylPlus
205	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	DLX4_HUMAN.H11MO.0.D	TAATTASNBHWWWWWD	7.88E-09	1.66E-06	MethylPlus
206	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXA5_HUMAN.H11MO.0.D	YTRATTAATG	4.89E-09	1.70E-06	MethylPlus
208	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NFAC4_HUMAN.H11MO.0.C	DGGAAAAYTW	2.38E-09	1.97E-06	MethylPlus

Rank	Motif DB	Motif ID	Consensus	p-value	adj p-value	Call (from Yin, et al)
209	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXC9_HUMAN.H11MO.0.C	WTTTATKGCH	4.18E-09	2.01E-06	MethylPlus
211	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	BARH1_HUMAN.H11MO.0.D	TAAWYGBYBHTAA	9.17E-09	2.39E-06	MethylMinus
212	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SHOX_HUMAN.H11MO.0.D	YTAATTRRYTWWWWWD	2.87E-08	3.59E-06	MethylPlus
215	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO2F3_HUMAN.H11MO.0.D	DTYATGCWWAWTTR	1.19E-08	4.27E-06	MethylPlus
218	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ONEC3_HUMAN.H11MO.0.D	TATYGATTTTTT	2.16E-08	4.66E-06	MethylMinus
220	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	E2F4_HUMAN.H11MO.0.A	SRGGGCGGGAARD	6.85E-09	5.01E-06	MethylMinus
221	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	DLX3_HUMAN.H11MO.0.C	GMTAATTRSW	2.25E-08	5.13E-06	MethylPlus
224	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO5F1_HUMAN.H11MO.1.A	WHATGCAVAT	1.41E-08	7.70E-06	MethylPlus
225	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SP2_HUMAN.H11MO.1.B	RGGRGGCGGGGC	1.30E-08	9.56E-06	MethylPlus
227	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	UNC4_HUMAN.H11MO.0.D	TAATTTAATTA	7.14E-08	1.10E-05	MethylPlus
229	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	DLX5_HUMAN.H11MO.0.D	TAATTAYAGGK	2.20E-08	1.17E-05	MethylPlus
230	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ETV2_HUMAN.H11MO.0.B	RRARRCAGGAARYRGS	3.53E-08	1.39E-05	MethylMinus
231	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GCR_HUMAN.H11MO.0.A	RGAACABMVTGTNCY	5.89E-08	1.48E-05	MethylPlus
233	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ETV4_HUMAN.H11MO.0.B	VRCAGGAARBR	1.54E-08	1.61E-05	MethylMinus
234	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	CDX1_HUMAN.H11MO.0.C	HTTTATGD	2.59E-08	1.95E-05	MethylPlus
237	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	EGR4_HUMAN.H11MO.0.D	GGSGGYRGGGM	4.48E-08	2.29E-05	MethylPlus
238	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	FLI1_HUMAN.H11MO.1.A	RRRCAGGAAGBRG	3.27E-08	2.42E-05	MethylMinus
242	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HME1_HUMAN.H11MO.0.D	VCTAATTRSYHAWWWWW	2.96E-07	4.00E-05	MethylPlus
243	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD8_HUMAN.H11MO.0.D	WAATWTGCAYN	6.04E-08	4.17E-05	MethylPlus
244	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	E2F1_HUMAN.H11MO.0.A	RRNKGGCGGGAARR	6.36E-08	4.53E-05	MethylMinus
245	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TF7L1_HUMAN.H11MO.0.B	DBBCTTTGAWSTB	6.52E-08	4.62E-05	MethylMinus
246	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HME2_HUMAN.H11MO.0.D	NNTRRTTARNB	6.96E-08	5.48E-05	MethylPlus
250	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ALX3_HUMAN.H11MO.0.D	TAATTARATTA	3.86E-07	7.99E-05	MethylPlus
252	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	E2F4_HUMAN.H11MO.1.A	GGCGGGAADTTBNR	1.92E-07	8.92E-05	MethylMinus
253	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ELF3_HUMAN.H11MO.0.A	RAABVAGGAAGTRR	2.16E-07	9.17E-05	MethylMinus
255	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ELF5_HUMAN.H11MO.0.A	VGRASSAGGAAGTRV	1.59E-07	1.01E-04	MethylMinus
260	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD3_HUMAN.H11MO.0.D	NGKTAATTAGK	3.68E-07	1.25E-04	MethylPlus
263	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	LBX2_HUMAN.H11MO.0.D	RYTAATTRGBY	9.26E-07	1.74E-04	MethylPlus
264	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	OSR2_HUMAN.H11MO.0.C	GBYSCWGCTNCTGYDG	2.74E-07	1.94E-04	Inconclusive
265	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	TCF7_HUMAN.H11MO.0.A	RMMMACATCAAAGGVA	3.65E-07	2.44E-04	MethylMinus
266	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SRBP2_HUMAN.H11MO.0.B	VSGTGGGGWGABG	2.92E-07	2.48E-04	MethylMinus
267	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NFAC2_HUMAN.H11MO.0.B	TGGAAAAWY	3.64E-07	2.70E-04	MethylPlus
268	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PBX1_HUMAN.H11MO.1.C	WAWTTGATTGATGDB	6.87E-07	2.88E-04	MethylPlus
269	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PTF1A_HUMAN.H11MO.0.B	SCAGCTGBYBBYBYYCHC	6.12E-07	3.56E-04	Little effect
270	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	CUX2_HUMAN.H11MO.0.D	WRATCAATRAA	5.94E-07	3.56E-04	Inconclusive
274	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GCR_HUMAN.H11MO.1.A	HRRGAACAGWVW	1.05E-06	5.24E-04	MethylPlus
278	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXB7_HUMAN.H11MO.0.C	TTGATTRATK	7.61E-06	6.85E-04	MethylPlus
279	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ETV3_HUMAN.H11MO.0.D	KWMRTTTGCACTTC	2.08E-06	7.13E-04	MethylMinus
280	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	XBP1_HUMAN.H11MO.0.D	GSKGACGTGGC	2.06E-06	7.24E-04	MethylMinus
282	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF9_HUMAN.H11MO.0.C	GAAAGCGAAAYT	5.66E-06	9.78E-04	MethylPlus and Little effect
285	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	PO2F2_HUMAN.H11MO.0.A	AYATGCAAATK	3.08E-06	1.19E-03	MethylPlus
286	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	IRF5_HUMAN.H11MO.0.D	KAAAGGRAARCMAAAWSWGA	2.47E-05	1.26E-03	Little effect
289	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	VSX2_HUMAN.H11MO.0.D	YTAATTAGCYTV	3.10E-05	1.45E-03	MethylPlus
292	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	SIX1_HUMAN.H11MO.0.A	TGWAAYHTGAKMCY	6.05E-06	1.73E-03	Little effect
293	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	ASCL2_HUMAN.H11MO.0.D	VCASCTGCT	3.57E-06	1.74E-03	MethylMinus
297	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	CLOCK_HUMAN.H11MO.0.C	CWGHCACGTGMVVM	4.53E-06	2.44E-03	MethylMinus and MethylPlus
302	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	COT1_HUMAN.H11MO.0.C	GRGGKSARAGGTCAGGG	6.41E-06	4.54E-03	MethylPlus
303	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	MYF6_HUMAN.H11MO.0.C	CASCTGC	7.29E-06	4.66E-03	MethylPlus
308	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	NFAC3_HUMAN.H11MO.0.B	TGGAAAAHY	1.10E-05	7.43E-03	MethylPlus
312	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	HXD13_HUMAN.H11MO.0.D	TTTATTAGRGA	1.50E-05	8.83E-03	MethylPlus
314	db/HUMAN/HOCOMOCOv11_full_HUMAN_mono_meme_format.meme	GATA6_HUMAN.H11MO.0.A	DVAGATAAGAWAA	5.68E-05	1.17E-02	Little effect

Color Key
MethylPlus
MethylMinus
Little effect
Inconclusive
Other

Appendix C: Chapter 4 extended data

C.1 Individual patient primary to metastasis DMR GO enrichment

C.1.1 Primary to metastasis hyperDMR enriched terms

	Trio 0	Trio 10	Trio 11	Trio 1	Trio 2	Trio 3	Trio 4	Trio 5	Trio 6	Trio 7	Trio 8	Trio 9
	response to muramyl dipeptide		epithelial-mesenchymal signaling involved in prostate gland development	brain development	somatic stem cell maintenance	embryonic organ morphogenesis	pattern specification process	pattern specification process		pattern specification process	embryonic skeletal system morphogenesis	metanephros development
			regulation of cell adhesion	forebrain development	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	embryonic organ development	embryonic morphogenesis	embryonic morphogenesis		regulation of transcription from RNA polymerase II promoter	embryonic organ morphogenesis	pattern specification process
			lung epithelium development	central nervous system development	stem cell maintenance	skeletal system morphogenesis	regionalization	embryonic organ morphogenesis		embryonic morphogenesis		embryonic morphogenesis
			lung morphogenesis	neurogenesis	negative regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	embryonic skeletal system morphogenesis	embryo development	regionalization		embryo development		regionalization
			respiratory system development	generation of neurons	regulation of BMP signaling pathway	cell differentiation in spinal cord	embryonic organ morphogenesis	embryonic organ development		regionalization		organ morphogenesis
			regulation of transcription from RNA polymerase II promoter	immune system process	nephron development	spinal cord association neuron differentiation	organ morphogenesis	skeletal system morphogenesis		embryonic organ morphogenesis		embryo development
			negative regulation of gene expression	negative regulation of macromolecule biosynthetic process	epithelial cell proliferation	embryonic skeletal system development	embryonic organ development	embryonic skeletal system morphogenesis		embryonic organ development		positive regulation of mesenchymal to epithelial transition involved in metanephros morphogenesis
			lung development	negative regulation of RNA metabolic process	limb bud formation	dorsal spinal cord development	positive regulation of transcription from RNA polymerase II promoter	anterior/posterior pattern specification		positive regulation of transcription from RNA polymerase II promoter		nephron morphogenesis
			regulation of anatomical structure morphogenesis	negative regulation of gene expression	positive regulation of establishment of protein localization to plasma membrane	eye morphogenesis	cell fate commitment	embryonic skeletal system development		embryonic skeletal system development		nephron epithelium morphogenesis
			respiratory tube development	negative regulation of macromolecule metabolic process	prostate gland epithelium morphogenesis	forebrain regionalization	epithelium development	tube morphogenesis		embryo development ending in birth or egg hatching		positive regulation of ureteric bud formation
			negative regulation of transcription, DNA-dependent	negative regulation of biosynthetic process	regulation of morphogenesis of a branching structure	spinal cord development	negative regulation of transcription, DNA-dependent	embryonic forelimb morphogenesis		cell fate commitment		kidney development

			negative regulation of RNA metabolic process	negative regulation of transcription, DNA-dependent	ureteric bud morphogenesis	cerebellar Purkinje cell-granule cell precursor cell signaling involved in regulation of granule cell precursor cell proliferation	negative regulation of transcription from RNA polymerase II promoter	cell fate commitment		negative regulation of transcription, DNA-dependent		cell fate commitment
			negative regulation of transcription from RNA polymerase II promoter	neuron projection morphogenesis	response to growth hormone stimulus	embryonic eye morphogenesis	morphogenesis of an epithelium	embryonic cranial skeleton morphogenesis		organ morphogenesis		metanephric nephron development
				negative regulation of cellular biosynthetic process	kidney epithelium development	determination of left/right symmetry	embryo development ending in birth or egg hatching	epithelial tube morphogenesis		negative regulation of RNA metabolic process		kidney epithelium development
				response to oxygen-containing compound	epithelial-mesenchymal cell signaling	cardiac chamber morphogenesis	neuron fate commitment	forelimb morphogenesis		chordate embryonic development		kidney morphogenesis
				negative regulation of nitrogen compound metabolic process	regulation of gliogenesis	ventricular cardiac muscle tissue development	tissue morphogenesis	forebrain regionalization		anterior/posterior pattern specification		nephron epithelium development
				cell morphogenesis involved in neuron differentiation	artery development	neuron fate commitment	chordate embryonic development	ear morphogenesis		skeletal system development		positive regulation of secretion
				cell morphogenesis involved in differentiation	negative regulation of osteoblast differentiation	camera-type eye morphogenesis	anterior/posterior pattern specification	central nervous system neuron differentiation		embryonic skeletal system morphogenesis		ureteric bud formation
				axonogenesis	negative regulation of gliogenesis	positive regulation of cerebellar granule cell precursor proliferation	epithelial tube morphogenesis	branching morphogenesis of an epithelial tube		neuron fate commitment		embryonic organ morphogenesis
				negative regulation of nucleobase-containing compound metabolic process	nephron epithelium development	regulation of cerebellar granule cell precursor proliferation	tube morphogenesis	dorsal/ventral pattern formation		negative regulation of transcription from RNA polymerase II promoter		skeletal system development
Total DMRs	**122**	**13**	**113**	**134**	**6207**	**2251**	**7311**	**2532**	**108**	**2886**	**53**	**1190**

C.1.2 Primary to metastasis hypoDMR enriched terms

	Trio 0	Trio 10	Trio 11	Trio 1	Trio 2	Trio 3	Trio 4	Trio 5	Trio 6	Trio 7	Trio 8	Trio 9
			embryonic organ morphogenesis	axon ensheathment	regulation of autophagy		loop of Henle development	negative regulation of gliogenesis		mitotic anaphase	embryonic skeletal system morphogenesis	regulation of insulin receptor signaling pathway
			morphogenesis of a polarized epithelium	myelination			nephron tubule development			actin filament bundle assembly	embryonic organ morphogenesis	negative regulation of insulin receptor signaling pathway
			embryonic eye morphogenesis	metanephric nephron tubule development			renal tubule development			regulation of RNA stability		specification of nephron tubule identity
			anterior/posterior pattern specification	metanephric tubule development			negative regulation of gliogenesis			positive regulation of Rac GTPase activity		nephron tubule formation
			intermediate filament organization	regulation of action potential in neuron			nephron epithelium development					nephron development
			cell differentiation involved in salivary gland development	metanephric epithelium development			negative regulation of glial cell differentiation					nephron epithelium morphogenesis
			neural precursor cell proliferation	positive regulation of gliogenesis			nephron tubule formation					nephron morphogenesis
			oligodendrocyte differentiation	nephron tubule development			forebrain neuron differentiation					metanephric nephron development
			tachykinin receptor signaling pathway	renal tubule development			O-glycan processing					kidney epithelium development
			excretion	loop of Henle development			specification of nephron tubule identity					
			metanephric nephron tubule development	compartment pattern specification			kidney epithelium development					
			metanephric tubule development	nephron epithelium development			negative regulation of glial cell proliferation					
			water transport	lung-associated mesenchyme development			glial cell development					
			fluid transport	response to peptide			neuron fate commitment					
				metanephros morphogenesis			nephron tubule morphogenesis					
				regulation of ARF protein signal transduction			forebrain generation of neurons					
				regulation of action potential			regulation of gliogenesis					
				endothelial cell development			positive regulation of protein sumoylation					
				kidney epithelium development			cell differentiation in spinal cord					
				regulation of gliogenesis			renal tubule morphogenesis					
Total DMRs	**27**	**40**	**1017**	**2728**	**1546**	**363**	**7790**	**3158**	**66**	**3366**	**272**	**1066**

C.2 Motifs enriched within shared DMRs

Motif database used in all cases:
db/HUMAN/HOCOMOCOv11_core_HUMAN_mono_meme_format.meme

C.2.1 Motifs enriched within hyperDMRs shared by at least 2 patients

rank	motif_ID	consensus	p-value	adj p-value	E-value	tests	FASTA max	pos	neg	PWM min	TP	%TP	FP	%FP
1	ZBT14_HUMAN.H11MO.0.C	GGAGCGCGC	2.93E-91	3.12E-88	1.26E-85	1066	3037	3037	1869	1.02	926	30.49	138	7.38
2	MECP2_HUMAN.H11MO.0.C	SCCGGRR	2.08E-63	4.27E-60	1.71E-57	2051	3037	3037	1869	1.33	1207	39.74	326	17.44
3	AP2B_HUMAN.H11MO.0.B	GCCYGVGGGS	7.84E-63	1.81E-59	7.26E-57	2304	3037	3037	1869	1.01	1702	56.04	593	31.73
4	E2F4_HUMAN.H11MO.0.A	SRGGGCGGGAARD	3.56E-57	6.23E-54	2.50E-51	1748	3037	3037	1869	1.02	1336	43.99	411	21.99
5	E2F1_HUMAN.H11MO.0.A	RRNKGGCGGGAARR	2.46E-51	3.95E-48	1.59E-45	1608	3037	3037	1869	1.02	1225	40.34	374	20.01
6	PATZ1_HUMAN.H11MO.0.C	GGGGMGGGGGMKGGRRVGGVRG	5.91E-49	1.86E-45	7.47E-43	3146	3037	3037	1869	2.38	1650	54.33	616	32.96
7	WT1_HUMAN.H11MO.0.C	RGGVGGGGGAGGRGGVGGRG	1.16E-46	2.72E-43	1.09E-40	2341	3037	3037	1869	1.16	1649	54.3	625	33.44
8	MBD2_HUMAN.H11MO.0.B	SSGKCCGGMGR	1.10E-44	1.14E-41	4.57E-39	1034	3037	3037	1869	1.04	820	27	203	10.86
9	SP2_HUMAN.H11MO.0.A	GGSSVGGGGGCGGGGCCDGSGS	1.15E-39	3.06E-36	1.23E-33	2660	3037	3037	1869	6.37	705	23.21	168	8.99
10	TFDP1_HUMAN.H11MO.0.C	RRRRGGCGGGAARN	1.85E-39	3.62E-36	1.45E-33	1956	3037	3037	1869	1.03	1420	46.76	525	28.09
11	SP3_HUMAN.H11MO.0.B	SGVVGGGGGCGGGGCBRGSS	2.83E-33	8.03E-30	3.23E-27	2837	3037	3037	1869	5.2	745	24.53	206	11.02
12	ZN467_HUMAN.H11MO.0.C	GGGGAGGGGRRGGRGRRGGRRR	1.39E-27	3.00E-24	1.21E-21	2159	3037	3037	1869	1.15	1501	49.42	630	33.71
13	ZBT7A_HUMAN.H11MO.0.A	VRGGGKCKY	1.59E-27	3.21E-24	1.29E-21	2027	3037	3037	1869	1.01	1435	47.25	592	31.67
14	E2F2_HUMAN.H11MO.0.B	GGCGCGAAAC	3.55E-26	1.69E-23	6.81E-21	478	3037	3037	1869	1.04	384	12.64	76	4.07
15	MAZ_HUMAN.H11MO.0.A	GGGGGAGGGGGGDGRGRRRGRRG	1.14E-26	3.75E-23	1.51E-20	3295	3037	3037	1869	6.8	1129	37.17	426	22.79
16	KLF15_HUMAN.H11MO.0.A	RGGGMGGRGVDGGGGGRGG	6.32E-24	1.60E-20	6.42E-18	2526	3037	3037	1869	5.68	901	29.67	320	17.12
17	E2F6_HUMAN.H11MO.0.A	DGGGCGGGARRVR	5.77E-24	1.60E-20	6.43E-18	2774	3037	3037	1869	2.28	1075	35.4	411	21.99
18	ZN281_HUMAN.H11MO.0.A	RGGWGGGGGAGGGGV	7.81E-23	1.33E-19	5.35E-17	1704	3037	3037	1869	1.06	1201	39.55	486	26
19	THAP1_HUMAN.H11MO.0.C	SGCCGCCATSTYGGSYBCGGGC	1.38E-20	1.81E-17	7.29E-15	1310	3037	3037	1869	1.08	943	31.05	359	19.21
20	ZFX_HUMAN.H11MO.0.A	SCAGGCCBCR	4.82E-18	9.73E-15	3.91E-12	2019	3037	3037	1869	1.97	878	28.91	339	18.14
21	SP4_HUMAN.H11MO.0.A	SRGVARGRGGCGGRGCHDRR	4.08E-18	1.22E-14	4.89E-12	2980	3037	3037	1869	3.01	1137	37.44	479	25.63
22	KLF3_HUMAN.H11MO.0.B	GGRGVDGGGCGGGGCYGGG	1.19E-17	2.51E-14	1.01E-11	2106	3037	3037	1869	1.96	1010	33.26	412	22.04
23	EGR1_HUMAN.H11MO.0.A	VDRDGCGKGGGYGGRRR	1.96E-17	4.57E-14	1.84E-11	2331	3037	3037	1869	2.59	858	28.25	332	17.76
24	FEV_HUMAN.H11MO.0.B	GCVGGAAGYG	7.72E-17	2.94E-13	1.18E-10	3809	3037	3037	1869	1.67	1457	47.97	672	35.96
25	ZIC1_HUMAN.H11MO.0.B	KGGGWGGTS	1.75E-16	3.98E-13	1.60E-10	2274	3037	3037	1869	1.03	1508	49.65	705	37.72
26	E2F5_HUMAN.H11MO.0.B	SGCGCSAAAH	2.19E-15	6.24E-13	2.51E-10	285	3037	3037	1869	1.47	212	6.98	40	2.14
27	KLF6_HUMAN.H11MO.0.A	GSRRGGGHGGGGMHGGGRV	3.47E-16	1.07E-12	4.31E-10	3085	3037	3037	1869	2.52	1215	40.01	536	28.68
28	SP1_HUMAN.H11MO.0.A	VGGGGGCGGGGCCKGGGGGGGG	2.61E-15	2.45E-12	9.84E-10	938	3037	3037	1869	3.99	401	13.2	118	6.31
29	CTCFL_HUMAN.H11MO.0.A	GCCDSYAGGGGGCGCYV	7.36E-15	5.03E-12	2.02E-09	684	3037	3037	1869	1.66	379	12.48	110	5.89
30	VEZF1_HUMAN.H11MO.0.C	GGRRRRGRRGGAGGGGGRGRRR	3.22E-15	1.08E-11	4.33E-09	3345	3037	3037	1869	8.37	1086	35.76	470	25.15
31	GABPA_HUMAN.H11MO.0.A	GVVRCCGGAAGTGV	5.25E-12	5.46E-09	2.20E-06	1040	3037	3037	1869	1.08	714	23.51	290	15.52
32	ZN263_HUMAN.H11MO.0.A	GGGAGGAGGRRGRGGRGGRR	2.39E-11	5.83E-08	2.34E-05	2438	3037	3037	1869	3.33	975	32.1	437	23.38
33	KAISO_HUMAN.H11MO.0.A	SARRYCTCGCGAGAV	2.82E-10	8.81E-08	3.54E-05	313	3037	3037	1869	1.15	223	7.34	60	3.21
34	EGR2_HUMAN.H11MO.0.A	GRGRRKGWGKGGGHGGRG	3.86E-11	1.01E-07	4.05E-05	2611	3037	3037	1869	1.04	1726	56.83	883	47.24
35	ERG_HUMAN.H11MO.0.A	RRRSAGGAAGYGG	1.99E-10	4.58E-07	1.84E-04	2294	3037	3037	1869	14.1	306	10.08	96	5.14
36	TAF1_HUMAN.H11MO.0.A	RARRWGGCGGMGGMGR	3.04E-10	9.67E-07	3.89E-04	3183	3037	3037	1869	8.11	418	13.76	150	8.03
37	ZN341_HUMAN.H11MO.0.C	GRGGRRRRGGRGRRGGAASAGC	1.54E-09	5.00E-06	2.01E-03	3248	3037	3037	1869	10	729	24	316	16.91
38	MZF1_HUMAN.H11MO.0.B	RRDRGGGAWTKRR	1.44E-08	3.14E-05	1.26E-02	2179	3037	3037	1869	1.08	1398	46.03	709	37.93
39	KLF12_HUMAN.H11MO.0.C	VGGGGCGGGGC	4.04E-08	5.32E-05	2.14E-02	1317	3037	3037	1869	4.39	340	11.2	124	6.63
40	E2F3_HUMAN.H11MO.0.A	DDGGCGGGAAA	2.12E-08	6.18E-05	2.48E-02	2918	3037	3037	1869	2.4	817	26.9	374	20.01
41	KLF1_HUMAN.H11MO.0.A	DGGGYGKGGCYGGG	8.03E-08	1.42E-04	5.71E-02	1768	3037	3037	1869	2.55	699	23.02	314	16.8
42	E2F7_HUMAN.H11MO.0.B	GDGGCGGGAARDR	5.39E-08	1.73E-04	6.95E-02	3211	3037	3037	1869	4.29	542	17.85	228	12.2
43	ZBT17_HUMAN.H11MO.0.A	SRRGGWGGGGGAGGGGMRR	2.26E-07	8.07E-04	3.24E-01	3568	3037	3037	1869	32.2	290	9.55	104	5.56
44	NRF1_HUMAN.H11MO.0.A	CWSTGCGCATGCGCRDS	8.62E-06	2.22E-03	8.93E-01	258	3037	3037	1869	1.08	190	6.26	65	3.48
45	ETS1_HUMAN.H11MO.0.A	RRRCMGGAAGTGG	1.34E-06	2.31E-03	9.29E-01	1728	3037	3037	1869	8.77	256	8.43	92	4.92
46	ETV1_HUMAN.H11MO.0.A	SRCCGGAAGYG	2.38E-06	3.11E-03	1.25E+00	1312	3037	3037	1869	1.12	827	27.23	400	21.4
47	ELK4_HUMAN.H11MO.0.A	SVCCGGAAGTGV	6.24E-06	4.74E-03	1.90E+00	761	3037	3037	1869	1.12	493	16.23	219	11.72
48	ELF2_HUMAN.H11MO.0.C	RAVCCGGAAGTGR	5.01E-06	7.65E-03	3.07E+00	1532	3037	3037	1869	1.88	650	21.4	304	16.27
49	FLI1_HUMAN.H11MO.0.A	SRRGGMAGGAAGGRRRGR	4.21E-06	1.01E-02	4.04E+00	2401	3037	3037	1869	7.71	664	21.86	311	16.64

C.2.2 Expression of genes with motifs enriched within shared hyperDMRs

Gene	log2FoldChange	padj
ZBTB14	0.118060993	0.95599
MECP2	0.316027308	0.87567
TFAP2B	-0.172245181	NA
E2F4	0.276267587	0.83256
E2F1	0.103573819	0.96708
PATZ1	-0.013734124	0.99251
WT1	-1.845265935	0.02135
MBD2	-0.484980898	0.73298
SP2	0.204616953	0.91908
TFDP1	0.232607286	0.91908
SP3	0.124292086	0.94351
ZNF467	0.371744374	0.89388
ZBTB7A	0.084485471	0.96538
E2F2	0.35605037	0.88922
MAZ	-0.210126485	0.91908
KLF15	-0.429696356	0.87975
E2F6	0.165770689	0.92786
ZNF281	-0.029071169	0.99053
THAP1	0.048361318	0.98806
ZFX	0.280304043	0.88102
SP4	0.253737837	0.92933
KLF3	0.574321442	0.65603
EGR1	-0.789025937	0.48755
FEV	-0.216804368	0.95106
ZIC1	2.050807286	0.00278
E2F5	0.326763286	0.90661
KLF6	0.003558394	0.99904
SP1	0.404129746	0.86308
CTCFL	-0.304861009	NA
VEZF1	0.335552436	0.88262
GABPA	-0.065963893	0.98071
ZNF263	0.323021046	0.82398
ZBTB33	0.28612073	0.90381
EGR2	-1.515284611	0.01138
ERG	-1.320705005	0.0833
TAF1	0.499651987	0.78264
ZNF341	0.123953694	0.94861
MZF1	0.49099283	0.59727
KLF12	-0.002151813	0.99947
E2F3	0.150292739	0.93833
KLF1	0.167378274	NA
E2F7	0.412971497	0.88913
ZBTB17	0.002803937	0.99904
NRF1	-0.153160912	0.93948
ETS1	-0.911383757	0.06488
ETV1	0.034686079	0.99071
ELK4	0.145804816	0.93833
ELF2	0.309617831	0.87942
FLI1	-0.806628685	0.43666

C.2.3 Ontology enrichment of genes with enriched motifs in shared hyperDMRs

All "Categories" are: GOTERM_BP_DIRECT

Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0006351~transcription, DNA-templated	32	65.306	4.82E-18	E2F1, E2F3, ZBTB33, ELF2, E2F4, ZNF467, E2F5, E2F6, E2F7, ZBTB14, ZBTB17, ELK4, THAP1, PATZ1, TFDP1, ZNF263, ZNF281, KLF6, ZBTB7A, ERG, EGR2, KLF12, ZFX, MECP2, MBD2, ZNF341, SP2, SP3, SP4, MZF1, KLF1, KLF3	49	1955	16792	5.609311551	1.66E-15	1.66E-15	6.57E-15
GO:0045944~positive regulation of transcription from RNA polymerase II promoter	22	44.898	5.01E-14	E2F1, EGR1, TAF1, E2F4, EGR2, KLF12, E2F5, E2F7, GABPA, CTCFL, KLF15, ZIC1, WT1, NRF1, FLI1, SP1, ETS1, SP3, MZF1, TFAP2B, ETV1, TFDP1	49	981	16792	7.685285735	1.72E-11	8.61E-12	6.83E-11
GO:0006366~transcription from RNA polymerase II promoter	17	34.694	4.18E-13	EGR1, TAF1, E2F4, EGR2, GABPA, CTCFL, KLF15, ZIC1, WT1, FEV, NRF1, FLI1, ETS1, MZF1, TFAP2B, ETV1, VEZF1	49	513	16792	11.35632733	1.44E-10	4.79E-11	5.70E-10
GO:0045893~positive regulation of transcription, DNA-templated	17	34.694	4.44E-13	E2F1, ZNF281, EGR1, KLF6, E2F3, ELF2, EGR2, MECP2, CTCFL, ZBTB17, ZIC1, WT1, SP1, ETS1, SP3, TFAP2B, KLF1	49	515	16792	11.31222508	1.53E-10	3.81E-11	6.05E-10
GO:0000122~negative regulation of transcription from RNA polymerase II promoter	16	32.653	7.69E-10	ZNF281, EGR1, E2F1, ZBTB7A, KLF12, E2F6, E2F7, GABPA, MECP2, ZBTB14, MBD2, WT1, ELK4, SP3, MZF1, TFAP2B	49	720	16792	7.615419501	2.65E-07	5.29E-08	1.05E-06
GO:0045892~negative regulation of transcription, DNA-templated	13	26.531	1.09E-08	ZNF263, E2F1, ZNF281, ZBTB7A, ZBTB33, ELF2, SP3, TFAP2B, MECP2, ZBTB14, PATZ1, MBD2, WT1	49	499	16792	8.92789661	3.76E-06	6.27E-07	1.49E-05
GO:0006357~regulation of transcription from RNA polymerase II promoter	11	22.449	3.77E-07	ERG, NRF1, ELF2, FLI1, KLF12, SP2, SP4, VEZF1, WT1, FEV, TFDP1	49	441	16792	8.547919848	1.30E-04	1.85E-05	5.15E-04
GO:0006355~regulation of transcription, DNA-templated	17	34.694	2.24E-06	ZNF263, E2F1, E2F2, EGR2, ZNF467, ZFX, WT1, ZNF341, MAZ, SP1, SP3, MZF1, TFAP2B, THAP1, PATZ1, KLF1, KLF3	49	1504	16792	3.87353452	7.70E-04	9.63E-05	0.00305
GO:0070345~negative regulation of fat cell proliferation	3	6.1224	7.96E-05	E2F1, E2F3, TFDP1	49	5	16792	205.6163265	0.027003717	0.003037	0.1085
GO:0030154~cell differentiation	8	16.327	3.16E-04	ZBTB7A, ERG, ELF2, FLI1, ELK4, GABPA, ZIC1, FEV	49	462	16792	5.934093118	0.103130745	0.0108255	0.43071
GO:1901653~cellular response to peptide	3	6.1224	4.33E-04	KLF15, KLF1, KLF3	49	11	16792	93.4619666	0.138386154	0.0134495	0.58893
GO:0006977~DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest	4	8.1633	7.36E-04	E2F1, E2F4, E2F7, TFDP1	49	62	16792	22.10928242	0.223875112	0.0208987	1.00001
GO:0030217~T cell differentiation	3	6.1224	0.00331	EGR1, SP3, PATZ1	49	30	16792	34.26938776	0.68005249	0.083929	4.41859
GO:0007049~cell cycle	5	10.204	0.00338	E2F2, TAF1, E2F3, CTCFL, THAP1	49	217	16792	7.896172294	0.688313418	0.0798958	4.51769
GO:0071930~negative regulation of transcription involved in G1/S transition of mitotic cell cycle	2	4.0816	0.00855	E2F1, E2F7	49	3	16792	228.462585	0.947888868	0.1787752	11.0556
GO:0006367~transcription initiation from RNA polymerase II promoter	4	8.1633	0.00934	E2F2, TAF1, E2F3, MAZ	49	152	16792	9.018259936	0.960349358	0.1826828	12.0142
GO:1990086~lens fiber cell apoptotic process	2	4.0816	0.01139	E2F1, E2F2	49	4	16792	171.3469388	0.980537902	0.206834	14.4626
GO:0072112~glomerular visceral epithelial cell differentiation	2	4.0816	0.01984	KLF15, WT1	49	7	16792	97.91253644	0.998986523	0.318201	23.9213
GO:0009887~organ morphogenesis	3	6.1224	0.02845	FLI1, E2F4, E2F5	49	92	16792	11.17480035	0.999951267	0.4070154	32.5477
GO:0042551~neuron maturation	2	4.0816	0.03101	MECP2, FEV	49	11	16792	62.30797774	0.999980308	0.4182794	34.9285
GO:0008284~positive regulation of cell proliferation	5	10.204	0.04367	TAF1, E2F3, ETS1, TFAP2B, MECP2	49	466	16792	3.676972935	0.999999787	0.5188252	45.6211
GO:0006349~regulation of gene expression by genetic imprinting	2	4.0816	0.04479	MECP2, CTCFL	49	16	16792	42.83673469	0.999999857	0.5115364	46.4785
GO:0016571~histone methylation	2	4.0816	0.04752	MECP2, CTCFL	49	17	16792	40.31692677	0.999999947	0.5172199	48.5302
GO:0044849~estrous cycle	2	4.0816	0.04752	EGR1, ETS1	49	17	16792	40.31692677	0.999999947	0.5172199	48.5302
GO:0007275~multicellular organism development	5	10.204	0.06107	ZBTB7A, ERG, NRF1, ZFX, KLF3	49	521	16792	3.288808806	1	0.5947035	57.6648
GO:0000083~regulation of transcription involved in G1/S transition of mitotic cell cycle	2	4.0816	0.06376	E2F4, E2F6	49	23	16792	29.79946761	1	0.5960845	59.2922
GO:0072332~intrinsic apoptotic signaling pathway by p53 class mediator	2	4.0816	0.07443	E2F1, E2F2	49	27	16792	25.38473167	1	0.6406295	65.1871
GO:1900740~positive regulation of protein insertion into mitochondrial membrane involved in apoptotic signaling pathway	2	4.0816	0.08236	E2F1, TFDP1	49	30	16792	22.8462585	1	0.6654924	69.0417
GO:0001666~response to hypoxia	3	6.1224	0.08667	EGR1, ETS1, MECP2	49	172	16792	5.977218794	1	0.6716869	70.9662
GO:0016573~histone acetylation	2	4.0816	0.09283	TAF1, MECP2	49	34	16792	20.15846339	1	0.685148	73.5265

C.2.4 Pathway enrichment of genes with enriched motifs in shared hyperDMRs

All "Categories" are: KEGG_PATHWAY

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
KEGG_PATHWAY	hsa05202:Transcriptional misregulation in cancer	9	18.3673	6.15E-09	ERG, FLI1, SP1, ELK4, ETV1, ZBTB17, WT1, FEV, KLF3	20	167	6879	18.53622754	1.54E-07	1.54E-07	4.99E-06
KEGG_PATHWAY	hsa04110:Cell cycle	7	14.2857	6.81E-07	E2F1, E2F2, E2F3, E2F4, E2F5, ZBTB17, TFDP1	20	124	6879	19.41653226	1.70E-05	8.51E-06	5.52E-04
KEGG_PATHWAY	hsa05166:HTLV-I infection	7	14.2857	4.32E-05	E2F1, EGR1, E2F2, E2F3, EGR2, ELK4, ETS1	20	254	6879	9.478937008	0.001079904	3.60E-04	0.03506
KEGG_PATHWAY	hsa04350:TGF-beta signaling pathway	4	8.16327	0.00148	E2F4, SP1, E2F5, TFDP1	20	84	6879	16.37857143	0.036301548	0.009201614	1.19283
KEGG_PATHWAY	hsa05219:Bladder cancer	3	6.12245	0.00556	E2F1, E2F2, E2F3	20	41	6879	25.16707317	0.130079842	0.027485962	4.4217
KEGG_PATHWAY	hsa05161:Hepatitis B	4	8.16327	0.00694	E2F1, E2F2, E2F3, EGR2	20	145	6879	9.488275862	0.159776692	0.028597715	5.49302
KEGG_PATHWAY	hsa05223:Non-small cell lung cancer	3	6.12245	0.01019	E2F1, E2F2, E2F3	20	56	6879	18.42589286	0.2258135	0.03590285	7.97047
KEGG_PATHWAY	hsa05214:Glioma	3	6.12245	0.01356	E2F1, E2F2, E2F3	20	65	6879	15.87461538	0.289086826	0.041753878	10.482
KEGG_PATHWAY	hsa05212:Pancreatic cancer	3	6.12245	0.01356	E2F1, E2F2, E2F3	20	65	6879	15.87461538	0.289086826	0.041753878	10.482
KEGG_PATHWAY	hsa05218:Melanoma	3	6.12245	0.01604	E2F1, E2F2, E2F3	20	71	6879	14.53309859	0.33246363	0.043913437	12.2924
KEGG_PATHWAY	hsa05220:Chronic myeloid leukemia	3	6.12245	0.01647	E2F1, E2F2, E2F3	20	72	6879	14.33125	0.339738524	0.040662114	12.6038
KEGG_PATHWAY	hsa05222:Small cell lung cancer	3	6.12245	0.02252	E2F1, E2F2, E2F3	20	85	6879	12.13941176	0.434104766	0.050442086	16.8704
KEGG_PATHWAY	hsa05215:Prostate cancer	3	6.12245	0.02403	E2F1, E2F2, E2F3	20	88	6879	11.72556818	0.455546584	0.04940237	17.906
KEGG_PATHWAY	hsa05200:Pathways in cancer	4	8.16327	0.09092	E2F1, E2F2, E2F3, ETS1	20	393	6879	3.500763359	0.907722643	0.167485254	53.853

C.2.5 Motifs enriched within hypoDMRs shared by at least 2 patients

rank	motif_ID	consensus	p-value	adj_p-value	E-value	tests	FASTA max	pos	neg	PWM_min	TP	%TP	FP	%FP
1	ZN770_HUMAN.H11MO.0.C	GGGAGGCYGAGGBRGGAGGATC	5.72E-127	1.37E-123	5.49E-121	2388	1869	1869	3037	1.98E+03	5328.52	131	4.31	
2	ZN121_HUMAN.H11MO.0.C	CTGGGCAACABAGYRAGACC	1.29E-104	9.51E-102	3.82E-99	736	1869	1869	3037	2.37	527	28.2	172	5.66
3	SRY_HUMAN.H11MO.0.B	YWTTGTTTH	1.83E-91	3.93E-88	1.58E-85	2147	1869	1869	3037	5.02	711	38.04	3912.87	
4	PAX5_HUMAN.H11MO.0.A	SCSKGGGCARCVRAGCGWGAC	3.03E-81	3.91E-78	1.57E-75	1289	1869	1869	3037	17.3	411	21.99	129	4.25
5	FOSB_HUMAN.H11MO.0.A	RTGAGTCAB	8.69E-75	8.28E-72	3.33E-69	952	1869	1869	3037	1.01	612	32.74	3411.2	
6	FOXK1_HUMAN.H11MO.0.A	TGTTTMCHTT	3.24E-74	7.29E-71	2.93E-68	2250	1869	1869	3037	1.04	1154	61.74	1068	35.17
7	MEF2B_HUMAN.H11MO.0.A	DKKKCTATTTTTRG	1.07E-73	1.83E-70	7.34E-68	1709	1869	1869	3037	8.46	589	31.51	3110.5	
8	ARI5B_HUMAN.H11MO.0.C	DBYKDGTATTSKR	2.96E-69	3.59E-66	1.44E-63	1211	1869	1869	3037	2.16	520	27.82	268.63	
9	FOXJ2_HUMAN.H11MO.0.C	TGTTTRTTTW	3.93E-68	5.14E-65	2.07E-62	1306	1869	1869	3037	1.44	736	39.38	5116.83	
10	FOXM1_HUMAN.H11MO.0.A	TGTTTRCTYWKB	1.91E-67	3.68E-64	1.48E-61	1927	1869	1869	3037	1.04	1011	54.09	8829.21	
11	FOXO1_HUMAN.H11MO.0.A	KBBYTGTTTHCN	1.62E-66	4.07E-63	1.64E-60	2507	1869	1869	3037	1.44	1075	57.52	9832.53	
12	SOX5_HUMAN.H11MO.0.C	YATTGTTW	7.67E-66	1.11E-62	4.47E-60	1449	1869	1869	3037	1.01	820	43.87	6220.71	
13	ANDR_HUMAN.H11MO.0.A	TGTTCTTKKYTGTTTRYW	8.27E-63	1.36E-59	5.46E-57	1644	1869	1869	3037	1.17	882	47.19	728	23.97
14	NANOG_HUMAN.H11MO.0.A	BYYWTTGWNATGCAAAT	1.23E-62	1.91E-59	7.68E-57	1556	1869	1869	3037	1.06	858	45.91	696	22.92
15	FOXP1_HUMAN.H11MO.0.A	TBTGTTTMY	1.68E-62	4.00E-59	1.61E-56	2385	1869	1869	3037	3.66	745	39.86	5418.04	
16	MAF_HUMAN.H11MO.0.A	VAAWBTGCTGASTHWGCMD	6.98E-62	9.60E-59	3.86E-56	1376	1869	1869	3037	1.09	770	41.2	583	19.2
17	BATF3_HUMAN.H11MO.0.B	BRSTTTCADTWTGASWM	1.33E-61	2.05E-58	8.22E-56	1536	1869	1869	3037	1.08	843	45.1	681	22.42
18	RXRB_HUMAN.H11MO.0.C	TSAGGTCASR	5.67E-61	9.03E-58	3.63E-55	1593	1869	1869	3037	1.12	825	44.14	6621.73	
19	NR2E3_HUMAN.H11MO.0.C	AAAGTCAAAAGTCA	9.03E-61	1.23E-57	4.93E-55	1359	1869	1869	3037	1.21	744	39.81	5518.27	
20	RARA_HUMAN.H11MO.0.A	GGRRGRTSMRRAGKTCAAGG	3.22E-60	4.77E-57	1.92E-54	1483	1869	1869	3037	13.4	351	18.78	130	4.28
21	FOXJ3_HUMAN.H11MO.0.A	TTGTTTATKKTTW	8.56E-60	8.10E-57	3.26E-54	946	1869	1869	3037	15	440	23.54	217.01	
22	MEF2A_HUMAN.H11MO.0.A	DKCTATTTTTRGM	9.48E-60	1.45E-56	5.81E-54	1525	1869	1869	3037	12.1	452	24.18	227.41	
23	SOX17_HUMAN.H11MO.0.C	SCCATTGTKBY	1.10E-59	2.04E-56	8.18E-54	1851	1869	1869	3037	1.04	961	51.42	8528.22	
24	FEZF1_HUMAN.H11MO.0.C	DYTGYYCTTTTH	1.46E-57	3.41E-54	1.37E-51	2341	1869	1869	3037	1.88	896	47.94	7725.55	
25	COT2_HUMAN.H11MO.0.A	CARAGGTCARRGD	6.52E-57	1.06E-53	4.24E-51	1618	1869	1869	3037	2.94	538	28.79	3210.73	
26	MEF2D_HUMAN.H11MO.0.A	DKCTATTTTTAG	2.32E-55	2.88E-52	1.16E-49	1242	1869	1869	3037	49.1	290	15.52	923.03	
27	MEF2C_HUMAN.H11MO.0.A	DKCTATTTTTRGM	1.84E-54	2.83E-51	1.14E-48	1535	1869	1869	3037	2.93	647	34.62	4615.28	
28	SRBP1_HUMAN.H11MO.0.A	DGTSRGGTGAB	5.14E-54	9.66E-51	3.88E-48	1882	1869	1869	3037	1.17	871	46.6	7625.06	
29	Z354A_HUMAN.H11MO.0.C	RTTTARWYCATTTACATTTAATGT	1.26E-53	1.62E-50	6.50E-48	1280	1869	1869	3037	1.94	6836.6	514	16.92	
30	LEF1_HUMAN.H11MO.0.A	BCCTTTSWWNTBYH	1.21E-52	3.47E-49	1.39E-46	2856	1869	1869	3037	2.05	1024	54.79	9932.7	
31	SOX2_HUMAN.H11MO.0.A	BBCCTTTGTYYYB	5.77E-52	1.41E-48	5.65E-46	2436	1869	1869	3037	4.2	757	40.5	6120.32	
32	NR1D1_HUMAN.H11MO.0.B	RRRNMASTGGGTCAGDRDR	1.23E-51	2.60E-48	1.04E-45	2116	1869	1869	3037	1.48	859	45.96	758	24.96
33	MAFB_HUMAN.H11MO.0.B	TGCTGASTYAD	5.89E-51	1.02E-47	4.10E-45	1732	1869	1869	3037	1.01	904	48.37	8227.2	
34	MAFF_HUMAN.H11MO.0.B	TGCTGASTCAGCABTTTT	4.44E-50	5.26E-47	2.12E-44	1185	1869	1869	3037	1.1	665	35.58	508	16.73
35	SOX3_HUMAN.H11MO.0.B	BCCWTTGTBYB	4.42E-47	8.86E-44	3.56E-41	2006	1869	1869	3037	1.29	936	50.08	8929.54	
36	SOX4_HUMAN.H11MO.0.B	BYCTTTGTYYYB	1.02E-46	2.48E-43	9.95E-41	2437	1869	1869	3037	4.6	687	36.76	5518.24	
37	NKX21_HUMAN.H11MO.0.A	BTKGAGTGSY	1.73E-46	3.23E-43	1.30E-40	1871	1869	1869	3037	2.51	549	29.37	3812.64	
38	ZFP28_HUMAN.H11MO.0.C	TTCTATTTCTTCTTGWGTCA	1.40E-45	1.26E-42	5.08E-40	904	1869	1869	3037	4.5	432	23.11	256	8.43
39	TBX21_HUMAN.H11MO.0.A	GDGDRGGTGTGRV	4.73E-46	1.46E-42	5.85E-40	3078	1869	1869	3037	1.02	1398	74.8	1665	54.82
40	FOS_HUMAN.H11MO.0.A	VTGAGTCAB	2.56E-45	1.68E-42	6.74E-40	654	1869	1869	3037	1.01	415	22.2	237.87	
41	PIT1_HUMAN.H11MO.0.C	YTMATGAATATWY	5.77E-45	7.78E-42	3.13E-39	1348	1869	1869	3037	1.09	710	37.99	5919.56	
42	TCF7_HUMAN.H11MO.0.A	RMMMACATCAAAGGVA	9.35E-45	1.50E-41	6.04E-39	1607	1869	1869	3037	1.05	832	44.52	762	25.09
43	ZN350_HUMAN.H11MO.0.C	CAGYYYTTTTATDWCCHR	9.91E-45	2.20E-41	8.85E-39	2220	1869	1869	3037	1.16	1052	56.29	1089	35.86
44	IRF1_HUMAN.H11MO.0.A	RAAANWGAAASTGAAASYRR	3.04E-43	3.85E-40	1.55E-37	1268	1869	1869	3037	111	311	16.64	143	4.71
45	AIRE_HUMAN.H11MO.0.C	HTTGGWWWWAWTGGDTWH	9.12E-43	1.09E-39	4.37E-37	1192	1869	1869	3037	1.05	656	35.1	536	17.65

rank	motif_ID	consensus	p-value	adj_p-value	E-value	tests	FASTA max	pos	neg	PWM_min	TP	%TP	FP	%FP
46	IKZF1_HUMAN.H11MO.0.C	BTGGGARR	5.23E-43	1.62E-39	6.51E-37	3096	1869	1869	3037	3.4	674	36.06	5518.37	
47	STAT2_HUMAN.H11MO.0.A	RRGRAAAHGAAACTGAAAV	8.15E-42	1.13E-38	4.55E-36	1387	1869	1869	3037	39.9	378	20.22	214	7.05
48	ZN554_HUMAN.H11MO.0.C	GCWGAGYCANGTRGDKKRCT	1.57E-41	2.30E-38	9.23E-36	1463	1869	1869	3037	1.2	706	37.77	609	20.05
49	SMAD2_HUMAN.H11MO.0.A	TGTCTGNCWCCT	4.19E-41	8.55E-38	3.44E-35	2039	1869	1869	3037	1.12	943	50.45	9431.22	
50	ZN394_HUMAN.H11MO.0.C	NRARWRGAAWNGAMWGRAAK	4.85E-41	1.11E-37	4.44E-35	2277	1869	1869	3037	21.3	410	21.94	251	8.26
51	GFI1_HUMAN.H11MO.0.C	GCWSTGATTT	7.45E-40	9.31E-37	3.74E-34	1250	1869	1869	3037	1.02	672	35.96	5718.9	
52	FOXA1_HUMAN.H11MO.0.A	TGTTTACWYWGB	1.26E-39	1.27E-36	5.12E-34	1010	1869	1869	3037	1.03	568	30.39	4414.55	
53	SMAD4_HUMAN.H11MO.0.B	CWGTCTGDCACCT	1.38E-39	2.05E-36	8.24E-34	1488	1869	1869	3037	1.04	766	40.98	7023.15	
54	BACH1_HUMAN.H11MO.0.A	TGCTGAGTCABSS	2.39E-39	3.17E-36	1.27E-33	1326	1869	1869	3037	1.4	601	32.16	4815.97	
55	IRF2_HUMAN.H11MO.0.A	RAAAVHGAAAGTGAAASTRV	6.89E-39	6.68E-36	2.69E-33	969	1869	1869	3037	2.49	476	25.47	335	11.03
56	RORA_HUMAN.H11MO.0.C	AWSTRGGTCARRG	1.07E-38	2.24E-35	9.01E-33	2103	1869	1869	3037	1.68	710	37.99	6320.81	
57	FOXA3_HUMAN.H11MO.0.B	CHTGTTTACWTWG	1.32E-37	1.33E-34	5.33E-32	1007	1869	1869	3037	1.06	555	29.7	4314.39	
58	CDX1_HUMAN.H11MO.0.C	HTTTATGD	9.46E-38	1.64E-34	6.58E-32	1730	1869	1869	3037	1.01	868	46.44	8628.35	
59	FOXO4_HUMAN.H11MO.0.C	TTGTTTAYK	2.23E-36	2.62E-33	1.05E-30	1178	1869	1869	3037	1.02	633	33.87	5417.91	
60	ZN418_HUMAN.H11MO.0.C	TGCTTYTRGCYTCTKVY	2.19E-36	3.29E-33	1.32E-30	1504	1869	1869	3037	1.4	673	36.01	597	19.66
61	FOXA2_HUMAN.H11MO.0.A	TGTTTACWYWGB	7.42E-36	6.73E-33	2.70E-30	907	1869	1869	3037	2.02	418	22.36	289.32	
62	NKX31_HUMAN.H11MO.0.C	NTWAAGTGBTTD	6.04E-35	7.21E-32	2.90E-29	1193	1869	1869	3037	1.02	636	34.03	5518.34	
63	MITF_HUMAN.H11MO.0.A	TCACGTGACC	5.48E-34	3.31E-31	1.33E-28	604	1869	1869	3037	1.02	368	19.69	237.77	
64	ZNF85_HUMAN.H11MO.0.C	AWAGAARWCTGMWGKAATCT	1.78E-33	1.18E-30	4.72E-28	659	1869	1869	3037	1.2	392	20.97	265	8.73
65	BATF_HUMAN.H11MO.0.A	RBTYTCRDWATGASTCAB	7.67E-33	7.10E-30	2.85E-27	925	1869	1869	3037	1.06	513	27.45	412	13.57
66	TBP_HUMAN.H11MO.0.A	MTATAAAARS	1.19E-32	1.74E-29	6.98E-27	1460	1869	1869	3037	1.04	731	39.11	7023.11	
67	SMAD3_HUMAN.H11MO.0.B	SYCTSHCWSCWS	7.97E-33	3.07E-29	1.23E-26	3854	1869	1869	3037	1.3	1459	78.06	1883	62
68	PRDM6_HUMAN.H11MO.0.C	RRRARGAAAAAAA	2.07E-32	5.23E-29	2.10E-26	2534	1869	1869	3037	12.2	560	29.96	4715.61	
69	ZNF41_HUMAN.H11MO.0.C	MACAAGGRARTMARVCACCATGAG	6.51E-31	4.55E-28	1.83E-25	700	1869	1869	3037	1.39	4021.46	288	9.48	
70	FOXP2_HUMAN.H11MO.0.C	BTGTTTACH	7.38E-31	1.05E-27	4.24E-25	1427	1869	1869	3037	1.06	708	37.88	6822.49	
71	FOXO3_HUMAN.H11MO.0.B	BYTGTTTACH	7.06E-30	8.78E-27	3.53E-24	1243	1869	1869	3037	1.1	624	33.39	5718.93	
72	SMCA1_HUMAN.H11MO.0.C	CMARGAAGMATD	6.57E-30	1.35E-26	5.44E-24	2062	1869	1869	3037	2.81	604	32.32	5418.04	
73	NFAC1_HUMAN.H11MO.0.B	AWGGAAARWVWGAMW	3.48E-29	9.31E-26	3.74E-23	2673	1869	1869	3037	4.55	676	36.17	6521.44	
74	FOXC1_HUMAN.H11MO.0.C	CDWTGTTTACTTWRG	3.66E-28	1.43E-25	5.76E-23	392	1869	1869	3037	1.08	252	13.48	134.58	
75	IRF3_HUMAN.H11MO.0.B	RAAARGGAAAVDGAAASDGA	6.57E-29	1.46E-25	5.88E-23	2228	1869	1869	3037	841	227	12.15	112	3.69
76	CEBPA_HUMAN.H11MO.0.A	DRTTGTGCAAYV	2.94E-27	2.48E-24	9.98E-22	845	1869	1869	3037	1.05	457	24.45	3712.38	
77	IRF7_HUMAN.H11MO.0.C	GAAASYGAAA	2.21E-27	4.55E-24	1.83E-21	2056	1869	1869	3037	2.03	778	41.63	8126.67	
78	TF7L1_HUMAN.H11MO.0.B	DBBCTTTGAWSTB	3.56E-27	5.70E-24	2.29E-21	1602	1869	1869	3037	1.02	783	41.89	8126.97	
79	ZN384_HUMAN.H11MO.0.C	GSAAAAAVYMKS	1.13E-26	1.83E-23	7.34E-21	1621	1869	1869	3037	1.02	789	42.22	8327.4	
80	JUNB_HUMAN.H11MO.0.A	DVTGAGTCABH	6.33E-26	3.69E-23	1.48E-20	582	1869	1869	3037	1.97	251	13.43	144.84	
81	NR4A1_HUMAN.H11MO.0.A	AAAGGKCAV	5.11E-26	1.20E-22	4.81E-20	2339	1869	1869	3037	1.03	1048	56.07	1235	40.67
82	ZFP82_HUMAN.H11MO.0.C	TTCCADTTCAYTAATTYTCTCTTC	7.68E-25	9.37E-22	3.77E-19	1220	1869	1869	3037	1.47	6133.01	603	19.86	
83	ZEB1_HUMAN.H11MO.0.A	BVCAGGTGWG	7.27E-25	1.63E-21	6.53E-19	2237	1869	1869	3037	1.01	1026	54.9	1211	39.87
84	ERR3_HUMAN.H11MO.0.B	TCAAGGTCA	3.74E-24	2.96E-21	1.19E-18	792	1869	1869	3037	1.16	400	21.4	3210.7	
85	GATA3_HUMAN.H11MO.0.A	RVAGATAAVAD	2.74E-24	5.21E-21	2.09E-18	1902	1869	1869	3037	1.02	891	47.67	1006	33.12
86	CEBPD_HUMAN.H11MO.0.C	VTTGCRCMAYC	8.13E-24	8.22E-21	3.30E-18	1011	1869	1869	3037	1.17	470	25.15	4113.66	
87	PBX2_HUMAN.H11MO.0.C	KRAWKTGATTGATKS	9.16E-24	9.40E-21	3.78E-18	1026	1869	1869	3037	1.18	496	26.54	4414.78	
88	THA_HUMAN.H11MO.0.C	RGGTCABYTSAGGTCAK	9.50E-24	1.25E-20	5.03E-18	1317	1869	1869	3037	2.85	284	15.2	192	6.32
89	PO2F1_HUMAN.H11MO.0.C	AWKBRWATGSAAWYAW	1.36E-23	1.43E-20	5.75E-18	1050	1869	1869	3037	1.05	539	28.84	507	16.69
90	VDR_HUMAN.H11MO.0.A	RRGGTCABYGRGKTCA	1.50E-23	1.59E-20	6.41E-18	1060	1869	1869	3037	53.3	77	4.12	9	0.3
91	TGIF1_HUMAN.H11MO.0.A	TGACAGS	2.80E-23	5.63E-20	2.26E-17	2014	1869	1869	3037	1.34	694	37.13	724	23.84
92	NFIA_HUMAN.H11MO.0.C	TTGGCHNRNWGCCAR	5.52E-23	6.03E-20	2.43E-17	1092	1869	1869	3037	1.41	415	22.2	3511.56	
93	JUND_HUMAN.H11MO.0.A	RRTGAGTCAYY	5.59E-22	2.71E-19	1.09E-16	485	1869	1869	3037	1.02	284	15.2	206.62	
94	JUN_HUMAN.H11MO.0.A	DRTGAGTCAYH	8.77E-22	3.71E-19	1.49E-16	423	1869	1869	3037	1.1	244	13.06	155.2	
95	IRF8_HUMAN.H11MO.0.B	RAAARRGGAASTGAAASTDR	3.61E-22	4.03E-19	1.62E-16	1115	1869	1869	3037	2.75	436	23.33	383	12.61
96	RARG_HUMAN.H11MO.0.B	RGGTCASVRTGVCMTGRV	5.13E-21	7.37E-18	2.96E-15	1438	1869	1869	3037	1.04	693	37.08	743	24.46
97	CEBPB_HUMAN.H11MO.0.A	DRTTRYGCAATM	1.31E-20	8.07E-18	3.24E-15	617	1869	1869	3037	1.1	334	17.87	268.79	
98	HNF1A_HUMAN.H11MO.0.C	DGTTAATKATTAACH	2.08E-20	1.90E-17	7.65E-15	916	1869	1869	3037	1.1	467	24.99	4314.36	
99	CLOCK_HUMAN.H11MO.0.C	CWGHCACGTGMVVM	2.13E-20	3.40E-17	1.37E-14	1597	1869	1869	3037	1.02	754	40.34	8327.59	
100	NF2L1_HUMAN.H11MO.0.C	HGTCATN	1.36E-19	1.40E-16	5.64E-14	1034	1869	1869	3037	1.05	499	26.7	485	15.97
101	LHX2_HUMAN.H11MO.0.A	KBYTAATTRGYK	1.94E-19	2.30E-16	9.25E-14	1183	1869	1869	3037	1.02	582	31.14	6019.76	
102	FOSL1_HUMAN.H11MO.0.A	KRVTGAGTCAYH	8.17E-19	2.91E-16	1.17E-13	356	1869	1869	3037	1.04	214	11.45	144.61	
103	OLIG2_HUMAN.H11MO.0.B	CCAKCTGYTYYYYWBCHB	1.61E-19	3.15E-16	1.26E-13	1958	1869	1869	3037	1.11	861	46.07	1008	33.19
104	BC11A_HUMAN.H11MO.0.A	RAAAGAGGAAGTGARAV	2.70E-19	7.53E-16	3.03E-13	2790	1869	1869	3037	24.6	313	16.75	250	8.23
105	CEBPG_HUMAN.H11MO.0.B	RTTGCATCAKHH	7.08E-18	3.01E-15	1.21E-12	425	1869	1869	3037	1.18	228	12.2	165.27	
106	BACH2_HUMAN.H11MO.0.A	TGCTGAGTCAB	5.20E-18	3.66E-15	1.47E-12	704	1869	1869	3037	1.08	359	19.21	3110.37	
107	FOSL2_HUMAN.H11MO.0.A	NDRTGAGTCAYH	1.11E-17	3.83E-15	1.54E-12	345	1869	1869	3037	1.02	207	11.08	134.54	
108	OTX2_HUMAN.H11MO.0.A	RRRGGATTARV	5.64E-18	4.94E-15	1.99E-12	876	1869	1869	3037	1.03	441	23.6	4213.86	
109	RORG_HUMAN.H11MO.0.C	RRAASTRGGTCA	7.28E-18	5.91E-15	2.38E-12	812	1869	1869	3037	1.02	419	22.42	3912.94	
110	NFAC2_HUMAN.H11MO.0.B	TGGAAAAWY	1.24E-17	2.12E-14	8.52E-12	1710	1869	1869	3037	18.3	131	7.01	622.04	
111	THB_HUMAN.H11MO.0.C	GRSSYSASVTSAGGTCA	2.19E-17	2.22E-14	8.91E-12	1014	1869	1869	3037	1.1	491	26.27	493	16.23
112	OSR2_HUMAN.H11MO.0.C	GBYSCWGCTNCTGYDG	1.05E-17	2.77E-14	1.12E-11	2636	1869	1869	3037	1.41	1000	53.5	1246	41.03
113	PO2F2_HUMAN.H11MO.0.A	AYATGCAAATK	3.23E-17	2.79E-14	1.12E-11	865	1869	1869	3037	1.23	392	20.97	3611.95	
114	PO5F1_HUMAN.H11MO.0.A	CYWTTSWYATGCAAAT	4.35E-17	3.56E-14	1.43E-11	818	1869	1869	3037	1.15	402	21.51	377	12.41
115	ATF2_HUMAN.H11MO.0.B	RRTGABGTCAY	6.62E-17	4.64E-14	1.87E-11	701	1869	1869	3037	1.02	366	19.58	3310.93	
116	PO3F1_HUMAN.H11MO.0.C	VATKSWWATGCAAW	5.97E-17	5.84E-14	2.35E-11	978	1869	1869	3037	1.05	483	25.84	4816	
117	SNAI1_HUMAN.H11MO.0.C	SCAGGTGK	3.60E-17	6.65E-14	2.67E-11	1849	1869	1869	3037	1.61	769	41.14	8929.44	
118	ITF2_HUMAN.H11MO.0.C	DGCAGSTGKS	4.04E-17	1.19E-13	4.78E-11	2942	1869	1869	3037	1.05	1235	66.08	1641	54.03
119	CRX_HUMAN.H11MO.0.B	WAAGRGGATTARV	2.01E-16	1.75E-13	7.03E-11	868	1869	1869	3037	1.02	438	23.43	4314.16	
120	BRAC_HUMAN.H11MO.0.A	YTVRCAKSDARGTGTGAADT	1.65E-16	1.75E-13	7.03E-11	1057	1869	1869	3037	1.08	510	27.29	527	17.35
121	TEAD4_HUMAN.H11MO.0.A	RCATTCCWGRVAK	1.69E-16	1.83E-13	7.37E-11	1087	1869	1869	3037	1.03	528	28.25	5518.18	
122	ATF3_HUMAN.H11MO.0.A	GGTSACGTGAB	4.29E-16	2.41E-13	9.68E-11	562	1869	1869	3037	1.08	290	15.52	248	
123	NKX61_HUMAN.H11MO.0.B	GAYWAATKRSHWWTTAATK	3.06E-16	3.61E-13	1.45E-10	1181	1869	1869	3037	1.08	569	30.44	612	20.15
124	TYY1_HUMAN.H11MO.0.A	CAARATGGCGGC	1.06E-15	3.72E-13	1.50E-10	351	1869	1869	3037	4.95	130	6.96	682.24	
125	NR1H3_HUMAN.H11MO.0.B	CHRRGGTYACTNNAGGTCA	6.64E-16	6.64E-13	2.67E-10	999	1869	1869	3037	1.06	487	26.06	501	16.5
126	TFE3_HUMAN.H11MO.0.B	RGTCACGTGA	6.33E-15	2.32E-12	9.31E-10	366	1869	1869	3037	1.02	210	11.24	155.14	
127	KLF8_HUMAN.H11MO.0.C	CAGGGKGTG	7.59E-15	9.17E-12	3.69E-09	1209	1869	1869	3037	9.93	166	8.88	108	3.56
128	P73_HUMAN.H11MO.0.A	RRRCAWGYCCARRCWTGYC	1.85E-14	1.03E-11	4.12E-09	554	1869	1869	3037	1.08	292	15.62	258.46	
129	DBP_HUMAN.H11MO.0.B	VTTATGYAABM	3.86E-14	2.74E-11	1.10E-08	708	1869	1869	3037	1.05	352	18.83	336	11.06
130	SIX2_HUMAN.H11MO.0.A	KGAAAYHTGAKMC	2.62E-14	3.04E-11	1.22E-08	1161	1869	1869	3037	1.03	550	29.43	605	19.92
131	SNAI2_HUMAN.H11MO.0.A	RRCAGGTGCN	3.13E-14	3.26E-11	1.31E-08	1043	1869	1869	3037	4.03	377	20.17	368	12.12
132	NR2C1_HUMAN.H11MO.0.C	GSYVAGAGGTCAB	9.73E-14	6.82E-11	2.74E-08	701	1869	1869	3037	1.39	313	16.75	289	9.52
133	HTF4_HUMAN.H11MO.0.A	DSCAGSTGKG	3.27E-14	8.22E-11	3.31E-08	2519	1869	1869	3037	3.75	486	26	516	16.99
134	ESR1_HUMAN.H11MO.0.A	AGGTCASMSTGACCY	2.84E-13	2.35E-10	9.46E-08	829	1869	1869	3037	1.45	331	17.71	316	10.41
135	PO3F2_HUMAN.H11MO.0.A	YYCATGCATWTYCATTM	3.71E-13	3.26E-10	1.31E-07	878	1869	1869	3037	1.25	387	20.71	3912.87	
136	GFI1B_HUMAN.H11MO.0.A	GCWSWGATTT	3.21E-13	3.92E-10	1.58E-07	1223	1869	1869	3037	1.04	564	30.18	637	20.97
137	ZN250_HUMAN.H11MO.0.C	TACCATRCTGTTTTGRTTAC	2.26E-11	8.12E-10	3.27E-07	36	1869	1869	3037	2.36	33	1.77	3	0.1
138	NFAC3_HUMAN.H11MO.0.B	TGGAAAAHY	6.72E-13	1.09E-09	4.37E-07	1617	1869	1869	3037	12.1	141	7.54	91	3
139	NR1I2_HUMAN.H11MO.0.C	DDRGGTYAANNRAGTTCAB	2.01E-12	1.16E-09	4.68E-07	579	1869	1869	3037	1.12	289	15.46	269	8.86
140	CEBPE_HUMAN.H11MO.0.A	TTKCGHAATMTB	1.96E-12	1.68E-09	6.77E-07	857	1869	1869	3037	1.02	415	22.2	436	14.36

rank	motif_ID	consensus	p-value	adj_p-value	E-value	tests	FASTA max	pos	neg	PWM_min	TP	%TP	FP	%FP
141	LYL1_HUMAN.H11MO.0.A	CCAKCTGYTYYCTB	6.19E-13	2.09E-09	8.40E-07	3375	1869	1869	3037	1.17	1276	68.27	1766	58.15
142	TEAD1_HUMAN.H11MO.0.A	RCATTCCWGSCW	4.67E-12	4.26E-09	1.71E-06	912	1869	1869	3037	1.02	439	23.49	473	15.57
143	GRHL2_HUMAN.H11MO.0.A	AACCDGTYTKDC	6.49E-12	4.87E-09	1.96E-06	751	1869	1869	3037	1.02	370	19.8	380	12.51
144	ZFP42_HUMAN.H11MO.0.A	GGCMGCCATTTT	9.30E-12	6.67E-09	2.68E-06	717	1869	1869	3037	1.04	351	18.78	356	11.72
145	NFKB2_HUMAN.H11MO.0.B	GGGRAAKYCCC	4.86E-12	6.72E-09	2.70E-06	1384	1869	1869	3037	1.24	582	31.14	677	22.29
146	PDX1_HUMAN.H11MO.0.A	TGATTGATK	7.87E-12	7.10E-09	2.85E-06	902	1869	1869	3037	1.04	425	22.74	456	15.01
147	TBX3_HUMAN.H11MO.0.C	DRAGGTGBSAR	3.43E-12	1.01E-08	4.05E-06	2938	1869	1869	3037	1.06	1195	63.94	1639	53.97
148	FOXQ1_HUMAN.H11MO.0.C	HATTGTTTATTT	6.51E-11	1.61E-08	6.49E-06	248	1869	1869	3037	2.06	103	5.51	61	2.01
149	ZN214_HUMAN.H11MO.0.C	WCMWTYYTRWGSMHYTTGATGA	2.82E-11	2.45E-08	9.86E-06	868	1869	1869	3037	1.19	399	21.35	426	14.03
150	ZIM3_HUMAN.H11MO.0.C	KTRGGTTTCTGTTKYYTKY	7.90E-11	3.90E-08	1.57E-05	493	1869	1869	3037	2.96	188	10.06	156	5.14
151	SOX10_HUMAN.H11MO.0.B	MAARRRVDBCWTTGTT	5.39E-11	4.59E-08	1.84E-05	852	1869	1869	3037	1.5	341	18.25	350	11.52
152	ALX1_HUMAN.H11MO.0.B	RTAATTRVATTA	6.40E-11	4.92E-08	1.98E-05	768	1869	1869	3037	1.04	370	19.8	390	12.84
153	HXA13_HUMAN.H11MO.0.C	GDTTTTATTGG	5.87E-11	5.67E-08	2.28E-05	966	1869	1869	3037	1.02	456	24.4	509	16.76
154	TF7L2_HUMAN.H11MO.0.A	BCCTTTGATSTBB	1.52E-10	1.47E-07	5.89E-05	966	1869	1869	3037	1.02	454	24.29	511	16.83
155	RUNX2_HUMAN.H11MO.0.A	KGKYTGTGGTTTKK	1.18E-10	1.52E-07	6.13E-05	1293	1869	1869	3037	1.04	587	31.41	702	23.11
156	ATF1_HUMAN.H11MO.0.B	HRRTGACGYMA	3.06E-10	1.87E-07	7.53E-05	612	1869	1869	3037	1.03	298	15.94	300	9.88
157	PAX6_HUMAN.H11MO.0.C	TYMYGCTTSABT	2.49E-10	2.24E-07	9.02E-05	902	1869	1869	3037	1.86	221	11.82	200	6.59
158	HXA10_HUMAN.H11MO.0.C	KATRRTTTATGW	2.78E-10	2.26E-07	9.10E-05	815	1869	1869	3037	1.15	368	19.69	394	12.97
159	SOX9_HUMAN.H11MO.0.B	AADGGNNBCWTTGTTY	5.18E-10	4.11E-07	1.65E-04	794	1869	1869	3037	1.43	324	17.34	337	11.1
160	PRGR_HUMAN.H11MO.0.A	RGAACATTYTGTTCTB	5.42E-10	5.03E-07	2.02E-04	928	1869	1869	3037	1.08	426	22.79	478	15.74
161	CREB1_HUMAN.H11MO.0.A	NRRTGACGTMA	1.34E-09	6.78E-07	2.73E-04	505	1869	1869	3037	1.03	253	13.54	247	8.13
162	ZBT18_HUMAN.H11MO.0.C	KCCAGMTGTKB	5.08E-10	8.84E-07	3.55E-04	1741	1869	1869	3037	2.35	417	22.31	465	15.31
163	ZN281_HUMAN.H11MO.0.A	RGGWGGGGGAGGGGV	5.40E-10	1.08E-06	4.34E-04	2000	1869	1869	3037	669	56	3	21	0.69
164	PBX1_HUMAN.H11MO.0.A	TGABTGACAG	2.14E-09	2.68E-06	1.08E-03	1254	1869	1869	3037	1.06	538	28.79	647	21.3
165	MEIS2_HUMAN.H11MO.0.B	TTGACAGBTGTCM	2.37E-08	6.88E-06	2.77E-03	290	1869	1869	3037	1.43	140	7.49	116	3.82
166	HXB13_HUMAN.H11MO.0.A	TTTTATDGSH	7.04E-09	8.62E-06	3.47E-03	1224	1869	1869	3037	1.08	531	28.41	644	21.21
167	SIX1_HUMAN.H11MO.0.A	TGWAAYHTGAKMCY	1.41E-08	9.00E-06	3.62E-03	636	1869	1869	3037	1.21	278	14.87	290	9.55
168	HNF1B_HUMAN.H11MO.0.A	DGTTAATRATTAACH	3.58E-08	1.79E-05	7.20E-03	500	1869	1869	3037	1.73	198	10.59	189	6.22
169	ZN143_HUMAN.H11MO.0.A	DGCMTKCTGGGARWTGTAGTYY	8.80E-08	2.09E-05	8.42E-03	238	1869	1869	3037	10.7	57	3.05	29	0.95
170	ZNF18_HUMAN.H11MO.0.C	GGTGTGAACHSR	3.24E-08	2.59E-05	1.04E-02	801	1869	1869	3037	1.12	348	18.62	390	12.84
171	THA11_HUMAN.H11MO.0.B	KGCMTBCTGGGARTTGTAGTYY	1.47E-07	2.67E-05	1.07E-02	182	1869	1869	3037	18.7	45	2.41	19	0.63
172	NFIC_HUMAN.H11MO.0.A	CYTGGCHBYMWGCCARG	2.05E-08	3.48E-05	1.40E-02	1696	1869	1869	3037	2.38	391	20.92	448	14.75
173	NFE2_HUMAN.H11MO.0.A	VVRTGACTCAGCA	7.34E-08	3.69E-05	1.48E-02	503	1869	1869	3037	1.03	247	13.22	256	8.43
174	KLF4_HUMAN.H11MO.0.A	WGGGYGKGGC	2.21E-08	4.02E-05	1.61E-02	1815	1869	1869	3037	11	232	12.41	231	7.61
175	CDX2_HUMAN.H11MO.0.A	TTTTATTGCWBT	5.92E-08	5.04E-05	2.03E-02	851	1869	1869	3037	1.03	391	20.92	454	14.95
176	TWST1_HUMAN.H11MO.0.A	MCATCTGKWWYYMATTA	9.54E-08	5.52E-05	2.22E-02	578	1869	1869	3037	1.09	278	14.87	299	9.85
177	FOXH1_HUMAN.H11MO.0.A	TGTGGATTS	8.54E-08	7.13E-05	2.87E-02	835	1869	1869	3037	1.02	386	20.65	449	14.78
178	DLX3_HUMAN.H11MO.0.C	GMTAATTRSW	1.38E-07	8.16E-05	3.28E-02	593	1869	1869	3037	1.02	284	15.2	309	10.17
179	GCR_HUMAN.H11MO.0.A	RGAACABMVTGTNCY	1.26E-07	9.31E-05	3.74E-02	736	1869	1869	3037	1.05	342	18.3	389	12.81
180	EVI1_HUMAN.H11MO.0.B	WAGAYAAGATAANAKA	6.40E-07	2.05E-04	8.23E-02	320	1869	1869	3037	1.12	161	8.61	153	5.04
181	Z324A_HUMAN.H11MO.0.C	KVTCAAACCATCCTTKSCTKSCW	1.66E-07	2.16E-04	8.69E-02	1302	1869	1869	3037	1.4	490	26.22	604	19.89
182	TFE2_HUMAN.H11MO.0.A	SMCAGCTGCWB	1.10E-07	2.23E-04	8.96E-02	2021	1869	1869	3037	1.05	848	45.37	1149	37.83
183	PKNX1_HUMAN.H11MO.0.B	TGABTGRCAGS	2.60E-07	2.62E-04	1.06E-01	1009	1869	1869	3037	1.68	286	15.3	315	10.37
184	ASCL1_HUMAN.H11MO.0.A	CHVCASCTGCYBYY	1.97E-07	3.50E-04	1.41E-01	1780	1869	1869	3037	1.02	761	40.72	1017	33.49
185	MEIS1_HUMAN.H11MO.0.A	RTGATTKATKRB	6.46E-07	4.96E-04	1.99E-01	767	1869	1869	3037	1.02	353	18.89	414	13.63
186	PBX3_HUMAN.H11MO.0.A	TGABTGRCRGS	3.59E-07	5.11E-04	2.06E-01	1423	1869	1869	3037	2.78	197	10.54	197	6.49
187	IRF4_HUMAN.H11MO.0.A	RAAWGRGGAASTGAAASY	4.93E-07	5.44E-04	2.19E-01	1104	1869	1869	3037	1.43	425	22.74	516	16.99
188	NR4A2_HUMAN.H11MO.0.C	AAAGGTCAS	6.43E-07	5.69E-04	2.29E-01	886	1869	1869	3037	2.46	249	13.32	269	8.86
189	IRF9_HUMAN.H11MO.0.C	GAAAGCGAAAYT	1.72E-06	6.88E-04	2.77E-01	399	1869	1869	3037	1.42	175	9.36	175	5.76
190	PRDM1_HUMAN.H11MO.0.A	RAAAGTGAAAGTGR	5.49E-07	8.50E-04	3.42E-01	1549	1869	1869	3037	1.12	649	34.72	852	28.05
191	ZNF76_HUMAN.H11MO.0.C	RRKGCMTKCTGGGARWTGTAGT	1.63E-06	8.97E-04	3.61E-01	549	1869	1869	3037	1.25	255	13.64	282	9.29
192	ERR2_HUMAN.H11MO.0.A	YCAAGGTCA	1.23E-06	1.05E-03	4.24E-01	854	1869	1869	3037	3.25	173	9.26	171	5.63
193	BMAL1_HUMAN.H11MO.0.A	KSCACGTGHSY	1.31E-06	1.23E-03	4.96E-01	946	1869	1869	3037	1.07	413	22.1	505	16.63
194	SMCA5_HUMAN.H11MO.0.C	GAATBGAATGGAAWC	1.45E-06	1.36E-03	5.46E-01	939	1869	1869	3037	1.06	418	22.36	513	16.89
195	TFEB_HUMAN.H11MO.0.C	RGTCACGTG	1.43E-05	1.44E-03	5.80E-01	101	1869	1869	3037	61.4	22	1.18	6	0.2
196	P53_HUMAN.H11MO.0.A	DGRCATGYCCAGRCATGYCY	1.09E-05	1.94E-03	7.78E-01	177	1869	1869	3037	2.98	55	2.94	36	1.19
197	ERR1_HUMAN.H11MO.0.A	BYCAAGGTCAMABWV	4.72E-06	2.96E-03	1.19E+00	627	1869	1869	3037	1.28	253	13.54	285	9.38
198	NF2L2_HUMAN.H11MO.0.A	WWYTGCTGAGTCAT	1.75E-05	3.34E-03	1.34E+00	191	1869	1869	3037	1.08	100	5.35	89	2.93
199	ZNF8_HUMAN.H11MO.0.C	TGTGGTATATCCATACAATGGA	4.57E-04	5.01E-03	2.02E+00	11	1869	1869	3037	2.4	10	0.54	1	0.03
200	MYF6_HUMAN.H11MO.0.C	CASCTGC	2.58E-06	5.43E-03	2.18E+00	2110	1869	1869	3037	1.15	875	46.82	1219	40.14
201	ZN257_HUMAN.H11MO.0.C	GMYCTTGCCTCW	5.32E-06	5.70E-03	2.29E+00	1074	1869	1869	3037	1.25	407	21.78	506	16.66
202	NKX32_HUMAN.H11MO.0.C	TTAAGTGBWT	1.24E-05	5.93E-03	2.39E+00	481	1869	1869	3037	4.21	91	4.87	77	2.54
203	ISL1_HUMAN.H11MO.0.A	VCTAATKGV	5.40E-06	7.55E-03	3.04E+00	1403	1869	1869	3037	1.05	574	30.71	756	24.89
204	CREM_HUMAN.H11MO.0.C	CRVTGACGTCA	1.77E-05	8.13E-03	3.27E+00	462	1869	1869	3037	1.32	185	9.9	199	6.55
205	GATA6_HUMAN.H11MO.0.A	DVAGATAAGAWAA	1.65E-05	8.33E-03	3.35E+00	508	1869	1869	3037	1.19	227	12.15	256	8.43
206	PEBB_HUMAN.H11MO.0.C	TYTGTGGTYWB	7.73E-06	9.66E-03	3.88E+00	1256	1869	1869	3037	1.12	504	26.97	653	21.5
207	ESR2_HUMAN.H11MO.0.A	RGGTCASCSTGMCCY	9.33E-06	1.17E-02	4.69E+00	1258	1869	1869	3037	1.15	499	26.7	647	21.3
208	HLF_HUMAN.H11MO.0.C	GRTTRYGTAAYMS	4.58E-05	1.20E-02	4.81E+00	263	1869	1869	3037	1.04	128	6.85	128	4.21
209	HXA9_HUMAN.H11MO.0.B	RTGATTTATKR	1.43E-05	1.20E-02	4.82E+00	841	1869	1869	3037	1.43	312	16.69	375	12.35
210	BHE40_HUMAN.H11MO.0.A	DGCACGTGAS	1.91E-05	1.39E-02	5.61E+00	735	1869	1869	3037	9.76	51	2.73	33	1.09
211	HNF6_HUMAN.H11MO.0.B	WTTATTGATTK	5.64E-05	1.44E-02	5.80E+00	258	1869	1869	3037	4.28	71	3.8	58	1.91
212	LHX3_HUMAN.H11MO.0.C	AAAATTAATTARY	5.41E-05	1.85E-02	7.45E+00	346	1869	1869	3037	122	24	1.28	9	0.3
213	AP2C_HUMAN.H11MO.0.A	VTGSCCTGRGGSHR	1.22E-05	2.14E-02	8.62E+00	1772	1869	1869	3037	28	48	2.57	29	0.95
214	NDF1_HUMAN.H11MO.0.A	RRCAGATGGY	2.08E-05	2.24E-02	9.02E+00	1090	1869	1869	3037	1.02	474	25.36	616	20.28
215	REL_HUMAN.H11MO.0.B	DRWRGGGRAATKCCA	1.66E-05	2.30E-02	9.25E+00	1405	1869	1869	3037	4.03	219	11.72	245	8.07

C.2.6 Expression of genes with motifs enriched within shared hypoDMRs

Gene	log2FoldChange	padj
ZNF770	0.350544467	0.87567
ZNF121	0.231363889	0.93456
SRY	-0.092466573	NA
PAX5	-1.071794951	0.27083
FOSB	-1.45428891	0.06651
FOXK1	0.309384184	0.73768
MEF2B	-0.029820392	NA
ARID5B	-0.243562091	0.91171
FOXJ2	0.238616248	0.87468
FOXM1	-0.106347295	0.95534
FOXO1	-0.783122365	0.42889
SOX5	0.154263551	0.95437
AR	-1.019265835	0.47795
NANOG	-0.035036221	NA
FOXP1	0.088580526	0.96557
MAF	-0.950563675	0.01507
BATF3	-0.515755151	0.85494
RXRB	0.130191349	0.92869
NR2E3	0.298454043	0.92786
RARA	-0.063740402	0.97185
FOXJ3	0.308769963	0.85823
MEF2A	0.110523743	0.95277
SOX17	-0.549123418	0.84752
FEZF1	0.223248762	0.94861
NR2F2	-0.042389411	0.9841
MEF2D	0.103476515	0.95106
MEF2C	-0.013610416	0.99559
SREBF1	-0.051763005	0.97791
ZNF354A	0.467920888	0.84752
LEF1	-1.143696649	0.07356
SOX2	0.40434347	0.90381
NR1D1	0.607348357	0.66454
MAFB	0.047341795	0.98441
MAFF	-0.295253825	0.84752
SOX3	0.037089935	NA
SOX4	-0.371086949	0.86491
NKX2-1	0.040029071	0.99071
ZFP28	0.118011502	0.96557
TBX21	-1.044756904	0.48981
FOS	-0.483331644	0.79648
POU1F1	0.105979912	NA
TCF7	-0.202858265	0.91997
ZNF350	0.454820219	0.84752
IRF1	-0.524260115	0.74039
AIRE	-0.082869942	NA
IKZF1	-0.888028637	0.32973
STAT2	-0.102506769	0.94861
ZNF554	0.42031352	0.84345
SMAD2	0.309786056	0.88262
ZNF394	0.362268583	0.85033
GFI1	-0.482363386	0.84697
FOXA1	0.307223825	0.91473
SMAD4	0.306062473	0.87788
BACH1	-0.026818573	0.99071
IRF2	-0.059976921	0.97636
RORA	0.097466651	0.97185
FOXA3	0.020188513	0.99545
CDX1	-0.432299024	NA
FOXO4	0.405588611	0.83256
ZNF418	0.594253781	0.82662
FOXA2	0.11363366	0.97185
NKX3-1	0.115600243	0.96964
MITF	-0.397319586	0.86308
ZNF85	0.24287082	0.94056
BATF	0.187163965	0.9536
TBP	0.293934611	0.90345
SMAD3	0.006342187	0.99823
PRDM6	-1.242327453	0.20007
ZNF41	0.424196482	0.85634
FOXP2	0.107839916	0.97185
FOXO3	-0.205522863	0.92223
SMARCA4	0.103485113	0.96883

Gene	log2FoldChange	padj
NFATC1	-1.0290368	0.04281
FOXC1	-0.164605128	0.94964
IRF3	0.17006171	0.90661
CEBPA	-0.191012801	0.94627
IRF7	-0.177103503	0.91148
TCF7L1	-0.257813003	0.91908
ZNF384	0.427896902	0.8145
JUNB	-0.719038275	0.0945
NR4A1	-1.066497096	0.19882
ZFP82	-0.001615495	0.99955
ZEB1	-0.011500357	0.99545
ESRRG	-0.16255953	0.95902
GATA3	-0.694674365	0.74688
CEBPD	-0.207622005	0.92933
PBX2	-0.293169941	0.86308
THRA	0.819297093	0.20428
POU2F1	0.215386851	0.91419
VDR	-0.494330213	0.79988
TGIF1	-0.070365591	0.9772
NFIA	0.511490864	0.78299
JUND	-0.666539741	0.5293
JUN	-0.454673071	0.72514
IRF8	-0.494916356	0.83192
RARG	0.50570711	0.84752
CEBPB	-0.439774846	0.84575
HNF1A	-0.102390038	0.97262
CLOCK	0.380877112	0.84752
NFE2L1	0.304262482	0.77718
LHX2	0.113371659	0.97185
FOSL1	-0.040331386	0.9905
OLIG2	0.474868966	0.87909
BCL11A	-0.147121035	0.96162
CEBPG	0.00552572	0.99833
BACH2	-0.795913004	0.59954
FOSL2	-0.579069834	0.77325
OTX2	-0.066379367	NA
RORC	0.702707479	0.75079
NFATC2	-0.686484338	0.38014
THRB	0.480083666	0.85477
OSR2	-0.250314301	0.94222
POU2F2	-0.57267021	0.76394
POU5F1	1.559016286	0.08385
ATF2	0.320428189	0.88262
POU3F1	0.598255797	0.84752
SNAI1	0.222752041	0.94453
TCF4	-0.978691588	0.12186
CRX	0.044137066	NA
T	0.061472288	NA
TEAD4	-0.302383703	0.91171
ATF3	-1.490802157	0.01316
NKX6-1	0.012831911	0.99715
YY1	-0.003512954	0.99904
NR1H3	-0.341765458	0.87567
TFE3	-0.019086451	0.98952
KLF8	-0.01251551	0.997
TP73	0.586897536	0.78264
DBP	0.237892917	0.91921
SIX2	0.135557094	0.96595
SNAI2	-0.507379228	0.83256
NR2C1	0.523364111	0.83085
TCF12	0.172352734	0.93401
ESR1	-0.17390553	0.93181
POU3F2	-0.476112474	0.87673
GFI1B	-0.203068189	NA
ZNF250	0.418887639	0.83256
NFATC3	0.003433477	0.99904
NR1I2	0.362588354	NA
CEBPE	-0.159191197	0.96067
LYL1	-0.389232636	0.87567
TEAD1	0.133248299	0.94894
GRHL2	0.339495882	0.87941
ZFP42	-0.098458632	NA

Gene	log2FoldChange	padj
NFKB2	-0.549535681	0.63452
PDX1	0.036879208	0.99111
TBX3	-0.860492457	0.53875
FOXQ1	0.122758875	0.96927
ZNF214	0.053445926	0.98867
ZIM3	NA	NA
SOX10	1.195779606	0.35225
ALX1	-0.082280441	NA
HOXA13	0.304671771	0.92786
TCF7L2	0.182527071	0.93545
RUNX2	-0.754622381	0.63452
ATF1	0.280779315	0.91811
PAX6	0.224541147	0.94323
HOXA10	0.527990949	0.84575
SOX9	0.418061009	0.88539
PGR	-0.858738835	0.45791
CREB1	0.1234924	0.95534
ZBTB18	0.289561906	0.88624
ZNF281	-0.029071169	0.99053
PBX1	0.273588083	0.92389
MEIS2	-0.110366018	0.96557
HOXB13	0.910002126	0.66834
SIX1	-0.353592343	0.90381
HNF1B	-0.014849103	0.99697
ZNF143	0.070027711	0.97885
ZNF18	0.141966992	0.9511
THAP11	0.247206328	0.88456
NFIC	-0.017089052	0.99371
NFE2	-0.962000432	0.58538
KLF4	-0.170257548	0.93918
CDX2	0.205908909	NA
TWIST1	-1.029207392	0.42084
FOXH1	-0.037305247	0.99093
DLX3	0.022426329	0.99503
NR3C1	-0.299091564	0.84752
MECOM	-0.210673044	0.93753
ZNF324	0.355331892	0.79241
TCF3	-0.137074324	0.93948
PKNOX1	0.216317077	0.91663
ASCL1	-0.421474272	0.897
MEIS1	-0.890352861	0.39298
PBX3	-0.094893788	0.96927
IRF4	-1.316192146	0.14171
NR4A2	-0.330673521	0.91663
IRF9	-0.254241246	0.90794
PRDM1	-0.911538608	0.34484
ZNF76	0.310383364	0.86935
ESRRB	0.037211426	NA
ARNTL	0.124278137	0.95534
SMARCA5	0.107023259	0.96312
TFEB	0.120683667	0.95534
TP53	-0.104776391	0.96312
ESRRA	0.392843488	0.83074
NFE2L2	-0.41451441	0.80815
ZNF8	0.116292606	0.96883
MYF6	0.013679566	NA
ZNF257	0.131966662	0.96557
NKX3-2	-0.587653543	NA
ISL1	-0.025306531	NA
CREM	-0.453572915	0.83074
GATA6	-1.20400582	0.14557
CBFB	0.155439424	0.93485
ESR2	0.140726445	0.94861
HLF	0.59858003	0.83692
HOXA9	0.154709754	NA
BHLHE40	-0.943038296	0.12349
ONECUT1	-0.513215872	0.87468
LHX3	0.171927111	NA
TFAP2C	0.581213257	0.8145
NEUROD1	0.084311384	NA
REL	-0.079358283	0.97185

C.2.7 Ontology enrichment of genes with enriched motifs in shared hypoDMRs

All "Categories" are: GOTERM_BP_DIRECT

Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0045944~positive regulation of transcription from RNA polymerase II promoter	141	65.5814	1.15E-120	MEF2C, MEF2B, CDX1, MEF2A, THRA, THRB, FOXO1, PDX1, NR2E3, FOXO3, RORA, FOXO4, CBFB, BATF, PGR, GATA6, GATA3, RARA, TWIST1, RARG, RXRB, YY1, SIX2, ARNTL, GRHL2, JUNB, ASCL1, JUN, SIX1, NFE2L1, TFAP2C, NFE2L2, ONECUT1, TFE3, SOX2, TFEB, SOX4, SOX9, MEIS1, ATF1, ATF2, FOXH1, VDR, MEIS2, LHX2, JUND, LHX3, HOXA10, NKX3-1, NKX2-1, TCF4, TCF3, NR1H3, SREBF1, MAFF, ESRRA, IKZF1, MAFB, ESRRB, FOXA1, ESRRG, NR4A2, SMAD4, SMAD3, NR4A1, SMAD2, NKX6-1, FOXP1, DLX3, NR1I2, ATF3, AIRE, TCF12, KLF4, BACH1, HLF, HNF1B, FOSL2, HNF1A, FOXA2, FOXA3, MITF, PAX6, PAX5, NFKB2, ZEB1, CRX, FOS, ZNF350,	215	981	16792	11.2257165	1.97E-117	1.97E-117	1.94E-117
GO:0006366~transcription from RNA polymerase II promoter	107	49.7674	2.98E-104	MEF2C, MEF2B, CDX1, MEF2A, CDX2, THRA, TBP, NR2E3, FOXO3, PDX1, FOXO4, CBFB, BATF, GATA6, GATA3, SIX2, GRHL2, JUNB, JUN, SIX1, NFE2L1, TFAP2C, NFE2L2, ONECUT1, SOX2, SOX5, SOX4, SOX9, MEIS1, ATF1, ATF2, FOXH1, MEIS2, JUND, NKX3-2, HOXA10, RUNX2, MAFF, MAFB, FOXA1, SMAD4, SMAD2, NKX6-1, DLX3, ATF3, AIRE, TCF12, KLF4, BACH1, HLF, BACH2, HNF1A, MITF, PAX6, PAX5, NFKB2, CRX, FOS, OSR2, GFI1B, POU5F1, FOSL1, ALX1, BATF3, SOX10, FOXJ2, OTX2, TP53, ESR1, LEF1, FOSB, ZNF143, FOXC1, CLOCK, FOXM1, NR3C1, POU1F1, REL, POU2F2, POU3F2, NFATC2, NFATC3, NFATC1, CEBPA, MAF, MYF6, CEBPB, CEBPE, CEBPD,	215	513	16792	16.2903486	5.11E-101	2.56E-101	5.03E-101
GO:0000122~negative regulation of transcription from RNA polymerase II promoter	97	45.1163	4.36E-74	MEF2C, MEF2A, CDX2, THRB, FOXO1, NR2E3, FOXO3, PDX1, GATA6, GATA3, RARA, OLIG2, TWIST1, RARG, YY1, ZNF8, JUNB, ASCL1, SIX1, TGIF1, PRDM1, TFAP2C, SOX3, SOX2, ATF2, NR2C1, FOXH1, VDR, T, MEIS2, JUND, NKX3-2, NKX2-1, BHLHE40, TCF4, TCF3, NR1H3, SREBF1, ESRRA, TCF7, IKZF1, FOXA1, SMAD4, NR4A2, SMAD3, SMAD2, FOXP1, NKX6-1, FOXP2, FEZF1, ATF3, KLF4, BACH1, HNF1B, BACH2, HNF1A, MITF, PAX6, PAX5, ZEB1, NFKB2, ZNF350, OSR2, GFI1B, POU5F1, SOX17, NR2F2, ALX1, BATF3, ZNF281, ARID5B, ESR1, TP53, LEF1, ESR2, FOSB, FOXM1, POU1F1, TCF7L2, ZBTB18, REL, NR1D1, BCL11A, GFI1, NFATC2, MAF, CEBPA, TBX3, ISL1, SNAI2, SNAI1, TP73,	215	720	16792	10.5221189	7.47E-71	2.49E-71	7.36E-71

Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0006351~transcription, DNA-templated	128	59.5349	8.08E-64	MEF2C, MEF2B, MEF2A, THRA, THRB, ZNF250, RORC, FOXO1, NR2E3, RORA, PGR, RARA, OLIG2, ZNF394, TWIST1, RARG, RXRB, YY1, ZNF8, ARNTL, MECOM, ASCL1, ZFP82, PRDM6, ZNF384, TGIF1, PRDM1, NFE2L2, SMARCA1, ZIM3, ZNF41, HOXA13, ZNF76, SOX3, TFE3, TFEB, HOXB13, NR2C1, SRY, T, VDR, FOXQ1, ZNF324, LYL1, LHX2, TEAD4, LHX3, NKX3-1, HOXA9, NKX2-1, BHLHE40, THAP11, TCF3, NR1H3, ESRRA, TCF7, KLF8, NFE2, IKZF1, ESRRB, ZNF121, ESRRG, NR4A2, SMAD4, SMAD3, NR4A1, SMAD2, FOXP1, FOXP2, FEZF1, NR1I2, ZNF214, ZNF418, ZNF85, HNF1B, FOSL2, ZNF18, FOXA2, FOXK1, FOXA3, TBX21, ZFP42, PAX6, PAX5, ZEB1, ZNF350, POU5F1, NR2F2,	215	1955	16792	5.11360495	1.39E-60	3.46E-61	1.36E-60
GO:0045893~positive regulation of transcription, DNA-templated	79	36.7442	3.28E-63	MEF2C, HNF1B, CDX2, HNF1A, FOXK1, FOXA2, TBX21, MITF, PAX6, RORC, FOXO1, TBP, FOXO3, RORA, FOS, OSR2, GATA6, GATA3, RARA, SOX17, NR2F2, ALX1, ZNF281, AR, FOXJ2, OTX2, TP53, ESR1, LEF1, ARNTL, ESR2, MECOM, JUN, SIX1, SMARCA5, FOXC1, NFE2L1, SMARCA1, CLOCK, FOXM1, TFE3, SOX2, TFEB, SOX4, SOX9, POU1F1, FOXH1, SRY, NR1D1, LYL1, LHX2, LHX3, NKX3-1, NKX2-1, POU3F1, TCF4, NFATC2, RUNX2, TCF3, NFATC3, NR1H3, NFATC1, NFE2, TBX3, CREB1, SMAD4, ESRRG, SMAD3, SMAD2, SNAI1, TP73, FEZF1, NR1I2, IRF7, AIRE, IRF1, NEUROD1, IRF4, KLF4	215	515	16792	11.9807451	5.63E-60	1.13E-60	5.54E-60
GO:0006355~regulation of transcription, DNA-templated	96	44.6512	1.79E-43	MEF2C, BACH1, ZNF85, CDX1, MEF2A, CDX2, THRA, THRB, TBX21, ZFP42, MITF, PAX6, PAX5, ZNF250, NFKB2, RORA, FOXO4, CRX, ZNF350, GFI1B, POU5F1, ZNF394, SOX17, AR, NANOG, ZNF354A, RXRB, OTX2, TP53, ESR1, ZNF8, ARNTL, ESR2, ZFP28, MECOM, ZFP82, SIX1, ZNF384, TGIF1, FOXC1, CLOCK, ZIM3, ZNF41, ZNF554, HOXA13, ONECUT1, CREM, SOX2, TFEB, SOX5, SOX4, HOXB13, NR3C1, POU1F1, MEIS1, TCF7L1, NR2C1, ATF2, VDR, MEIS2, ZNF324, LYL1, LHX2, LHX3, POU2F2, HOXA10, HOXA9, NKX3-1, GFI1, BHLHE40, NFATC2, RUNX2, THAP11, TCF3, NR1H3, ESRRA, CEBPB, KLF8, ESRRB, ZNF121, CREB1, SMAD4, ESRRG, SMAD3, ISL1, IRF9, DLX3,	215	1504	16792	4.98525482	3.07E-40	5.12E-41	3.02E-40
GO:0006357~regulation of transcription from RNA polymerase II promoter	57	26.5116	3.53E-40	ZNF18, THRA, HNF1A, FOSL2, FOXK1, FOXA2, MITF, FOXO3, ZEB1, NFKB2, CBFB, BATF, FOS, NR2F2, FOSL1, SOX10, ZNF354A, YY1, FOSB, FOXJ3, ZNF143, JUNB, SMARCA5, NFE2L1, TFAP2C, CLOCK, ZNF76, TCF7L2, TCF7L1, ATF1, ATF2, SRY, FOXQ1, LYL1, JUND, TEAD4, RUNX2, NFATC3, MAF, MYF6, SREBF1, MAFF, TCF7, NFE2, TBX3, MAFB, CREB1, CEBPG, ZNF770, SMAD3, ISL1, FOXP1, NKX6-1, STAT2, ATF3, DBP, TCF12	215	441	16792	10.0948584	6.05E-37	8.65E-38	5.96E-37
GO:0045892~negative regulation of transcription, DNA-templated	50	23.2558	1.01E-29	THRA, FOXK1, FOXO1, ZEB1, ZNF350, GATA6, GATA3, RARA, NR2F2, TWIST1, ALX1, ZNF281, SOX10, ARID5B, TP53, LEF1, ARNTL, MECOM, ASCL1, PRDM6, JUN, CLOCK, CREM, FOXM1, SOX9, TCF7L2, ZBTB18, VDR, NR1D1, POU2F1, NKX2-1, NKX3-1, BHLHE40, GFI1, RUNX2, MYF6, CEBPA, CEBPB, IKZF1, TBX3, CEBPD, SMAD4, SMAD2, FOXP1, FOXP2, FEZF1, NR1I2, ZNF418, IRF1, KLF4	215	499	16792	7.82588433	1.72E-26	2.16E-27	1.70E-26

Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0043401~steroid hormone mediated signaling pathway	22	10.2326	8.40E-26	ESRRA, THRA, RARG, THRB, ESRRB, RXRB, NR4A2, ESR1, ESRRG, LEF1, NR4A1, RORC, RORA, NR2E3, NR2C1, PGR, VDR, NR1I2, NR1D1, RARA, NR2F2, NR1H3	215	57	16792	30.1447572	1.44E-22	1.60E-23	1.42E-22
GO:0006367~transcription initiation from RNA polymerase II promoter	27	12.5581	3.77E-22	THRA, THRB, RORC, TBP, NR3C1, RORA, NR2E3, NR2C1, PGR, VDR, NR1D1, TEAD4, RARA, TCF4, NR1H3, AR, ESRRA, RARG, RXRB, ESRRB, NR4A2, ESR1, ESRRG, TEAD1, NR4A1, ESR2, NR1I2	215	152	16792	13.8734394	6.47E-19	6.47E-20	6.37E-19
GO:0030522~intracellular receptor signaling pathway	14	6.51163	6.49E-16	ESRRA, AR, THRA, THRB, ESRRB, NR4A2, NR4A1, RORC, NR2E3, RORA, NR1I2, NR1D1, NR2F2, NR1H3	215	38	16792	28.774541	1.14E-12	1.04E-13	1.12E-12
GO:0060070~canonical Wnt signaling pathway	12	5.5814	8.05E-09	T, TCF7, RARG, NR4A2, SMAD3, SOX4, LEF1, SOX17, SNAI2, TCF7L2, TCF7L1, KLF4	215	83	16792	11.2919025	1.38E-05	1.15E-06	1.36E-05
GO:0031018~endocrine pancreas development	8	3.72093	2.65E-08	HNF1B, HNF1A, FOXA2, ONECUT1, FOXO1, SOX4, NEUROD1, SOX9	215	26	16792	24.0314848	4.55E-05	3.50E-06	4.48E-05
GO:0001889~liver development	11	5.11628	3.23E-08	CEBPA, HNF1A, ONECUT1, DBP, GATA6, JUN, CEBPG, ARID5B, SMAD3, RARA, PDX1	215	74	16792	11.6098052	5.54E-05	3.96E-06	5.46E-05
GO:0045665~negative regulation of neuron differentiation	10	4.65116	3.38E-08	ASCL1, SOX3, SOX2, PAX6, PBX1, FOXO3, OLIG2, ISL1, MEIS1, TP73	215	56	16792	13.9468439	5.80E-05	3.86E-06	5.71E-05
GO:0008285~negative regulation of cell proliferation	22	10.2326	3.77E-08	CEBPA, AR, RARG, TP53, SMAD4, SOX4, SMAD2, PDX1, NR2E3, ZEB1, POU1F1, FOXO4, TP73, FEZF1, VDR, JUN, GATA3, IRF1, NKX3-1, RARA, FOSL1, KLF4	215	396	16792	4.33901809	6.46E-05	4.04E-06	6.36E-05
GO:0060021~palate development	11	5.11628	4.20E-08	MEF2C, OSR2, TBX3, ARID5B, SMAD4, LEF1, SMAD2, SNAI2, SNAI1, ALX1, TWIST1	215	76	16792	11.304284	7.20E-05	4.24E-06	7.09E-05
GO:0010628~positive regulation of gene expression	18	8.37209	4.47E-08	MEF2C, AR, MITF, TP53, PAX6, SMAD3, LEF1, VDR, NR1I2, OSR2, POU5F1, NKX2-1, NKX3-1, PRDM1, NFE2L2, POU3F1, KLF4, TWIST1	215	262	16792	5.36580863	7.67E-05	4.26E-06	7.55E-05
GO:0045666~positive regulation of neuron differentiation	11	5.11628	5.42E-08	FEZF1, MEF2C, ASCL1, FOXA1, NEUROD1, RARA, ZEB1, TCF4, TCF12, TCF3, NKX6-1	215	78	16792	11.0144305	9.29E-05	4.89E-06	9.14E-05
GO:0048701~embryonic cranial skeleton morphogenesis	8	3.72093	1.00E-07	SIX1, SIX2, PAX5, SMAD3, SMAD2, RUNX2, GRHL2, TWIST1	215	31	16792	20.1554389	1.72E-04	8.61E-06	1.69E-04
GO:0035019~somatic stem cell population maintenance	10	4.65116	1.29E-07	NANOG, CDX2, POU5F1, SOX2, SMAD4, SOX4, SMAD2, PBX1, SOX9, KLF4	215	65	16792	12.0157424	2.21E-04	1.05E-05	2.18E-04
GO:0035914~skeletal muscle cell differentiation	9	4.18605	1.76E-07	MYF6, MEF2C, MEF2D, MAFF, HLF, FOS, ATF3, PAX5, NR4A1	215	49	16792	14.3453251	3.01E-04	1.37E-05	2.97E-04
GO:0009887~organ morphogenesis	11	5.11628	2.66E-07	CDX1, CDX2, TBX3, THRB, GATA3, LHX3, PAX6, PAX5, PDX1, NKX6-1, CRX	215	92	16792	9.33832154	4.56E-04	1.98E-05	4.49E-04
GO:0034097~response to cytokine	9	4.18605	2.84E-07	FOS, REL, JUN, JUND, RARA, PDX1, NFKB2, FOSL1, JUNB	215	52	16792	13.5177102	4.87E-04	2.03E-05	4.79E-04
GO:0045444~fat cell differentiation	10	4.65116	3.59E-07	CEBPA, SREBF1, CEBPD, ARID5B, NR4A2, FOXO1, NR4A1, TCF7L2, KLF4, ATF2	215	73	16792	10.6989487	6.16E-04	2.46E-05	6.06E-04
GO:0045597~positive regulation of cell differentiation	8	3.72093	3.68E-07	AR, CDX2, JUN, JUND, SOX2, NEUROD1, SOX17, JUNB	215	37	16792	16.8869893	6.31E-04	2.43E-05	6.22E-04
GO:0007623~circadian rhythm	10	4.65116	4.54E-07	SREBF1, NR1D1, DBP, JUN, CREB1, JUND, RORC, ARNTL, CLOCK, CRX	215	75	16792	10.4136434	7.79E-04	2.89E-05	7.67E-04
GO:0009952~anterior/posterior pattern specification	10	4.65116	7.94E-07	HNF1B, RARG, YY1, HOXA10, NEUROD1, HOXA9, SMAD2, PBX1, NR2F2, ALX1	215	80	16792	9.7627907	0.001361015	4.86E-05	0.001340848
GO:0021983~pituitary gland development	7	3.25581	1.19E-06	SOX3, CREB1, LHX3, SOX2, PAX6, NKX2-1, ISL1	215	28	16792	19.5255814	0.002039765	7.04E-05	0.002010217
GO:0010629~negative regulation of gene expression	12	5.5814	1.45E-06	PGR, MEF2C, ZNF281, REL, POU5F1, YY1, CREB1, ESR1, NKX3-1, SMAD2, POU3F2, KLF4	215	137	16792	6.84107961	0.002482016	8.28E-05	0.002446598
GO:0003151~outflow tract morphogenesis	8	3.72093	1.74E-06	FOXH1, MEF2C, TBX3, JUN, SIX1, SOX17, ISL1, ATF2	215	46	16792	13.5830131	0.00297518	9.61E-05	0.002933443
GO:0051591~response to cAMP	8	3.72093	1.74E-06	SREBF1, FOS, JUN, CREM, JUND, FOSB, FOSL1, JUNB	215	46	16792	13.5830131	0.00297518	9.61E-05	0.002933443
GO:0042493~response to drug	17	7.90698	1.86E-06	SREBF1, HNF1B, CREB1, PDX1, FOSB, JUNB, NKX6-1, TP73, FOS, GATA6, JUN, GATA3, JUND, TGIF1, NEUROD1, NFATC2, FOSL1	215	304	16792	4.36756426	0.003184384	9.97E-05	0.003140037
GO:0008584~male gonad development	10	4.65116	3.12E-06	GATA6, GATA3, ARID5B, ZFP42, HOXA10, ESR1, HOXA9, NKX3-1, TFAP2C, SOX9	215	94	16792	8.30875804	0.005341269	1.62E-04	0.005272533
GO:0023019~signal transduction involved in regulation of gene expression	6	2.7907	3.23E-06	T, PAX6, SMAD3, NEUROD1, SMAD2, SOX17	215	19	16792	24.6638923	0.005516195	1.63E-04	0.005445681

Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0010718~positive regulation of epithelial to mesenchymal transition	7	3.25581	3.32E-06	SMAD4, SMAD3, LEF1, SMAD2, SNAI1, ALX1, TWIST1	215	33	16792	16.56716	0.005675834	1.63E-04	0.005603725
GO:0001701~in utero embryonic development	13	6.04651	5.09E-06	SOX10, MAFF, AR, TBX3, GATA6, GATA3, SMAD4, SMAD3, SMAD2, FOXC1, FOSL1, FOXP1, TWIST1	215	187	16792	5.42957344	0.008685685	2.42E-04	0.008588202
GO:0031016~pancreas development	6	2.7907	5.53E-06	MEIS2, NKX3-2, SMAD2, ISL1, TCF7L2, NKX6-1	215	21	16792	22.3149502	0.009436181	2.56E-04	0.009333768
GO:0048469~cell maturation	7	3.25581	5.65E-06	CEBPA, ASCL1, SOX10, GATA3, POU2F2, RUNX2, CBFB	215	36	16792	15.1865633	0.009649213	2.55E-04	0.009545503
GO:0001649~osteoblast differentiation	10	4.65116	7.23E-06	MEF2C, MEF2D, SOX2, LEF1, SNAI2, RUNX2, SNAI1, JUNB, CBFB, TWIST1	215	104	16792	7.509839	0.012317704	3.18E-04	0.01220157
GO:0032922~circadian regulation of gene expression	8	3.72093	7.62E-06	HNF1B, NR1D1, CREM, RORC, BHLHE40, RORA, ARNTL, CLOCK	215	57	16792	10.9617299	0.012985959	3.27E-04	0.012867827
GO:0051091~positive regulation of sequence-specific DNA binding transcription factor activity	10	4.65116	7.82E-06	CEBPG, ARID5B, FOXA1, ESR1, NEUROD1, ESR2, TCF3, FOSL1, NKX6-1, ATF2	215	105	16792	7.43831672	0.013318616	3.27E-04	0.013199655
GO:0001764~neuron migration	10	4.65116	7.82E-06	FEZF1, MEF2C, ASCL1, GATA3, PAX6, NR4A2, NKX2-1, NR2F2, NKX6-1, TWIST1	215	105	16792	7.43831672	0.013318616	3.27E-04	0.013199655
GO:0001568~blood vessel development	7	3.25581	7.84E-06	MEF2C, DLX3, CDX2, TBX3, PAX6, FOXO1, TCF7L2	215	38	16792	14.3872705	0.013363735	3.20E-04	0.01324467
GO:0008284~positive regulation of cell proliferation	20	9.30233	8.60E-06	AR, CDX2, RARG, TBX3, FOXM1, SOX4, LEF1, PDX1, POU1F1, ISL1, SOX9, T, OSR2, ATF3, NKX3-1, RARA, PBX1, POU3F2, FOSL1, RUNX2	215	466	16792	3.35203114	0.01463607	3.43E-04	0.014514917
GO:0048839~inner ear development	7	3.25581	1.24E-05	MAF, CEBPA, CEBPD, LHX3, SIX1, SOX2, NEUROD1	215	41	16792	13.3345434	0.021002187	4.82E-04	0.020895077
GO:0060337~type I interferon signaling pathway	8	3.72093	1.66E-05	IRF9, IRF7, IRF8, IRF1, IRF2, IRF3, IRF4, STAT2	215	64	16792	9.7627907	0.028079635	6.33E-04	0.02803654
GO:0007517~muscle organ development	9	4.18605	1.78E-05	MEF2C, MEF2D, MEF2B, MEF2A, FOXK1, TEAD4, FOXO4, TCF12, TWIST1	215	89	16792	7.89798798	0.030051584	6.63E-04	0.030035498
GO:0001714~endodermal cell fate specification	4	1.86047	2.00E-05	HNF1B, NANOG, POU5F1, SOX2	215	5	16792	62.4818605	0.033768091	7.31E-04	0.033813871
GO:0060065~uterus development	5	2.32558	2.32E-05	GATA3, SMAD4, HOXA10, ESR1, HOXA9	215	14	16792	27.8936877	0.039051149	8.30E-04	0.039209714
GO:0071277~cellular response to calcium ion	7	3.25581	4.46E-05	MEF2C, FOS, MEF2A, JUN, JUND, FOSB, JUNB	215	51	16792	10.719927	0.073615891	0.00155932	0.075254179
GO:0007519~skeletal muscle tissue development	7	3.25581	4.46E-05	MYF6, MEF2C, SIX1, NR2F2, ZBTB18, FOXP1, FOXP2	215	51	16792	10.719927	0.073615891	0.00155932	0.075254179
GO:0048384~retinoic acid receptor signaling pathway	5	2.32558	6.82E-05	RARG, RXRB, CREM, ESRRG, RARA	215	18	16792	21.6950904	0.110407762	0.002337106	0.115114555
GO:2000144~positive regulation of DNA-templated transcription, initiation	4	1.86047	6.88E-05	JUN, MITF, FOSL1, TWIST1	215	7	16792	44.6299003	0.111286383	0.002310664	0.116086288
GO:0032808~lacrimal gland development	4	1.86047	6.88E-05	SOX10, PAX6, FOXC1, SOX9	215	7	16792	44.6299003	0.111286383	0.002310664	0.116086288
GO:0050680~negative regulation of epithelial cell proliferation	7	3.25581	7.62E-05	MEF2C, AR, SOX2, PAX6, NKX3-1, SOX9, ATF2	215	56	16792	9.7627907	0.122565632	0.002511324	0.128646041
GO:0006959~humoral immune response	7	3.25581	8.43E-05	MEF2C, GATA3, TFE3, POU2F2, TFEB, AIRE, PAX5	215	57	16792	9.59151367	0.134630422	0.002724557	0.142258713
GO:0042753~positive regulation of circadian rhythm	4	1.86047	1.09E-04	NKX2-1, RORC, RORA, ARNTL	215	8	16792	39.0511628	0.170553806	0.003456919	0.183932671
GO:0043433~negative regulation of sequence-specific DNA binding transcription factor activity	7	3.25581	1.13E-04	FOXA2, CEBPG, ESR1, PBX1, BHLHE40, TCF7L2, TWIST1	215	60	16792	9.11193798	0.175735646	0.003507724	0.190091056
GO:0007275~multicellular organism development	19	8.83721	1.28E-04	HLF, CDX1, CDX2, NFE2, FOXA3, TBX21, CREM, OTX2, MITF, TP53, PAX6, PAX5, DBP, LHX3, HOXA10, TGIF1, NKX3-1, HOXA9, ALX1	215	521	16792	2.84826139	0.197706047	0.003925849	0.216634942
GO:0001756~somitogenesis	6	2.7907	1.30E-04	MYF6, T, SMAD3, LEF1, NKX3-1, FOXC1	215	39	16792	12.0157424	0.19958114	0.003897999	0.218933601
GO:0060395~SMAD protein signal transduction	7	3.25581	1.35E-04	T, FOS, HNF1A, JUN, SMAD4, SMAD3, SMAD2	215	62	16792	8.8180045	0.207332874	0.003998054	0.2284933
GO:0030326~embryonic limb morphogenesis	6	2.7907	1.47E-04	HNF1A, HOXA10, LEF1, PBX1, PBX2, ALX1	215	40	16792	11.7153488	0.222606669	0.004258852	0.247603406
GO:0003309~type B pancreatic cell differentiation	4	1.86047	1.62E-04	GATA6, PAX6, PDX1, NKX6-1	215	9	16792	34.7121447	0.242602311	0.004620407	0.27319116
GO:0060487~lung epithelial cell differentiation	4	1.86047	2.29E-04	ASCL1, FOXA1, SOX9, GRHL2	215	10	16792	31.2409302	0.325131537	0.006425782	0.386401095
GO:0042118~endothelial cell activation	4	1.86047	2.29E-04	SMAD4, HOXA9, TCF4, FOXP1	215	10	16792	31.2409302	0.325131537	0.006425782	0.386401095
GO:0090090~negative regulation of canonical Wnt signaling pathway	10	4.65116	2.47E-04	SOX10, SOX2, FOXO1, LEF1, SOX17, FOXO3, ISL1, SNAI2, SOX9, TCF7L2	215	163	16792	4.79155372	0.345243314	0.006807235	0.416067312
GO:0007568~aging	10	4.65116	2.70E-04	SREBF1, FOS, JUN, CREB1, JUND, PAX5, NFKB2, FOXO3, NFE2L2, FOXO4	215	165	16792	4.73347428	0.370460823	0.007318599	0.454566547
GO:0060333~interferon-gamma-mediated signaling pathway	7	3.25581	2.87E-04	IRF9, IRF7, IRF8, IRF1, IRF2, IRF3, IRF4	215	71	16792	7.70022928	0.388773754	0.007662494	0.483494107
GO:0042593~glucose homeostasis	8	3.72093	3.09E-04	CEBPA, HNF1A, FOXA1, PAX6, SOX4, NEUROD1, PDX1, TCF7L2	215	101	16792	6.18632282	0.411437508	0.008121795	0.52050648
GO:0001501~skeletal system development	9	4.18605	3.74E-04	VDR, TBX3, HOXA13, TEAD4, HOXA10, NKX3-2, SOX4, FOXC1, SOX9	215	137	16792	5.13080971	0.473452168	0.009671309	0.629493374
GO:0045670~regulation of osteoclast differentiation	4	1.86047	4.12E-04	ESRRA, CEBPB, TFE3, MITF	215	12	16792	26.0341085	0.507141513	0.010504634	0.694158994

Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0001843~neural tube closure	7	3.25581	4.46E-04	T, RARG, LHX2, RARA, GRHL2, ALX1, TWIST1	215	77	16792	7.10021142	0.534702816	0.011188104	0.750404669
GO:0072539~T-helper 17 cell differentiation	3	1.39535	4.81E-04	BATF, RORC, RORA	215	3	16792	78.1023256	0.561747552	0.011884756	0.808899172
GO:0003215~cardiac right ventricle morphogenesis	4	1.86047	5.31E-04	FOXH1, GATA3, SOX4, ISL1	215	13	16792	24.0314848	0.597968048	0.012933121	0.893104477
GO:0071222~cellular response to lipopolysaccharide	8	3.72093	6.10E-04	MEF2C, CEBPB, NR1D1, CEBPE, IRF8, RARA, GFI1, NR1H3	215	113	16792	5.52936818	0.649009598	0.014638232	1.025492875
GO:0016569~covalent chromatin modification	8	3.72093	6.10E-04	GFI1B, FOXA2, IKZF1, FOXA3, FOXA1, SMARCA5, NR3C1, SMARCA1	215	113	16792	5.52936818	0.649009598	0.014638232	1.025492875
GO:0030073~insulin secretion	5	2.32558	6.16E-04	HNF1B, HNF1A, NEUROD1, SMAD2, PDX1	215	31	16792	12.5971493	0.652468791	0.014571964	1.035143481
GO:0048665~neuron fate specification	4	1.86047	6.70E-04	ASCL1, FOXA1, SIX1, ISL1	215	14	16792	22.3149502	0.683031312	0.01561586	1.124791117
GO:0007219~Notch signaling pathway	8	3.72093	6.78E-04	CEBPA, ASCL1, HNF1B, ONECUT1, FOXA1, FOXC1, SNAI2, SOX9	215	115	16792	5.43320526	0.687364142	0.015589585	1.13818858
GO:0035115~embryonic forelimb morphogenesis	5	2.32558	6.97E-04	OSR2, TBX3, HOXA9, RUNX2, TWIST1	215	32	16792	12.2034884	0.697590515	0.015819841	1.170552851
GO:0006338~chromatin remodeling	7	3.25581	8.05E-04	HNF1A, GATA3, FOXA1, SMARCA5, ESR1, SMARCA1, SOX9	215	86	16792	6.35716604	0.7484945	0.017997786	1.349729742
GO:2000679~positive regulation of transcription regulatory region DNA binding	4	1.86047	8.29E-04	GATA3, NEUROD1, KLF4, TWIST1	215	15	16792	20.8272868	0.758975138	0.018308956	1.391061266
GO:0045787~positive regulation of cell cycle	5	2.32558	8.81E-04	ASCL1, TBX3, RARA, TCF3, FOSL1	215	34	16792	11.4856361	0.77956514	0.019199866	1.477715603
GO:0009612~response to mechanical stimulus	6	2.7907	9.18E-04	MEIS2, JUN, JUND, FOSB, FOSL1, JUNB	215	59	16792	7.94260938	0.793141892	0.019748248	1.539356473
GO:0010957~negative regulation of vitamin D biosynthetic process	3	1.39535	9.54E-04	GFI1, SNAI2, SNAI1	215	4	16792	58.5767442	0.805326788	0.020247619	1.598189683
GO:0051098~regulation of binding	3	1.39535	9.54E-04	SMAD4, SMAD3, SMAD2	215	4	16792	58.5767442	0.805326788	0.020247619	1.598189683
GO:0060534~trachea cartilage development	3	1.39535	9.54E-04	RARG, RARA, SOX9	215	4	16792	58.5767442	0.805326788	0.020247619	1.598189683
GO:0060913~cardiac cell fate determination	3	1.39535	9.54E-04	POU5F1, SOX17, ISL1	215	4	16792	58.5767442	0.805326788	0.020247619	1.598189683
GO:0045669~positive regulation of osteoblast differentiation	6	2.7907	9.92E-04	MEF2C, CEBPA, CEBPB, CEBPD, JUND, RUNX2	215	60	16792	7.81023256	0.817582175	0.020786542	1.661162299
GO:0060216~definitive hemopoiesis	4	1.86047	0.001011136	LYL1, HOXA9, MEIS1, CBFB	215	16	16792	19.5255814	0.823595486	0.020935973	1.693609866
GO:0007179~transforming growth factor beta receptor signaling pathway	7	3.25581	0.001145627	FOXH1, FOS, JUN, CREB1, SMAD4, SMAD3, SMAD2	215	92	16792	5.94256825	0.859967036	0.023406976	1.91683317
GO:0032331~negative regulation of chondrocyte differentiation	4	1.86047	0.001216325	RARG, NKX3-2, SNAI2, SOX9	215	17	16792	18.3770178	0.875974273	0.024542227	2.033981904
GO:0001657~ureteric bud development	5	2.32558	0.001348137	SIX1, SMAD3, SMAD2, FOXC1, RARA	215	38	16792	10.2766218	0.901096896	0.026851897	2.252050399
GO:0007283~spermatogenesis	14	6.51163	0.00136923	AR, FOXA3, CREM, YY1, SMAD4, HOXA10, HOXA9, PAX5, TBP, RARA, SOX17, ARNTL, SOX9, CLOCK	215	385	16792	2.84008457	0.904615425	0.026953787	2.286903987
GO:0071456~cellular response to hypoxia	7	3.25581	0.001428179	GATA6, TP53, NKX3-1, RORA, FOXO3, NFE2L2, TWIST1	215	96	16792	5.69496124	0.913799431	0.027780149	2.384248791
GO:0060324~face development	4	1.86047	0.001445947	RARG, SOX3, RARA, GRHL2	215	18	16792	17.3560724	0.91639006	0.027806013	2.413570066
GO:0060037~pharyngeal system development	4	1.86047	0.001445947	GATA3, SIX1, NKX3-1, ISL1	215	18	16792	17.3560724	0.91639006	0.027806013	2.413570066
GO:0032570~response to progesterone	5	2.32558	0.001487375	SREBF1, FOS, FOSB, FOSL1, NR1H3	215	39	16792	10.0131187	0.922132598	0.02827512	2.481908753
GO:0002062~chondrocyte differentiation	5	2.32558	0.001487375	MEF2C, MEF2D, OSR2, SIX2, RUNX2	215	39	16792	10.0131187	0.922132598	0.02827512	2.481908753
GO:0060509~Type I pneumocyte differentiation	3	1.39535	0.001576202	THRA, THRB, CREB1	215	5	16792	46.8613953	0.93315103	0.029611822	2.628280206
GO:0060850~regulation of transcription involved in cell fate commitment	3	1.39535	0.001576202	CEBPB, RORC, RORA	215	5	16792	46.8613953	0.93315103	0.029611822	2.628280206
GO:0048645~organ formation	3	1.39535	0.001576202	AR, GATA6, NKX3-2	215	5	16792	46.8613953	0.93315103	0.029611822	2.628280206
GO:0030218~erythrocyte differentiation	5	2.32558	0.001636326	THRA, PKNOX1, IKZF1, GATA3, NFE2L1	215	40	16792	9.7627907	0.939710567	0.030392297	2.727237569
GO:0032332~positive regulation of chondrocyte differentiation	4	1.86047	0.001701022	SOX5, SMAD3, SOX9, RUNX2	215	19	16792	16.4425949	0.946052092	0.031237955	2.833612643
GO:0030111~regulation of Wnt signaling pathway	4	1.86047	0.001701022	HNF1B, HNF1A, LEF1, TCF7L1	215	19	16792	16.4425949	0.946052092	0.031237955	2.833612643
GO:0009749~response to glucose	6	2.7907	0.001743441	SREBF1, HNF1B, HNF1A, NEUROD1, SMAD2, TCF7L2	215	68	16792	6.89138167	0.949843652	0.031666382	2.903300282
GO:0045599~negative regulation of fat cell differentiation	5	2.32558	0.001964748	GATA3, SMAD3, FOXO1, RORA, ARNTL	215	42	16792	9.2978959	0.965708485	0.035245352	3.266112268
GO:0001892~embryonic placenta development	4	1.86047	0.00198252	CEBPA, CEBPB, ESRRB, TFEB	215	20	16792	15.6204651	0.966739872	0.035191092	3.295191741
GO:0042474~middle ear morphogenesis	4	1.86047	0.00198252	OSR2, SIX1, NKX3-2, SIX2	215	20	16792	15.6204651	0.966739872	0.035191092	3.295191741
GO:1902895~positive regulation of pri-miRNA transcription from RNA polymerase II promoter	4	1.86047	0.00198252	FOS, JUN, TEAD1, SMAD3	215	20	16792	15.6204651	0.966739872	0.035191092	3.295191741
GO:0017015~regulation of transforming growth factor beta receptor signaling pathway	4	1.86047	0.00198252	SMAD4, SMAD3, SMAD2, ZEB1	215	20	16792	15.6204651	0.966739872	0.035191092	3.295191741
GO:0071300~cellular response to retinoic acid	6	2.7907	0.001983166	MEF2C, T, RARG, RARA, SOX9, KLF4	215	70	16792	6.69448505	0.966776811	0.03484215	3.2962497
GO:0043525~positive regulation of neuron apoptotic process	5	2.32558	0.002144904	ASCL1, JUN, TP53, FOXO3, ATF2	215	43	16792	9.08166577	0.974839041	0.037251942	3.56051912
GO:0030879~mammary gland development	4	1.86047	0.00229136	TBX3, LEF1, HOXA9, SOX9	215	21	16792	14.8766334	0.980438496	0.039349686	3.799231854
GO:0007492~endoderm development	4	1.86047	0.00229136	ONECUT1, SMAD4, SMAD3, NKX2-1	215	21	16792	14.8766334	0.980438496	0.039349686	3.799231854
GO:0060430~lung saccule development	3	1.39535	0.002344462	GATA6, CREB1, NKX2-1	215	6	16792	39.0511628	0.982145025	0.039845764	3.885646955
GO:2000675~negative regulation of type B pancreatic cell apoptotic process	3	1.39535	0.002344462	NEUROD1, PDX1, TCF7L2	215	6	16792	39.0511628	0.982145025	0.039845764	3.885646955
GO:0060536~cartilage morphogenesis	3	1.39535	0.002344462	MEF2C, SNAI2, SNAI1	215	6	16792	39.0511628	0.982145025	0.039845764	3.885646955
GO:0048341~paraxial mesoderm formation	3	1.39535	0.002344462	HNF1A, LEF1, FOXC1	215	6	16792	39.0511628	0.982145025	0.039845764	3.885646955
GO:0009791~post-embryonic development	6	2.7907	0.002386726	GATA3, ARID5B, NR4A2, SMAD2, PRDM1, FOXP2	215	73	16792	6.41936923	0.983396264	0.040152898	3.954372616
GO:0032870~cellular response to hormone stimulus	5	2.32558	0.002538804	FOS, JUN, JUND, FOSB, JUNB	215	45	16792	8.67803618	0.987216312	0.042245896	4.201285692

Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0071542~dopaminergic neuron differentiation	4	1.86047	0.002628416	FOXA2, FOXA1, OTX2, NR4A2	215	22	16792	14.2004228	0.9890418	0.04328682	4.346498831
GO:0030900~forebrain development	5	2.32558	0.002753217	OTX2, SOX2, NKX2-1, POU3F1, NR2F2	215	46	16792	8.48938322	0.991158321	0.044867893	4.548390529
GO:0045165~cell fate commitment	5	2.32558	0.002753217	ONECUT1, GATA6, GATA3, SMAD2, PRDM1	215	46	16792	8.48938322	0.991158321	0.044867893	4.548390529
GO:0003148~outflow tract septum morphogenesis	4	1.86047	0.002994512	GATA6, SMAD4, RARA, ISL1	215	23	16792	13.5830131	0.994161563	0.048251759	4.937599278
GO:0060740~prostate gland epithelium morphogenesis	3	1.39535	0.003254731	AR, RARG, FOXA1	215	7	16792	33.4724252	0.99626858	0.051854493	5.355658843
GO:0060290~transdifferentiation	3	1.39535	0.003254731	MEF2C, SMAD3, PDX1	215	7	16792	33.4724252	0.99626858	0.051854493	5.355658843
GO:0050728~negative regulation of inflammatory response	6	2.7907	0.003368066	GATA3, SMAD3, RORA, ISL1, KLF4, NR1H3	215	79	16792	5.9318222	0.996929706	0.053121729	5.537198144
GO:0030099~myeloid cell differentiation	4	1.86047	0.003390429	CEBPA, IRF8, GFI1, CBFB	215	24	16792	13.0170543	0.997045615	0.052979226	5.572980441
GO:0009954~proximal/distal pattern formation	4	1.86047	0.003390429	HOXA10, HOXA9, PBX1, PBX2	215	24	16792	13.0170543	0.997045615	0.052979226	5.572980441
GO:0051149~positive regulation of muscle cell differentiation	4	1.86047	0.003390429	MYF6, MEF2C, MEF2A, TCF3	215	24	16792	13.0170543	0.997045615	0.052979226	5.572980441
GO:0048589~developmental growth	4	1.86047	0.003816903	SOX10, GATA3, SMAD4, SMAD3	215	25	16792	12.4963721	0.998582009	0.058919863	6.252930572
GO:0071158~positive regulation of cell cycle arrest	4	1.86047	0.003816903	GATA6, TP53, FOXO4, TP73	215	25	16792	12.4963721	0.998582009	0.058919863	6.252930572
GO:0030878~thyroid gland development	4	1.86047	0.003816903	THRA, SIX1, SMAD3, NKX2-1	215	25	16792	12.4963721	0.998582009	0.058919863	6.252930572
GO:0048709~oligodendrocyte differentiation	4	1.86047	0.003816903	SOX10, NKX2-1, SOX9, NKX6-1	215	25	16792	12.4963721	0.998582009	0.058919863	6.252930572
GO:0060261~positive regulation of transcription initiation from RNA polymerase II promoter	3	1.39535	0.004303297	HNF1B, HNF1A, IRF8	215	8	16792	29.2883721	0.999386347	0.065603065	7.022789888
GO:0030850~prostate gland development	3	1.39535	0.004303297	AR, RARA, SOX9	215	8	16792	29.2883721	0.999386347	0.065603065	7.022789888
GO:0003139~secondary heart field specification	3	1.39535	0.004303297	FOXH1, MEF2C, ISL1	215	8	16792	29.2883721	0.999386347	0.065603065	7.022789888
GO:0007183~SMAD protein complex assembly	3	1.39535	0.004303297	SMAD4, SMAD3, SMAD2	215	8	16792	29.2883721	0.999386347	0.065603065	7.022789888
GO:2000177~regulation of neural precursor cell proliferation	3	1.39535	0.004303297	FOXO1, FOXO3, SMARCA1	215	8	16792	29.2883721	0.999386347	0.065603065	7.022789888
GO:2000543~positive regulation of gastrulation	3	1.39535	0.004303297	OSR2, FOXA2, OTX2	215	8	16792	29.2883721	0.999386347	0.065603065	7.022789888
GO:0007411~axon guidance	8	3.72093	0.004318148	FEZF1, LHX2, CREB1, GATA3, OTX2, SMAD4, PAX6, NKX2-1	215	159	16792	3.92967676	0.999401844	0.065243891	7.046201397
GO:0048863~stem cell differentiation	4	1.86047	0.004764259	ZNF281, PDX1, FOXO4, RUNX2	215	27	16792	11.5707149	0.99972265	0.071129408	7.746901677
GO:0001822~kidney development	6	2.7907	0.004845682	HNF1B, GATA3, ARID5B, SIX1, SIX2, TP73	215	86	16792	5.44899946	0.999758959	0.071681013	7.874255697
GO:0048511~rhythmic process	5	2.32558	0.004930367	HLF, CREM, NKX2-1, NFKB2, TWIST1	215	54	16792	7.23169681	0.999791692	0.072268768	8.006534068
GO:0051726~regulation of cell cycle	7	3.25581	0.005133292	JUN, FOXM1, JUND, IRF1, ARNTL, MECOM, JUNB	215	124	16792	4.40900225	0.999853175	0.074502104	8.32278
GO:0042475~odontogenesis of dentin-containing tooth	5	2.32558	0.00526491	DLX3, LEF1, FOXC1, NFIC, RUNX2	215	55	16792	7.10021142	0.999882981	0.075704313	8.527350155
GO:0006974~cellular response to DNA damage stimulus	9	4.18605	0.005280785	BATF, IRF7, YY1, TP53, FOXO1, IRF3, NFATC2, TP73, ATF2	215	208	16792	3.37942755	0.99988614	0.075295029	8.551994566
GO:0043388~positive regulation of DNA binding	4	1.86047	0.005286403	CEBPG, PDX1, IRF4, ISL1	215	28	16792	11.1574751	0.999887238	0.074752738	8.560714899
GO:0035116~embryonic hindlimb morphogenesis	4	1.86047	0.005286403	OSR2, RARG, TBX3, TWIST1	215	28	16792	11.1574751	0.999887238	0.074752738	8.560714899
GO:0051569~regulation of histone H3-K4 methylation	3	1.39535	0.005486523	GFI1B, GATA3, GFI1	215	9	16792	26.0341085	0.999920145	0.076846837	8.870827707
GO:0048617~embryonic foregut morphogenesis	3	1.39535	0.005486523	SMAD3, SMAD2, SOX17	215	9	16792	26.0341085	0.999920145	0.076846837	8.870827707
GO:2000188~regulation of cholesterol homeostasis	3	1.39535	0.005486523	NR1D1, RORA, NR1H3	215	9	16792	26.0341085	0.999920145	0.076846837	8.870827707
GO:0042733~embryonic digit morphogenesis	5	2.32558	0.005614314	OSR2, TBX3, SMAD4, GRHL2, TWIST1	215	56	16792	6.97342193	0.999935939	0.077935555	9.068339331
GO:0043066~negative regulation of apoptotic process	14	6.51163	0.005694098	SOX10, RARG, TBX3, TP53, FOXO1, LEF1, SMAD3, SOX9, ASCL1, GATA6, LHX3, NKX3-2, RARA, TWIST1	215	455	16792	2.40314848	0.999944175	0.078369345	9.191448004
GO:0071773~cellular response to BMP stimulus	4	1.86047	0.006430488	GATA6, GATA3, SMAD4, RUNX2	215	30	16792	10.4136434	0.999984332	0.087381348	10.32033268
GO:0001654~eye development	4	1.86047	0.006430488	MEIS2, SOX2, PAX6, FOXC1	215	30	16792	10.4136434	0.999984332	0.087381348	10.32033268
GO:0048706~embryonic skeletal system development	4	1.86047	0.006430488	T, NKX3-2, HOXA9, PBX1	215	30	16792	10.4136434	0.999984332	0.087381348	10.32033268
GO:0030217~T cell differentiation	4	1.86047	0.006430488	PKNOX1, TBX21, SOX4, RUNX2	215	30	16792	10.4136434	0.999984332	0.087381348	10.32033268
GO:0009653~anatomical structure morphogenesis	6	2.7907	0.006436816	SOX10, ONECUT1, POU5F1, GATA3, SIX2, NFE2L1	215	92	16792	5.09362993	0.999984502	0.08677886	10.32997567
GO:0090009~primitive streak formation	3	1.39535	0.006800837	T, FOXA2, OTX2	215	10	16792	23.4306977	0.999991733	0.090762179	10.88307534
GO:0043524~negative regulation of neuron apoptotic process	7	3.25581	0.006924931	MEF2C, CEBPB, JUN, SIX1, NR4A2, ISL1, TP73	215	132	16792	4.14178999	0.999993328	0.091635382	11.07089056
GO:0008283~cell proliferation	12	5.5814	0.007444335	AR, NANOG, GFI1B, SIX1, TP53, SMAD4, SIX2, IRF2, FOXC1, ZEB1, GRHL2, TCF7L2	215	366	16792	2.56073199	0.99999728	0.097438406	11.85297274
GO:0001947~heart looping	5	2.32558	0.007594005	FOXH1, MEF2C, TBX3, SMAD3, SOX17	215	61	16792	6.40182997	0.9999979	0.098555958	12.07713095
GO:0007498~mesoderm development	4	1.86047	0.007711013	T, IKZF1, LHX2, SMAD4	215	32	16792	9.7627907	0.999998284	0.099254547	12.25199862
GO:0030857~negative regulation of epithelial cell differentiation	3	1.39535	0.008242738	TBX3, ZEB1, SOX9	215	11	16792	21.3006342	0.999999316	0.104969504	13.04254401
GO:0048646~anatomical structure formation involved in morphogenesis	3	1.39535	0.008242738	GATA3, FOXA1, NKX2-1	215	11	16792	21.3006342	0.999999316	0.104969504	13.04254401
GO:0010941~regulation of cell death	3	1.39535	0.008242738	JUN, JUND, JUNB	215	11	16792	21.3006342	0.999999316	0.104969504	13.04254401
GO:0009611~response to wounding	5	2.32558	0.008499427	POU5F1, SOX2, PAX6, HOXB13, PDX1	215	63	16792	6.19859727	0.999999561	0.107277238	13.42177598
GO:0034599~cellular response to oxidative stress	5	2.32558	0.008977481	NR4A2, FOXO1, NFE2L1, FOXO3, NFE2L2	215	64	16792	6.10174419	0.999999808	0.11216425	14.12390754
GO:0001837~epithelial to mesenchymal transition	4	1.86047	0.009131531	LEF1, SNAI2, SOX9, SNAI1	215	34	16792	9.18850889	0.999999853	0.113164299	14.34902056
GO:0007050~cell cycle arrest	7	3.25581	0.009438799	SOX2, TP53, IRF1, SMAD3, FOXO4, TCF7L2, TP73	215	141	16792	3.87742042	0.999999914	0.115926703	14.79637604

Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0042127~regulation of cell proliferation	8	3.72093	0.009651727	ESRRA, TCF7, JUN, FOXM1, JUND, TFAP2C, SOX9, JUNB	215	185	16792	3.37739786	0.99999994	0.117556912	15.10508972
GO:1990440~positive regulation of transcription from RNA polymerase II promoter in response to endoplasmic reticulum stress	3	1.39535	0.009808791	ATF3, CEBPB, TP53	215	12	16792	19.5255814	0.999999954	0.118524113	15.33213491
GO:0060707~trophoblast giant cell differentiation	3	1.39535	0.009808791	PRDM1, NR2F2, SNAI1	215	12	16792	19.5255814	0.999999954	0.118524113	15.33213491
GO:0030183~B cell differentiation	5	2.32558	0.009985464	ONECUT1, LYL1, CEBPG, POU1F1, TCF3	215	66	16792	5.91684285	0.999999966	0.119697752	15.5868427
GO:0050796~regulation of insulin secretion	5	2.32558	0.01051585	HNF1A, NEUROD1, ARNTL, CLOCK, NKX6-1	215	67	16792	5.82853176	0.999999987	0.124806371	16.34717118
GO:0048536~spleen development	4	1.86047	0.010695052	ONECUT1, NKX3-2, PBX1, NFKB2	215	36	16792	8.67803618	0.99999999	0.125938386	16.60260562
GO:0061029~eyelid development in camera-type eye	3	1.39535	0.011495629	OSR2, JUN, TWIST1	215	13	16792	18.0236136	0.999999998	0.133843645	17.73480824
GO:0043518~negative regulation of DNA damage response, signal transduction by p53 class mediator	3	1.39535	0.011495629	SNAI2, SNAI1, TWIST1	215	13	16792	18.0236136	0.999999998	0.133843645	17.73480824
GO:0014032~neural crest cell development	3	1.39535	0.011495629	FOXC1, SNAI2, SOX9	215	13	16792	18.0236136	0.999999998	0.133843645	17.73480824
GO:0006359~regulation of transcription from RNA polymerase III promoter	3	1.39535	0.011495629	ZNF76, TBP, ZNF143	215	13	16792	18.0236136	0.999999998	0.133843645	17.73480824
GO:0003198~epithelial to mesenchymal transition involved in endocardial cushion formation	3	1.39535	0.011495629	SMAD4, SNAI2, SNAI1	215	13	16792	18.0236136	0.999999998	0.133843645	17.73480824
GO:0071347~cellular response to interleukin-1	5	2.32558	0.012818386	CEBPB, YY1, NKX3-1, RORA, SOX9	215	71	16792	5.50016377	1	0.147154865	19.57386963
GO:0048704~embryonic skeletal system morphogenesis	4	1.86047	0.013313802	OSR2, SIX1, ZEB1, ALX1	215	39	16792	8.01049493	1	0.151418808	20.25264767
GO:0009615~response to virus	6	2.7907	0.013346032	MEF2C, BATF3, IRF7, TBX21, GATA3, FOSL1	215	110	16792	4.26012685	1	0.150767566	20.296619
GO:0043010~camera-type eye development	4	1.86047	0.014260624	PKNOX1, FOXC1, GRHL2, FOXP2	215	40	16792	7.81023256	1	0.159259359	21.53493968
GO:0043065~positive regulation of apoptotic process	10	4.65116	0.015003482	RARG, CREB1, TP53, FOXO1, SOX4, NEUROD1, NR4A1, FOXO3, FOSL1, TP73	215	300	16792	2.60341085	1	0.165815235	22.52739365
GO:0048485~sympathetic nervous system development	3	1.39535	0.015218507	ASCL1, GATA3, SOX4	215	15	16792	15.6204651	1	0.16693325	22.81245433
GO:0030330~DNA damage response, signal transduction by p53 class mediator	3	1.39535	0.015218507	BATF, TP53, FOXO3	215	15	16792	15.6204651	1	0.16693325	22.81245433
GO:0060441~epithelial tube branching involved in lung morphogenesis	3	1.39535	0.015218507	FOXA1, NKX2-1, SOX9	215	15	16792	15.6204651	1	0.16693325	22.81245433
GO:0032526~response to retinoic acid	4	1.86047	0.015244726	SREBF1, ASCL1, RARG, RARA	215	41	16792	7.61973908	1	0.166145888	22.84714578
GO:0032024~positive regulation of insulin secretion	4	1.86047	0.015244726	SOX4, ISL1, TCF7L2, NKX6-1	215	41	16792	7.61973908	1	0.166145888	22.84714578
GO:0030154~cell differentiation	13	6.04651	0.01564838	SRY, CDX1, NANOG, CDX2, OSR2, NR1D1, ONECUT1, FOXK1, FOXA3, CREM, YY1, TP53, MECOM	215	462	16792	2.19768449	1	0.169118637	23.37938564
GO:0030509~BMP signaling pathway	5	2.32558	0.016118528	T, SMAD4, LEF1, ZNF8, RUNX2	215	76	16792	5.13831089	1	0.172694723	23.99494657
GO:0030324~lung development	5	2.32558	0.016118528	CEBPA, SREBF1, LHX3, NKX2-1, SMAD2	215	76	16792	5.13831089	1	0.172694723	23.99494657
GO:0001658~branching involved in ureteric bud morphogenesis	4	1.86047	0.016266263	SIX1, SMAD4, PBX1, SOX9	215	42	16792	7.43831672	1	0.173074495	24.18741255
GO:0070848~response to growth factor	3	1.39535	0.017248131	MEIS2, GATA6, SOX2	215	16	16792	14.644186	1	0.18148144	25.45496438
GO:0006978~DNA damage response, signal transduction by p53 class mediator resulting in transcription of p21 class mediator	3	1.39535	0.017248131	FOXM1, TP53, TP73	215	16	16792	14.644186	1	0.18148144	25.45496438
GO:0003203~endocardial cushion morphogenesis	3	1.39535	0.017248131	ISL1, SOX9, TWIST1	215	16	16792	14.644186	1	0.18148144	25.45496438
GO:0048538~thymus development	4	1.86047	0.017325365	MAFB, GATA3, SIX1, PBX1	215	43	16792	7.26533261	1	0.181124084	25.55381901
GO:1904837~beta-catenin-TCF complex assembly	4	1.86047	0.017325365	TCF7, LEF1, TCF7L2, TCF7L1	215	43	16792	7.26533261	1	0.181124084	25.55381901
GO:0035264~multicellular organism growth	5	2.32558	0.019109759	AR, RARG, ARID5B, RARA, GRHL2	215	80	16792	4.88139535	1	0.196792273	27.803699
GO:0001503~ossification	5	2.32558	0.019109759	THRA, FOXC1, SOX9, RUNX2, TWIST1	215	80	16792	4.88139535	1	0.196792273	27.803699
GO:0002076~osteoblast development	3	1.39535	0.019385706	JUND, SMAD3, RUNX2	215	17	16792	13.7827633	1	0.198182912	28.14586565
GO:0021542~dentate gyrus development	3	1.39535	0.019385706	MEF2C, LEF1, NEUROD1	215	17	16792	13.7827633	1	0.198182912	28.14586565
GO:0042789~mRNA transcription from RNA polymerase II promoter	3	1.39535	0.019385706	SREBF1, POU5F1, SOX17	215	17	16792	13.7827633	1	0.198182912	28.14586565
GO:0048663~neuron fate commitment	3	1.39535	0.019385706	SMAD4, PAX6, OLIG2	215	17	16792	13.7827633	1	0.198182912	28.14586565
GO:0021522~spinal cord motor neuron differentiation	3	1.39535	0.019385706	SOX4, OLIG2, ISL1	215	17	16792	13.7827633	1	0.198182912	28.14586565
GO:0009948~anterior/posterior axis specification	3	1.39535	0.019385706	CDX1, CDX2, SIX2	215	17	16792	13.7827633	1	0.198182912	28.14586565
GO:0030308~negative regulation of cell growth	6	2.7907	0.019398716	FOXK1, TP53, SMAD4, SMAD3, SOX17, ESR2	215	121	16792	3.87284259	1	0.197143962	28.16196055
GO:0001755~neural crest cell migration	4	1.86047	0.019556661	SOX10, LEF1, ISL1, ALX1	215	45	16792	6.94242894	1	0.197439381	28.3570796
GO:0048666~neuron development	4	1.86047	0.020728994	MEF2C, SOX3, NEUROD1, PBX3	215	46	16792	6.79150657	1	0.206868958	29.78982957
GO:0001708~cell fate specification	3	1.39535	0.021628176	FOXA2, POU1F1, SOX9	215	18	16792	13.0170543	1	0.213671201	30.87043745
GO:0051412~response to corticosterone	3	1.39535	0.021628176	FOS, FOSB, FOSL1	215	18	16792	13.0170543	1	0.213671201	30.87043745
GO:0048557~embryonic digestive tract morphogenesis	3	1.39535	0.021628176	HNF1B, TCF7, SIX2	215	18	16792	13.0170543	1	0.213671201	30.87043745
GO:0031668~cellular response to extracellular stimulus	3	1.39535	0.021628176	FOS, NR4A2, FOSL1	215	18	16792	13.0170543	1	0.213671201	30.87043745
GO:0042462~eye photoreceptor cell development	3	1.39535	0.021628176	PAX6, NR2E3, PRDM1	215	18	16792	13.0170543	1	0.213671201	30.87043745
GO:0045600~positive regulation of fat cell differentiation	4	1.86047	0.021939172	CEBPA, CEBPB, CREB1, SNAI2	215	47	16792	6.64700643	1	0.215196584	31.24052793

Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0043011~myeloid dendritic cell differentiation	3	1.39535	0.023972547	BATF, BATF3, IRF4	215	19	16792	12.3319461	1	0.231548651	33.61471944
GO:0042832~defense response to protozoan	3	1.39535	0.023972547	BATF, IRF8, IRF4	215	19	16792	12.3319461	1	0.231548651	33.61471944
GO:0045638~negative regulation of myeloid cell differentiation	3	1.39535	0.023972547	MEIS2, HOXA9, MEIS1	215	19	16792	12.3319461	1	0.231548651	33.61471944
GO:0042752~regulation of circadian rhythm	4	1.86047	0.024473098	NR1D1, CREB1, CREM, NR1H3	215	49	16792	6.37570005	1	0.234521858	34.18722357
GO:0007493~endodermal cell fate determination	2	0.93023	0.025326663	GATA6, SOX17	215	2	16792	78.1023256	1	0.240402855	35.15278567
GO:0032916~positive regulation of transforming growth factor beta3 production	2	0.93023	0.025326663	CREB1, SMAD3	215	2	16792	78.1023256	1	0.240402855	35.15278567
GO:0035565~regulation of pronephros size	2	0.93023	0.025326663	HNF1B, HNF1A	215	2	16792	78.1023256	1	0.240402855	35.15278567
GO:0048340~paraxial mesoderm morphogenesis	2	0.93023	0.025326663	SMAD3, SMAD2	215	2	16792	78.1023256	1	0.240402855	35.15278567
GO:0036315~cellular response to sterol	2	0.93023	0.025326663	RORC, RORA	215	2	16792	78.1023256	1	0.240402855	35.15278567
GO:0035947~regulation of gluconeogenesis by regulation of transcription from RNA polymerase II promoter	2	0.93023	0.025326663	NR1D1, FOXO1	215	2	16792	78.1023256	1	0.240402855	35.15278567
GO:0034124~regulation of MyD88-dependent toll-like receptor signaling pathway	2	0.93023	0.025326663	IRF7, IRF1	215	2	16792	78.1023256	1	0.240402855	35.15278567
GO:0008050~female courtship behavior	2	0.93023	0.025326663	THRA, THRB	215	2	16792	78.1023256	1	0.240402855	35.15278567
GO:0007423~sensory organ development	2	0.93023	0.025326663	MAFB, SOX3	215	2	16792	78.1023256	1	0.240402855	35.15278567
GO:0021892~cerebral cortex GABAergic interneuron differentiation	2	0.93023	0.025326663	ASCL1, NKX2-1	215	2	16792	78.1023256	1	0.240402855	35.15278567
GO:0045655~regulation of monocyte differentiation	2	0.93023	0.025326663	IRF7, FOXP1	215	2	16792	78.1023256	1	0.240402855	35.15278567
GO:0030318~melanocyte differentiation	3	1.39535	0.026415883	MEF2C, SOX10, MITF	215	20	16792	11.7153488	1	0.248113495	36.36557468
GO:0032753~positive regulation of interleukin-4 production	3	1.39535	0.026415883	CEBPB, GATA3, RARA	215	20	16792	11.7153488	1	0.248113495	36.36557468
GO:0032481~positive regulation of type I interferon production	4	1.86047	0.0271583	IRF7, IRF1, IRF3, NFKB2	215	51	16792	6.12567259	1	0.252847118	37.17995343
GO:0071333~cellular response to glucose stimulus	4	1.86047	0.0285575	MEF2C, SOX4, NEUROD1, FOXO3	215	52	16792	6.0078712	1	0.262759202	38.68819182
GO:0097150~neuronal stem cell population maintenance	3	1.39535	0.028955306	SOX2, FOXO1, FOXO3	215	21	16792	11.1574751	1	0.26454438	39.11073835
GO:0032355~response to estradiol	5	2.32558	0.02901885	ESRRA, FOXA1, ESR1, RARA, NR2F2	215	91	16792	4.29133657	1	0.26367454	39.17797993
GO:0090103~cochlea morphogenesis	3	1.39535	0.031587995	SIX1, ZEB1, SOX9	215	22	16792	10.6503171	1	0.282233766	41.83885283
GO:0010719~negative regulation of epithelial to mesenchymal transition	3	1.39535	0.031587995	FOXA2, FOXA1, NKX2-1	215	22	16792	10.6503171	1	0.282233766	41.83885283
GO:0001502~cartilage condensation	3	1.39535	0.031587995	THRA, SOX9, ALX1	215	22	16792	10.6503171	1	0.282233766	41.83885283
GO:0030182~neuron differentiation	5	2.32558	0.033252419	MEF2C, CEBPB, POU3F2, SMARCA1, RUNX2	215	95	16792	4.11064871	1	0.293400391	43.50382258
GO:0021987~cerebral cortex development	4	1.86047	0.034529458	ASCL1, LHX2, PAX5, FOXP2	215	56	16792	5.57873754	1	0.301428438	44.75076963
GO:0007548~sex differentiation	3	1.39535	0.037122162	SRY, AR, PBX1	215	24	16792	9.7627907	1	0.318789064	47.20316222
GO:0071353~cellular response to interleukin-4	3	1.39535	0.037122162	TCF7, GATA3, LEF1	215	24	16792	9.7627907	1	0.318789064	47.20316222
GO:0060041~retina development in camera-type eye	4	1.86047	0.037738699	LHX2, PAX6, NR2E3, SOX9	215	58	16792	5.38636728	1	0.32164652	47.77106839
GO:0002068~glandular epithelial cell development	2	0.93023	0.03774954	RARG, RARA	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0060743~epithelial cell maturation involved in prostate gland development	2	0.93023	0.03774954	FOXA1, HOXB13	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0019218~regulation of steroid metabolic process	2	0.93023	0.03774954	RORC, RORA	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0039530~MDA-5 signaling pathway	2	0.93023	0.03774954	IRF7, IRF3	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0072193~ureter smooth muscle cell differentiation	2	0.93023	0.03774954	SIX1, SOX9	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0021912~regulation of transcription from RNA polymerase II promoter involved in spinal cord motor neuron fate specification	2	0.93023	0.03774954	PAX6, NKX6-1	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0070384~Harderian gland development	2	0.93023	0.03774954	RARG, SOX9	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:1990785~response to water-immersion restraint stress	2	0.93023	0.03774954	FOXO3, FOXO4	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0071899~negative regulation of estrogen receptor binding	2	0.93023	0.03774954	LEF1, NKX3-1	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:2000074~regulation of type B pancreatic cell development	2	0.93023	0.03774954	ARNTL, CLOCK	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0021986~habenula development	2	0.93023	0.03774954	PAX6, NR4A2	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0009786~regulation of asymmetric cell division	2	0.93023	0.03774954	POU5F1, PAX6	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0060745~mammary gland branching involved in pregnancy	2	0.93023	0.03774954	VDR, ESR1	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0051594~detection of glucose	2	0.93023	0.03774954	PDX1, NKX6-1	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0072095~regulation of branch elongation involved in ureteric bud branching	2	0.93023	0.03774954	HNF1B, SIX1	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0071895~odontoblast differentiation	2	0.93023	0.03774954	DLX3, LEF1	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0021530~spinal cord oligodendrocyte cell fate specification	2	0.93023	0.03774954	ASCL1, OLIG2	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0061138~morphogenesis of a branching epithelium	2	0.93023	0.03774954	SOX10, SOX9	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0001547~antral ovarian follicle growth	2	0.93023	0.03774954	ESR1, FOXO3	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:2001016~positive regulation of skeletal muscle cell differentiation	2	0.93023	0.03774954	MEF2C, ARNTL	215	3	16792	52.0682171	1	0.320182056	47.78100247
GO:0010468~regulation of gene expression	5	2.32558	0.039025818	ASCL1, NANOG, POU5F1, SOX2, TP73	215	100	16792	3.90511628	1	0.327612294	48.93819303
GO:0007338~single fertilization	4	1.86047	0.039398566	AR, SMAD4, HOXA10, HOXA9	215	59	16792	5.29507292	1	0.328654786	49.27158158
GO:0060612~adipose tissue development	3	1.39535	0.040018272	ARID5B, RORC, ATF2	215	25	16792	9.37227907	1	0.331381952	49.82132164
GO:0051090~regulation of sequence-specific DNA binding transcription factor activity	3	1.39535	0.040018272	FOS, JUN, ATF2	215	25	16792	9.37227907	1	0.331381952	49.82132164
GO:0001816~cytokine production	3	1.39535	0.040018272	MAF, BATF, NFATC2	215	25	16792	9.37227907	1	0.331381952	49.82132164

Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0050679~positive regulation of epithelial cell proliferation	4	1.86047	0.04109501	OSR2, SOX9, FOXP1, TWIST1	215	60	16792	5.20682171	1	0.337172266	50.76318931
GO:0002053~positive regulation of mesenchymal cell proliferation	3	1.39535	0.042996908	SOX9, FOXP1, FOXP2	215	26	16792	9.0118068	1	0.34835132	52.38635121
GO:0001958~endochondral ossification	3	1.39535	0.042996908	MEF2C, MEF2D, RUNX2	215	26	16792	9.0118068	1	0.34835132	52.38635121
GO:0045930~negative regulation of mitotic cell cycle	3	1.39535	0.042996908	SMAD3, NKX3-1, FOXC1	215	26	16792	9.0118068	1	0.34835132	52.38635121
GO:0060349~bone morphogenesis	3	1.39535	0.04605552	T, CDX1, OSR2	215	27	16792	8.67803618	1	0.366722774	54.89154619
GO:0032728~positive regulation of interferon-beta production	3	1.39535	0.04605552	IRF7, IRF1, IRF3	215	27	16792	8.67803618	1	0.366722774	54.89154619
GO:0030512~negative regulation of transforming growth factor beta receptor signaling pathway	4	1.86047	0.048242177	ONECUT1, SMAD3, NKX2-1, SMAD2	215	64	16792	4.88139535	1	0.378979385	56.60595666
GO:0009314~response to radiation	3	1.39535	0.049191606	JUN, JUND, JUNB	215	28	16792	8.36810631	1	0.383250256	57.33108867
GO:0001707~mesoderm formation	3	1.39535	0.049191606	SMAD3, SMAD2, SNAI1	215	28	16792	8.36810631	1	0.383250256	57.33108867
GO:0030858~positive regulation of epithelial cell differentiation	2	0.93023	0.050014809	PAX6, SOX9	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:0072676~lymphocyte migration	2	0.93023	0.050014809	TBX21, GATA3	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:0021913~regulation of transcription from RNA polymerase II promoter involved in ventral spinal cord interneuron specification	2	0.93023	0.050014809	PAX6, NKX6-1	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:1901203~positive regulation of extracellular matrix assembly	2	0.93023	0.050014809	SMAD3, SOX9	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:0003170~heart valve development	2	0.93023	0.050014809	PRDM1, SOX9	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:0072107~positive regulation of ureteric bud formation	2	0.93023	0.050014809	GATA3, SIX1	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:0033687~osteoblast proliferation	2	0.93023	0.050014809	OSR2, JUNB	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:0090403~oxidative stress-induced premature senescence	2	0.93023	0.050014809	TP53, ARNTL	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:2000741~positive regulation of mesenchymal stem cell differentiation	2	0.93023	0.050014809	SOX5, SOX9	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:0045064~T-helper 2 cell differentiation	2	0.93023	0.050014809	BATF, GATA3	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:0032909~regulation of transforming growth factor beta2 production	2	0.93023	0.050014809	SMAD4, SMAD3	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:0060486~Clara cell differentiation	2	0.93023	0.050014809	GATA6, NKX2-1	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:0045351~type I interferon biosynthetic process	2	0.93023	0.050014809	IRF7, IRF3	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:0048619~embryonic hindgut morphogenesis	2	0.93023	0.050014809	TCF7, HOXA13	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:0033152~immunoglobulin V(D)J recombination	2	0.93023	0.050014809	TCF3, FOXP1	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:0034121~regulation of toll-like receptor signaling pathway	2	0.93023	0.050014809	ESR1, GFI1	215	4	16792	39.0511628	1	0.386674587	57.95057155
GO:0071356~cellular response to tumor necrosis factor	5	2.32558	0.052184118	CEBPA, GATA3, NKX3-1, RORA, NFE2L2	215	110	16792	3.55010571	1	0.398194003	59.54275873
GO:0060412~ventricular septum morphogenesis	3	1.39535	0.052402715	TBX3, SMAD4, SOX4	215	29	16792	8.07955092	1	0.397822755	59.70001569
GO:0035690~cellular response to drug	4	1.86047	0.05797042	MEF2C, TP53, NKX3-1, NFE2L2	215	69	16792	4.52767105	1	0.428595524	63.51679844
GO:0032496~response to lipopolysaccharide	6	2.7907	0.059007689	FOS, JUN, JUND, NFKB2, JUNB, FOXP1	215	164	16792	2.85740216	1	0.432710178	64.18917675
GO:0030513~positive regulation of BMP signaling pathway	3	1.39535	0.059040445	GATA6, SMAD4, SMAD2	215	31	16792	7.55828957	1	0.431152799	64.21021887
GO:0071479~cellular response to ionizing radiation	3	1.39535	0.059040445	TP53, SNAI2, CLOCK	215	31	16792	7.55828957	1	0.431152799	64.21021887
GO:0010332~response to gamma radiation	3	1.39535	0.059040445	GATA3, TP53, TP73	215	31	16792	7.55828957	1	0.431152799	64.21021887
GO:0042795~snRNA transcription from RNA polymerase II promoter	4	1.86047	0.060019162	POU2F2, POU2F1, TBP, ZNF143	215	70	16792	4.46299003	1	0.434873566	64.83359367
GO:0033153~T cell receptor V(D)J recombination	2	0.93023	0.06212446	TCF7, LEF1	215	5	16792	31.2409302	1	0.444683868	66.14007245
GO:0060956~endocardial cell differentiation	2	0.93023	0.06212446	SMAD4, SOX17	215	5	16792	31.2409302	1	0.444683868	66.14007245
GO:0065004~protein-DNA complex assembly	2	0.93023	0.06212446	HNF1B, TCF4	215	5	16792	31.2409302	1	0.444683868	66.14007245
GO:0042634~regulation of hair cycle	2	0.93023	0.06212446	ARNTL, CLOCK	215	5	16792	31.2409302	1	0.444683868	66.14007245
GO:0072540~T-helper 17 cell lineage commitment	2	0.93023	0.06212446	BATF, IRF4	215	5	16792	31.2409302	1	0.444683868	66.14007245
GO:0060510~Type II pneumocyte differentiation	2	0.93023	0.06212446	GATA6, NKX2-1	215	5	16792	31.2409302	1	0.444683868	66.14007245
GO:0071864~positive regulation of cell proliferation in bone marrow	2	0.93023	0.06212446	MEF2C, LEF1	215	5	16792	31.2409302	1	0.444683868	66.14007245
GO:0071599~otic vesicle development	2	0.93023	0.06212446	GATA3, SIX1	215	5	16792	31.2409302	1	0.444683868	66.14007245
GO:0003211~cardiac ventricle formation	2	0.93023	0.06212446	MEF2C, SOX4	215	5	16792	31.2409302	1	0.444683868	66.14007245
GO:0030098~lymphocyte differentiation	2	0.93023	0.06212446	IKZF1, CBFB	215	5	16792	31.2409302	1	0.444683868	66.14007245
GO:0016202~regulation of striated muscle tissue development	2	0.93023	0.06212446	SMAD3, LEF1	215	5	16792	31.2409302	1	0.444683868	66.14007245
GO:0045595~regulation of cell differentiation	3	1.39535	0.062462407	NANOG, RUNX2, KLF4	215	32	16792	7.32209302	1	0.444772073	66.34548869
GO:0016573~histone acetylation	3	1.39535	0.069501233	HNF1A, CLOCK, ATF2	215	34	16792	6.89138167	1	0.479854469	70.36654935
GO:0040018~positive regulation of multicellular organism growth	3	1.39535	0.073113717	CREB1, POU3F2, POU1F1	215	35	16792	6.69448505	1	0.496069816	72.2502891
GO:0030501~positive regulation of bone mineralization	3	1.39535	0.073113717	MEF2C, OSR2, SMAD3	215	35	16792	6.69448505	1	0.496069816	72.2502891
GO:0001890~placenta development	3	1.39535	0.073113717	DLX3, HNF1A, LHX3	215	35	16792	6.69448505	1	0.496069816	72.2502891
GO:0003180~aortic valve morphogenesis	2	0.93023	0.07408046	GATA3, TWIST1	215	6	16792	26.0341085	1	0.498975134	72.73494916
GO:0071376~cellular response to corticotropin-releasing hormone stimulus	2	0.93023	0.07408046	NR4A2, NR4A1	215	6	16792	26.0341085	1	0.498975134	72.73494916
GO:0010694~positive regulation of alkaline phosphatase activity	2	0.93023	0.07408046	MEF2C, SMAD3	215	6	16792	26.0341085	1	0.498975134	72.73494916
GO:0002819~regulation of adaptive immune response	2	0.93023	0.07408046	IRF7, IRF1	215	6	16792	26.0341085	1	0.498975134	72.73494916
GO:0038092~nodal signaling pathway	2	0.93023	0.07408046	SMAD3, SMAD2	215	6	16792	26.0341085	1	0.498975134	72.73494916
GO:2000189~positive regulation of cholesterol homeostasis	2	0.93023	0.07408046	NR1D1, NR1H3	215	6	16792	26.0341085	1	0.498975134	72.73494916

Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0032873~negative regulation of stress-activated MAPK cascade	2	0.93023	0.07408046	FOXM1, FOXO1	215	6	16792	26.0341085	1	0.498975134	72.73494916
GO:0014033~neural crest cell differentiation	2	0.93023	0.07408046	MEF2C, SMAD4	215	6	16792	26.0341085	1	0.498975134	72.73494916
GO:0046632~alpha-beta T cell differentiation	2	0.93023	0.07408046	TCF7, LEF1	215	6	16792	26.0341085	1	0.498975134	72.73494916
GO:0071156~regulation of cell cycle arrest	2	0.93023	0.07408046	FOXM1, NEUROD1	215	6	16792	26.0341085	1	0.498975134	72.73494916
GO:0060527~prostate epithelial cord arborization involved in prostate glandular acinus morphogenesis	2	0.93023	0.07408046	ESR1, HOXB13	215	6	16792	26.0341085	1	0.498975134	72.73494916
GO:2000323~negative regulation of glucocorticoid receptor signaling pathway	2	0.93023	0.07408046	ARNTL, CLOCK	215	6	16792	26.0341085	1	0.498975134	72.73494916
GO:0055012~ventricular cardiac muscle cell differentiation	2	0.93023	0.07408046	MEF2C, RARA	215	6	16792	26.0341085	1	0.498975134	72.73494916
GO:0006352~DNA-templated transcription, initiation	3	1.39535	0.076785403	SMARCA5, TBP, TCF4	215	36	16792	6.50852713	1	0.510138551	74.04905956
GO:0035094~response to nicotine	3	1.39535	0.080514212	CREB1, PDX1, NKX6-1	215	37	16792	6.33262099	1	0.525691751	75.7631478
GO:0007507~heart development	6	2.7907	0.085144891	MEF2C, MEF2A, GATA3, SOX4, NKX3-1, FOXC1	215	183	16792	2.56073199	1	0.544648764	77.74367087
GO:2000049~positive regulation of cell-cell adhesion mediated by cadherin	2	0.93023	0.085884748	FOXA2, FOXA1	215	7	16792	22.3149502	1	0.546049602	78.04563105
GO:2000617~positive regulation of histone H3-K9 acetylation	2	0.93023	0.085884748	GATA3, SMAD4	215	7	16792	22.3149502	1	0.546049602	78.04563105
GO:0060575~intestinal epithelial cell differentiation	2	0.93023	0.085884748	CDX2, GATA6	215	7	16792	22.3149502	1	0.546049602	78.04563105
GO:0060576~intestinal epithelial cell development	2	0.93023	0.085884748	NKX3-2, PRDM1	215	7	16792	22.3149502	1	0.546049602	78.04563105
GO:0008595~anterior/posterior axis specification, embryo	2	0.93023	0.085884748	T, TBX3	215	7	16792	22.3149502	1	0.546049602	78.04563105
GO:0014807~regulation of somitogenesis	2	0.93023	0.085884748	CDX1, CDX2	215	7	16792	22.3149502	1	0.546049602	78.04563105
GO:0021902~commitment of neuronal cell to specific neuron type in forebrain	2	0.93023	0.085884748	ASCL1, PAX6	215	7	16792	22.3149502	1	0.546049602	78.04563105
GO:0051151~negative regulation of smooth muscle cell differentiation	2	0.93023	0.085884748	PRDM6, FOXO4	215	7	16792	22.3149502	1	0.546049602	78.04563105
GO:0021520~spinal cord motor neuron cell fate specification	2	0.93023	0.085884748	LHX3, ISL1	215	7	16792	22.3149502	1	0.546049602	78.04563105
GO:0003183~mitral valve morphogenesis	2	0.93023	0.085884748	SOX4, TWIST1	215	7	16792	22.3149502	1	0.546049602	78.04563105
GO:0021798~forebrain dorsal/ventral pattern formation	2	0.93023	0.085884748	PAX6, NKX2-1	215	7	16792	22.3149502	1	0.546049602	78.04563105
GO:0048793~pronephros development	2	0.93023	0.085884748	HNF1B, OSR2	215	7	16792	22.3149502	1	0.546049602	78.04563105
GO:0007596~blood coagulation	6	2.7907	0.086701106	MAFF, NFE2, GATA6, GATA3, IRF1, IRF2	215	184	16792	2.54681496	1	0.547766062	78.37433969
GO:0007223~Wnt signaling pathway, calcium modulating pathway	3	1.39535	0.088135098	LEF1, TCF7L2, NFATC1	215	39	16792	6.0078712	1	0.552109748	78.94055778
GO:0030890~positive regulation of B cell proliferation	3	1.39535	0.088135098	MEF2C, NFATC2, TCF3	215	39	16792	6.0078712	1	0.552109748	78.94055778
GO:0048468~cell development	3	1.39535	0.092023228	MAF, GATA6, ARID5B	215	40	16792	5.85767442	1	0.566629369	80.40644343
GO:0032731~positive regulation of interleukin-1 beta production	2	0.93023	0.097539241	SMAD3, ISL1	215	8	16792	19.5255814	1	0.587069188	82.32212887
GO:0001945~lymph vessel development	2	0.93023	0.097539241	FOXC1, NR2F2	215	8	16792	19.5255814	1	0.587069188	82.32212887
GO:2000020~positive regulation of male gonad development	2	0.93023	0.097539241	SRY, SOX9	215	8	16792	19.5255814	1	0.587069188	82.32212887
GO:0071499~cellular response to laminar fluid shear stress	2	0.93023	0.097539241	NFE2L2, KLF4	215	8	16792	19.5255814	1	0.587069188	82.32212887
GO:0033173~calcineurin-NFAT signaling cascade	2	0.93023	0.097539241	NFATC2, NFATC1	215	8	16792	19.5255814	1	0.587069188	82.32212887
GO:0002467~germinal center formation	2	0.93023	0.097539241	MEF2C, NFKB2	215	8	16792	19.5255814	1	0.587069188	82.32212887
GO:0040036~regulation of fibroblast growth factor receptor signaling pathway	2	0.93023	0.097539241	OTX2, RUNX2	215	8	16792	19.5255814	1	0.587069188	82.32212887
GO:0061469~regulation of type B pancreatic cell proliferation	2	0.93023	0.097539241	NR1D1, NR4A1	215	8	16792	19.5255814	1	0.587069188	82.32212887
GO:0009629~response to gravity	2	0.93023	0.097539241	FOS, FOSL1	215	8	16792	19.5255814	1	0.587069188	82.32212887
GO:0001541~ovarian follicle development	3	1.39535	0.099945304	CEBPB, SMAD4, FOXC1	215	42	16792	5.57873754	1	0.594626049	83.10128546
GO:0022008~neurogenesis	3	1.39535	0.099945304	ASCL1, PRDM6, NEUROD1	215	42	16792	5.57873754	1	0.594626049	83.10128546

C.2.8 Pathway enrichment of genes with enriched motifs in shared hypoDMRs

Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
hsa05202:Transcriptional misregulation in cancer	21	9.7674	4.32E-13	MAF, MEF2C, CEBPA, CEBPB, CEBPE, RXRB, TFE3, TP53, FOXO1, PAX5, MEIS1, ATF1, REL, LYL1, SIX1, HOXA10, RARA, PBX1, PBX3, TCF3, RUNX2	106	167	6879	8.1606033	4.75E-11	4.75E-11	4.88E-10
hsa04950:Maturity onset diabetes of the young	9	4.186	2.80E-09	HNF1B, HNF1A, FOXA2, ONECUT1, FOXA3, PAX6, NEUROD1, PDX1, NKX6-1	106	26	6879	22.464078	3.08E-07	1.54E-07	3.17E-06
hsa04550:Signaling pathways regulating pluripotency o	15	6.9767	2.05E-08	NANOG, HNF1A, ONECUT1, TBX3, ESRRB, SOX2, SMAD4, PAX6, SMAD3, SMAD2, ISL1, MEIS1, POU5F1, TCF3, KLF4	106	140	6879	6.9531671	2.26E-06	7.53E-07	2.32E-05
hsa04380:Osteoclast differentiation	14	6.5116	7.50E-08	FOSL2, CREB1, MITF, FOSB, NFKB2, JUNB, STAT2, IRF9, FOS, JUN, JUND, NFATC2, FOSL1, NFATC1	106	131	6879	6.9354746	8.25E-06	2.06E-06	8.48E-05
hsa05166:HTLV-I infection	18	8.3721	2.30E-07	CREM, CREB1, TP53, SMAD4, SMAD3, SMAD2, TBP, NFKB2, ATF1, ATF2, FOS, ATF3, JUN, NFATC2, NFATC3, FOSL1, TCF3, NFATC1	106	254	6879	4.5989452	2.53E-05	5.05E-06	2.59E-04
hsa05210:Colorectal cancer	10	4.6512	3.30E-07	FOS, TCF7, JUN, TP53, SMAD4, SMAD3, LEF1, SMAD2, TCF7L2, TCF7L1	106	62	6879	10.467133	3.63E-05	6.04E-06	3.72E-04
hsa04710:Circadian rhythm	7	3.2558	5.92E-06	NR1D1, CREB1, RORC, BHLHE40, RORA, ARNTL, CLOCK	106	31	6879	14.653987	6.51E-04	9.30E-05	0.00669
hsa04310:Wnt signaling pathway	12	5.5814	7.08E-06	TCF7, JUN, TP53, SMAD4, LEF1, SOX17, NFATC2, NFATC3, FOSL1, TCF7L2, TCF7L1, NFATC1	106	138	6879	5.6431501	7.79E-04	9.74E-05	0.008
hsa04520:Adherens junction	9	4.186	1.08E-05	TCF7, SMAD4, SMAD3, LEF1, SMAD2, SNAI2, SNAI1, TCF7L2, TCF7L1	106	71	6879	8.2262822	0.001189414	1.32E-04	0.01222
hsa05161:Hepatitis B	12	5.5814	1.14E-05	FOS, JUN, IRF7, CREB1, TP53, SMAD4, IRF3, NFATC2, NFATC3, NFATC1, STAT2, ATF2	106	145	6879	5.3707222	0.001252874	1.25E-04	0.01288
hsa04390:Hippo signaling pathway	12	5.5814	1.68E-05	TCF7, TEAD4, SOX2, SMAD4, TEAD1, SMAD3, LEF1, SMAD2, SNAI2, TCF7L2, TCF7L1, TP73	106	151	6879	5.157316	0.001842085	1.68E-04	0.01894
hsa05200:Pathways in cancer	19	8.8372	2.29E-05	CEBPA, AR, TCF7, RXRB, MITF, TP53, SMAD4, SMAD3, FOXO1, LEF1, SMAD2, NFKB2, MECOM, TCF7L2, TCF7L1, FOS, JUN, NKX3-1, RARA	106	393	6879	3.1374766	0.002518024	2.10E-04	0.02589

Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
hsa05321:Inflammatory bowel disease (IBD)	8	3.7209	4.80E-05	MAF, JUN, TBX21, SMAD3, RORC, SMAD2, RORA, NFATC1	106	64	6879	8.1120283	0.00526797	4.06E-04	0.05424
hsa05215:Prostate cancer	9	4.186	5.28E-05	AR, TCF7, CREB1, TP53, FOXO1, LEF1, NKX3-1, TCF7L2, TCF7L1	106	88	6879	6.6371141	0.005788207	4.15E-04	0.05961
hsa05216:Thyroid cancer	6	2.7907	6.68E-05	TCF7, RXRB, TP53, LEF1, TCF7L2, TCF7L1	106	29	6879	13.426805	0.007322578	4.90E-04	0.07546
hsa05213:Endometrial cancer	6	2.7907	0.00111	TCF7, TP53, LEF1, FOXO3, TCF7L2, TCF7L1	106	52	6879	7.4880261	0.11472544	0.0075872	1.24381
hsa05221:Acute myeloid leukemia	6	2.7907	0.00155	CEBPA, TCF7, LEF1, RARA, TCF7L2, TCF7L1	106	56	6879	6.9531671	0.156935585	0.0099916	1.73812
hsa05168:Herpes simplex infection	10	4.6512	0.00183	IRF9, FOS, JUN, IRF7, TP53, TBP, IRF3, ARNTL, CLOCK, STAT2	106	183	6879	3.5462419	0.182550753	0.0111357	2.04903
hsa04022:cGMP-PKG signaling pathway	9	4.186	0.00274	MEF2C, MEF2D, MEF2B, MEF2A, CREB1, NFATC2, NFATC3, NFATC1, ATF2	106	158	6879	3.6966205	0.260848372	0.0157821	3.05678
hsa05203:Viral carcinogenesis	10	4.6512	0.00394	IRF9, REL, JUN, IRF7, CREB1, TP53, TBP, IRF3, NFKB2, ATF2	106	205	6879	3.1656696	0.35228793	0.0214814	4.36281
hsa04010:MAPK signaling pathway	11	5.1163	0.00507	MEF2C, FOS, JUN, JUND, TP53, NR4A1, NFKB2, MECOM, NFATC3, NFATC1, ATF2	106	253	6879	2.8215751	0.427996764	0.0262498	5.57607
hsa05217:Basal cell carcinoma	5	2.3256	0.00903	TCF7, TP53, LEF1, TCF7L2, TCF7L1	106	54	6879	6.0089099	0.631393785	0.0443512	9.74297
hsa05160:Hepatitis C	7	3.2558	0.01577	IRF9, IRF7, TP53, IRF1, IRF3, STAT2, NR1H3	106	133	6879	3.4155909	0.826048114	0.0732231	16.4429
hsa04915:Estrogen signaling pathway	6	2.7907	0.01736	FOS, JUN, CREB1, ESR1, ESR2, ATF2	106	99	6879	3.9331046	0.854340698	0.0771331	17.9525
hsa05031:Amphetamine addiction	5	2.3256	0.01792	FOS, JUN, CREB1, FOSB, ATF2	106	66	6879	4.9163808	0.863180478	0.0764808	18.4784
hsa04916:Melanogenesis	6	2.7907	0.01806	TCF7, CREB1, MITF, LEF1, TCF7L2, TCF7L1	106	100	6879	3.8937736	0.865292675	0.0742045	18.6086
hsa04662:B cell receptor signaling pathway	5	2.3256	0.02077	FOS, JUN, NFATC2, NFATC3, NFATC1	106	69	6879	4.7026251	0.90057031	0.0819402	21.1079
hsa04917:Prolactin signaling pathway	5	2.3256	0.02281	FOS, IRF1, ESR1, ESR2, FOXO3	106	71	6879	4.5701568	0.920980689	0.0866582	22.9479
hsa04668:TNF signaling pathway	6	2.7907	0.02347	FOS, CEBPB, JUN, CREB1, JUNB, ATF2	106	107	6879	3.6390407	0.92663764	0.0861427	23.5335
hsa05133:Pertussis	5	2.3256	0.02725	FOS, JUN, IRF8, IRF1, IRF3	106	75	6879	4.3264151	0.952137583	0.0963509	26.8153
hsa04919:Thyroid hormone signaling pathway	6	2.7907	0.03082	THRA, THRB, RXRB, TP53, ESR1, FOXO1	106	115	6879	3.3858901	0.968059827	0.1051447	29.7933
hsa04925:Aldosterone synthesis and secretion	5	2.3256	0.03483	CREB1, NR4A2, NR4A1, ATF1, ATF2	106	81	6879	4.0059399	0.979747976	0.1147272	33.003
hsa05030:Cocaine addiction	4	1.8605	0.03836	JUN, CREB1, FOSB, ATF2	106	49	6879	5.2976511	0.986473876	0.1222539	35.7238
hsa05162:Measles	6	2.7907	0.05223	IRF9, IRF7, TP53, IRF3, TP73, STAT2	106	133	6879	2.9276493	0.997261485	0.159316	45.4489
hsa04068:FoxO signaling pathway	6	2.7907	0.05362	SMAD4, SMAD3, FOXO1, SMAD2, FOXO3, FOXO4	106	134	6879	2.9058012	0.997670281	0.1590308	46.3472
hsa04660:T cell receptor signaling pathway	5	2.3256	0.06614	FOS, JUN, NFATC2, NFATC3, NFATC1	106	100	6879	3.2448113	0.999461806	0.188679	53.8438
hsa05212:Pancreatic cancer	4	1.8605	0.07656	TP53, SMAD4, SMAD3, SMAD2	106	65	6879	3.9936139	0.999843295	0.2108412	59.3377
hsa04921:Oxytocin signaling pathway	6	2.7907	0.07881	MEF2C, FOS, JUN, NFATC2, NFATC3, NFATC1	106	150	6879	2.5958491	0.999880235	0.2115065	60.4451
hsa05412:Arrhythmogenic right ventricular cardiomyopa	4	1.8605	0.08219	TCF7, LEF1, TCF7L2, TCF7L1	106	67	6879	3.8744016	0.999920016	0.2148569	62.0517

C.3 DMV genes

C.3.1 List of genes encompassed within autosomal DMVs

ACTG1
ADAMTS1
ADAMTS9
ADAMTS9-AS2
ADCYAP1
ADGRB2
ADGRL2
ADM
AJAP1
ALCAM
ALOX12-AS1
ALOXE3
ALX4
ANKRD13B
ANP32A
ANXA2R
ANXA8
ANXA8L1
APC2
ARID2
ARID5B
ASCL1
ATXN1
BAHCC1
BARHL1
BARHL2
BARX1
BARX1-AS1
BARX2
BBC3
BCL11A
BCL11B
BCL6
BDNF
BHLHE22
BHLHE23
BHLHE40
BHLHE40-AS1
BHLHE41
BMP4
BOLA2B
BORCS6
BRD2
BRF1
BSX
BTBD11
C17orf102
C17orf49
C17orf82
C18orf65
C1orf94
C5orf38
C9orf172
CA11
CACNG8
CAND1.11
CASC11
CBLN1
CBX3
CBX4
CBX8
CCDC140
CCDC151
CCDC85B
CCND1
CCND2
CCND2-AS1
CCNO
CCR10
CDK5
CDK5R2
CDK6
CDKN1B
CDX2
CEBPA
CEBPA-AS1
CFL1
CHAT
CHD3
CHRM2
CIC
CNTNAP3
COL2A1
CPT1A
CRAMP1
CRLF1
CRMP1
CRNDE
CRYBA2
CSMD2
CSMD3
CTBP2
CTD-2194D22.4
CTDSP1
CXXC5
CYP1B1
CYP26A1
CYP26B1
DBP
DBX1
DDN
DDT
DDTL
DHH
DHRS3
DIGIT
DKFZP434H168
DKFZp686K1684
DLGAP2
DLX1
DLX2
DLX2-AS1
DLX5
DLX6
DLX6-AS1
DMRT2
DMRT3
DMRTA2
DNAH2
DUSP4
DUSP5P1
DUSP6
DUSP8
EBF2
EBF3
EFNA3
EFNA5
EFNB2
EGR2
EGR3
EHBP1L1
ELAVL2
EML2
EML2-AS1
EMX1
EMX2
EMX2OS
EN1
EN2
EOMES
EPHA4
EPHA7
ERF
ESAM
ESRP1
ESRRG
EVX1
EVX1-AS
EVX2
FAM27E3
FAM53B
FAM72A
FAM72B
FAM72D
FAM89B
FAM90A7P
FAT1
FAT4
FBXL15
FENDRR
FEV
FEZF1
FEZF1-AS1
FEZF2
FGF14
FGF5
FGF9
FHOD1
FIBP
FLJ31356
FLRT2
FOSL1
FOSL2
FOXA1
FOXA2
FOXB1
FOXB2
FOXC1
FOXD1
FOXD3
FOXD3-AS1
FOXD4
FOXD4L1
FOXD4L3
FOXD4L6
FOXE1
FOXF1
FOXF2
FOXG1
FOXG1-AS1
FOXL2
FOXL2NB
FOXO3
FOXP2
FSCN1
FST
FZD8
GAD1
GAD2
GADD45B
GAL3ST2
GATA2
GATA2-AS1
GATA3
GATA3-AS1
GATA4
GATA6
GATA6-AS1
GBX2
GDF6
GDNF
GDNF-AS1
GFI1
GFRA1
GIT1
GNAO1
GPAT2
GRID1
GRIK3
GSC
GSE1
GSTT2
GSTT2B
GSX1
GSX2
GTF2H2
GTF2IP4
HAND1
HAND2
HAND2-AS1
HELT
HES1
HES3
HES7
HGH1
HIST1H2AI
HIST1H2BL
HIST1H3H
HIST2H2AA3
HIST2H2AA4
HIST2H2AB
HIST2H2AC
HIST2H2BC
HIST2H2BE
HIST2H3A

HIST2H3C
HIST2H4A
HIST2H4B
HMGA2
HMX1
HMX2
HMX3
HNRNPA2B1
HNRNPL
HNRNPUL2
HNRNPUL2-BSCL2
HOTTIP
HOXA13
HOXB13
HOXB7
HOXB8
HOXC10
HOXC12
HOXC13
HOXC13-AS
HOXC4
HOXC5
HOXC6
HOXC9
HOXC-AS1
HOXC-AS3
HOXD10
HOXD11
HOXD12
HOXD8
HOXD9
HPSE2
HRAT92
HS3ST3B1
HTR1A
ICAM1
ICAM4
ICAM5
ID2
ID2-AS1
IER2
IER5L
IFT140
IGF1R
IGF2
IGF2-AS
IGFBP3
IGFBP5
IGSF21
IHH
ING1
INHBB
INS-IGF2
INSM1
IRAIN
IRF2BP2
IRX1
IRX2
IRX3
IRX4
IRX5
JADE2
JARID2
JARID2-AS1
JUN
JUNB
KCNC3
KCNK9
KCNN1
KDM6B
KIF26A
KIF26B
KISS1R
KLF10
KLF13
KLF4
KMT2A
KRTAP5-AS1
LARGE1
LARP1
LFNG
LHX1
LHX2
LHX3
LHX4
LHX5
LHX5-AS1
LHX8
LHX9
LIN28A
LINC00261
LINC00403
LINC00461
LINC00599
LINC00682
LINC00938
LINC00965
LINC00982
LINC01089
LINC01096
LINC01097
LINC01126
LINC01135
LINC01158
LINC01159
LINC01210
LINC01391
LINC01475
LINC01521
LINC01726
LINC01749
LINC01962
LINC-PINT
LINGO1
LMO4
LMX1A
LMX1B
LOC100287944
LOC100288748
LOC100505588
LOC100505824
LOC101927359
LOC101927497
LOC101927989
LOC101928433
LOC101928438
LOC101929076
LOC101929340
LOC102723471
LOC102724034
LOC102724623
LOC105371046
LOC105378933
LOC153684
LOC154761
LOC283856
LOC284395
LOC401463
LOC440461
LOC613038
LOC646743
LOC648987
LOC90768
LRFN5
LRP1
LTBP4
MAFA
MAFB
MAML3
MAP3K8
MARCKS
MEIS1
MEIS1-AS3
MEIS2
MEIS3
METAP1D
MGC12916
MIDN
MIR1199
MIR124-1
MIR124-2
MIR124-2HG
MIR124-3
MIR129-2
MIR1469
MIR3131
MIR3185
MIR3190
MIR3191
MIR3663
MIR3663HG
MIR4521
MIR4683
MIR4734
MIR4740
MIR497HG
MIR6068
MIR6720
MIR9-3
MIR9-3HG
MISP3
MKLN1
MKX
MKX-AS1
MLLT3
MLLT6
MMP17
MNX1
MNX1-AS1
MROH1
MSL2
MSX1
MSX2
MUS81
MXI1
MYADM
MYC
MYCN
MYCNOS
MYO3A
NAA38
NAB2
NACC2
NAT16
NBPF10
NBPF11
NBPF12
NDUFA4L2
NEAT1
NEFL
NEUROD1
NEUROD2
NEUROG3
NFIB
NFIL3
NKX1-1
NKX2-1
NKX2-1-AS1
NKX2-2
NKX2-3
NKX2-5
NKX2-6
NKX3-2
NKX6-1
NKX6-2
NOG
NOL4
NPIPB3
NPIPB4
NPR3
NPTX1
NR2E1
NR2F1
NR2F1-AS1
NR2F2
NR3C1
NR3C2
NR4A3
NR5A2
NRN1
NRSN2-AS1
NTF3
NTNG2
NTRK3
NTRK3-AS1
NXPH1
NXPH4
OLIG2
OLIG3
ONECUT1
ONECUT2
ONECUT3
OPLAH
OSR2
OTP
OTX1
OTX2
PATZ1
PAUPAR
PAX1
PAX2
PAX3
PAX5
PAX6

PAX7
PAX9
PCDH10
PCDH17
PCDH7
PCDH8
PCGF2
PDE10A
PDE4A
PDGFA
PDGFRA
PDPK1
PDX1
PDX1-AS1
PFKFB3
PHF12
PHF21B
PHOX2B
PHPT1
PITX2
PLEKHH3
POU3F2
POU3F3
POU4F1
POU4F2
POU4F3
PPARGC1A
PPP1R9B
PRAC1
PRAC2
PRDM13
PRDM16
PRDM6
PRKCG
PROX1
PROX1-AS1
PRR15
PRRX1
PSD
PSMB3
PTCH1
PTF1A
PTMA
PTPRG
PTPRN2
PYGM
RARG
RASA4
RASGRP2
RFX4
RGL3
RGMB
RGMB-AS1
RGS20
RHOU
RNA5S14
RNA5S15
RNA5S16
RNA5S17
RNA5S2
RNA5S3
RNA5S4
RNASEK
RNASEK-C17orf49
RNF219-AS1
RNF220
ROBO3
RORB
RORB-AS1
RPSAP52
RTBDN
RTN4RL2
RUNX1
RUNX2
SALL1
SALL3
SAMD4A
SATB2
SATB2-AS1
SCRT1
SEC1P
SEMA6A
SETD1B
SFMBT2
SFRP1
SFTA3
SGMS1
SGMS1-AS1
SHH
SHISA9
SHOX2
SIM1
SIM2
SIN3A
SIX1
SIX3
SIX3-AS1
SIX6
SKI
SKOR1
SKOR2
SLC11A1
SLC17A6
SLC18A3
SLC30A3
SLC32A1
SLC4A2
SLC6A5
SLC9A5
SLIT2
SLITRK1
SLITRK3
SLX1A
SLX1A-SULT1A3
SMAD7
SMARCA2
SNORA99
SNORD3B-1
SNORD3B-2
SNORD3D
SNTG2
SOCS2
SOCS2-AS1
SORCS3
SOX1
SOX12
SOX14
SOX2
SOX2-OT
SOX4
SOX9
SOX9-AS1
SP8
SP9
SPAG4
SPHK1
SPRY2
SSBP4
SSSCA1
SSSCA1-AS1
SSTR1
ST3GAL1
ST8SIA4
STAC3
STK3
STX10
SULT1A3
TARID
TBC1D3L
TBR1
TBX15
TBX18
TBX2
TBX2-AS1
TBX5
TBX5-AS1
TCAF2
TCF21
TCF4
TCF7L2
TCHH
TENM4
TFAP2A
TFAP2A-AS1
TFAP2B
TFAP2C
TFAP2D
TGIF1
TLE3
TLX1
TLX1NB
TLX3
TMEM132C
TMEM132E
TNIK
TNRC18
TOB2
TPPP
TRIM7
TRIM8
TSHZ1
TSPEAR
TTC6
TTC9C
TUBB2B
TUBB3
UBTF
UNCX
USP17L19
VAX1
VCAN
VGF
VSTM2B
VSX1
WASH7P
WNK1
WNT1
WNT10B
WNT11
WNT3A
WNT5A
WNT7B
WT1
WT1-AS
ZADH2
ZBTB7A
ZBTB7B
ZCCHC2
ZEB2
ZEB2-AS1
ZFP36L1
ZFP36L2
ZIC1
ZIC2
ZIC4
ZIC5
ZMIZ1
ZMIZ1-AS1
ZNF423
ZNF462
ZNF467
ZNF503
ZNF503-AS2
ZNF516
ZNF580
ZNF581

C.3.2 Ontology enrichment for DMV genes

3 gene IDs were not recognized in the input: *LINC01726*, *LINC01749*, *LINC01962*
All “Categories” were “GOTERM_BP_DIRECT”
Gene lists have been omitted for brevity. Only terms with a Benjamini Hochberg p-value<0.05 are reported for brevity.

Term	Count	%	PValue	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0045944~positive regulation of transcription from RNA polymerase II promoter	156	21.3114754	1.56E-68	526	981	16792	5.07659213	4.59E-65	4.59E-65	2.81E-65
GO:0000122~negative regulation of transcription from RNA polymerase II promoter	121	16.5300546	1.37E-54	526	720	16792	5.36499789	4.04E-51	2.02E-51	2.47E-51
GO:0006366~transcription from RNA polymerase II promoter	100	13.6612022	1.18E-50	526	513	16792	6.22299306	3.47E-47	1.16E-47	2.12E-47
GO:0006351~transcription, DNA-templated	162	22.1311475	7.74E-33	526	1955	16792	2.64536092	2.27E-29	5.69E-30	1.39E-29
GO:0006355~regulation of transcription, DNA-templated	128	17.4863388	2.58E-26	526	1504	16792	2.71693229	7.59E-23	1.52E-23	4.64E-23
GO:0045893~positive regulation of transcription, DNA-templated	71	9.69945355	6.81E-26	526	515	16792	4.40116653	2.00E-22	3.33E-23	1.22E-22
GO:0045892~negative regulation of transcription, DNA-templated	69	9.42622951	3.13E-25	526	499	16792	4.41433437	9.18E-22	1.31E-22	5.62E-22
GO:0009952~anterior/posterior pattern specification	24	3.27868852	1.83E-16	526	80	16792	9.57718631	6.52E-13	8.15E-14	4.00E-13
GO:0042472~inner ear morphogenesis	20	2.73224044	5.14E-16	526	52	16792	12.278444	1.63E-12	1.81E-13	9.99E-13
GO:0060021~palate development	22	3.00546448	9.88E-15	526	76	16792	9.24114469	2.90E-11	2.90E-12	1.78E-11
GO:0007411~axon guidance	29	3.96174863	9.03E-14	526	159	16792	5.82260803	2.65E-10	2.41E-11	1.62E-10
GO:0008285~negative regulation of cell proliferation	45	6.14754098	2.58E-13	526	396	16792	3.62772209	7.57E-10	6.31E-11	4.63E-10
GO:0006357~regulation of transcription from RNA polymerase II promoter	47	6.42076503	7.10E-13	526	441	16792	3.4023262	2.08E-09	1.60E-10	1.27E-09
GO:0009887~organ morphogenesis	21	2.86885246	5.74E-12	526	92	16792	7.28698959	1.69E-08	1.21E-09	1.03E-08
GO:0035019~somatic stem cell population maintenance	18	2.45901639	9.39E-12	526	65	16792	8.84047967	2.76E-08	1.84E-09	1.69E-08
GO:0001658~branching involved in ureteric bud morphogenesis	15	2.04918033	1.70E-11	526	42	16792	11.4014123	4.99E-08	3.12E-09	3.05E-08
GO:0021527~spinal cord association neuron differentiation	10	1.36612022	1.71E-11	526	13	16792	24.556888	5.01E-08	2.95E-09	3.07E-08
GO:0007399~nervous system development	35	4.78142077	2.52E-11	526	287	16792	3.89316517	7.40E-08	4.11E-09	4.52E-08
GO:0048665~neuron fate specification	10	1.36612022	4.65E-11	526	14	16792	22.8028246	1.36E-07	7.18E-09	8.35E-08
GO:0001764~neuron migration	21	2.86885246	7.44E-11	526	105	16792	6.38479087	2.19E-07	1.09E-08	1.34E-07
GO:0045665~negative regulation of neuron differentiation	16	2.18579235	1.09E-10	526	56	16792	9.12112982	3.20E-07	1.53E-08	1.96E-07
GO:0045666~positive regulation of neuron differentiation	18	2.45901639	2.16E-10	526	78	16792	7.36706639	6.36E-07	2.89E-08	3.89E-07
GO:0008284~positive regulation of cell proliferation	44	6.01092896	2.22E-10	526	466	16792	3.01427895	6.53E-07	2.84E-08	3.99E-07
GO:0021983~pituitary gland development	12	1.63934426	3.34E-10	526	28	16792	13.6816947	9.80E-07	4.08E-08	5.99E-07
GO:0045165~cell fate commitment	14	1.91256831	9.31E-10	526	46	16792	9.71598611	2.74E-06	1.09E-07	1.67E-06
GO:0030901~midbrain development	12	1.63934426	1.21E-09	526	31	16792	12.3576598	3.55E-06	1.36E-07	2.17E-06
GO:0048704~embryonic skeletal system morphogenesis	13	1.77595628	1.39E-09	526	39	16792	10.6413181	4.10E-06	1.52E-07	2.51E-06
GO:0008045~motor neuron axon guidance	10	1.36612022	1.86E-09	526	19	16792	16.8020812	5.48E-06	1.96E-07	3.35E-06
GO:0030326~embryonic limb morphogenesis	13	1.77595628	1.94E-09	526	40	16792	10.3752852	5.69E-06	1.96E-07	3.48E-06
GO:0031018~endocrine pancreas development	11	1.50273224	2.77E-09	526	26	16792	13.5062884	8.15E-06	2.72E-07	4.98E-06
GO:0021902~commitment of neuronal cell to specific neuron type in forebrain	7	0.95628415	6.19E-09	526	7	16792	31.9239544	1.82E-05	5.87E-07	1.11E-05

Term	Count	%	PValue	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0007275~multicellular organism development	44	6.01092896	6.80E-09	526	521	16792	2.69607292	2.00E-05	6.24E-07	1.22E-05
GO:0071542~dopaminergic neuron differentiation	10	1.36612022	9.23E-09	526	22	16792	14.5108884	2.71E-05	8.21E-07	1.66E-05
GO:0001501~skeletal system development	21	2.86885246	9.67E-09	526	137	16792	4.89345286	2.84E-05	8.35E-07	1.74E-05
GO:0042475~odontogenesis of dentin-containing tooth	14	1.91256831	1.03E-08	526	55	16792	8.12609748	3.02E-05	8.61E-07	1.84E-05
GO:0043066~negative regulation of apoptotic process	40	5.46448087	1.26E-08	526	455	16792	2.80650148	3.70E-05	1.03E-06	2.26E-05
GO:0048706~embryonic skeletal system development	11	1.50273224	1.40E-08	526	30	16792	11.7054499	4.12E-05	1.11E-06	2.52E-05
GO:0001657~ureteric bud development	12	1.63934426	1.41E-08	526	38	16792	10.0812487	4.13E-05	1.09E-06	2.53E-05
GO:0009953~dorsal/ventral pattern formation	11	1.50273224	2.85E-08	526	32	16792	10.9738593	8.36E-05	2.14E-06	5.11E-05
GO:0035115~embryonic forelimb morphogenesis	11	1.50273224	2.85E-08	526	32	16792	10.9738593	8.36E-05	2.14E-06	5.11E-05
GO:0043524~negative regulation of neuron apoptotic process	20	2.73224044	2.87E-08	526	132	16792	4.83696278	8.42E-05	2.11E-06	5.15E-05
GO:0001709~cell fate determination	9	1.2295082	2.87E-08	526	18	16792	15.9619772	8.44E-05	2.06E-06	5.16E-05
GO:0060070~canonical Wnt signaling pathway	16	2.18579235	3.75E-08	526	83	16792	6.1540153	1.10E-04	2.62E-06	6.74E-05
GO:0002053~positive regulation of mesenchymal cell proliferation	10	1.36612022	5.18E-08	526	26	16792	12.278444	1.52E-04	3.54E-06	9.31E-05
GO:0031016~pancreas development	9	1.2295082	1.23E-07	526	21	16792	13.6816947	3.61E-04	8.21E-06	2.21E-04
GO:0010628~positive regulation of gene expression	27	3.68852459	2.17E-07	526	262	16792	3.28987316	6.37E-04	1.42E-05	3.89E-04
GO:0048701~embryonic cranial skeleton morphogenesis	10	1.36612022	2.91E-07	526	31	16792	10.2980498	8.54E-04	1.86E-05	5.22E-04
GO:0048646~anatomical structure formation involved in morphogenesis	7	0.95628415	3.67E-07	526	11	16792	20.3152437	0.00107764	2.29E-05	6.59E-04
GO:0048663~neuron fate commitment	8	1.09289617	4.16E-07	526	17	16792	15.0230374	0.00122036	2.54E-05	7.47E-04
GO:0022008~neurogenesis	11	1.50273224	4.90E-07	526	42	16792	8.36103567	0.00143798	2.94E-05	8.80E-04
GO:0048538~thymus development	11	1.50273224	6.21E-07	526	43	16792	8.16659298	0.00182144	3.65E-05	0.00111476
GO:0051091~positive regulation of sequence-specific DNA binding transcription factor activity	16	2.18579235	9.12E-07	526	105	16792	4.86460257	0.00267582	5.25E-05	0.00163835
GO:0021766~hippocampus development	12	1.63934426	1.05E-06	526	56	16792	6.84084737	0.00306857	5.91E-05	0.00187919
GO:0042733~embryonic digit morphogenesis	12	1.63934426	1.05E-06	526	56	16792	6.84084737	0.00306857	5.91E-05	0.00187919
GO:0030182~neuron differentiation	15	2.04918033	1.44E-06	526	95	16792	5.04062437	0.0042072	7.95E-05	0.00257795
GO:0021549~cerebellum development	10	1.36612022	1.52E-06	526	37	16792	8.62809578	0.00444974	8.26E-05	0.0027269
GO:0042474~middle ear morphogenesis	8	1.09289617	1.53E-06	526	20	16792	12.7695817	0.00447517	8.15E-05	0.00274252
GO:0071300~cellular response to retinoic acid	13	1.77595628	1.57E-06	526	70	16792	5.92873438	0.00460251	8.24E-05	0.00282073
GO:0050679~positive regulation of epithelial cell proliferation	12	1.63934426	2.15E-06	526	60	16792	6.38479087	0.00629741	1.11E-04	0.00386275
GO:0007420~brain development	21	2.86885246	2.20E-06	526	190	16792	3.52843706	0.00643429	1.11E-04	0.00394698
GO:0072001~renal system development	7	0.95628415	2.20E-06	526	14	16792	15.9619772	0.00644667	1.10E-04	0.0039546
GO:0048485~sympathetic nervous system development	7	0.95628415	3.57E-06	526	15	16792	14.8978454	0.0104399	1.75E-04	0.00641699
GO:0030513~positive regulation of BMP signaling pathway	9	1.2295082	3.62E-06	526	31	16792	9.26824482	0.01057563	1.74E-04	0.00650086
GO:0001649~osteoblast differentiation	15	2.04918033	4.32E-06	526	104	16792	4.6044165	0.01259945	2.04E-04	0.00775277
GO:0072178~nephric duct morphogenesis	5	0.68306011	4.61E-06	526	5	16792	31.9239544	0.01344549	2.15E-04	0.00827688
GO:0045599~negative regulation of fat cell differentiation	10	1.36612022	4.73E-06	526	42	16792	7.60094152	0.01381082	2.17E-04	0.00850334
GO:0045880~positive regulation of smoothened signaling pathway	8	1.09289617	6.11E-06	526	24	16792	10.6413181	0.01779961	2.76E-04	0.01098127
GO:0009954~proximal/distal pattern formation	8	1.09289617	6.11E-06	526	24	16792	10.6413181	0.01779961	2.76E-04	0.01098127
GO:0048935~peripheral nervous system neuron development	6	0.81967213	6.49E-06	526	10	16792	19.1543726	0.01887391	2.89E-04	0.01165036

Term	Count	%	PValue	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0001701~in utero embryonic development	20	2.73224044	6.61E-06	526	187	16792	3.41432667	0.01923634	2.90E-04	0.01187626
GO:0008584~male gonad development	14	1.91256831	7.00E-06	526	94	16792	4.7546315	0.02034916	3.02E-04	0.01257037
GO:0021542~dentate gyrus development	7	0.95628415	8.37E-06	526	17	16792	13.1451577	0.02430012	3.56E-04	0.01504107
GO:0048565~digestive tract development	9	1.2295082	9.67E-06	526	35	16792	8.20901684	0.02801146	4.06E-04	0.017371
GO:0030900~forebrain development	10	1.36612022	1.05E-05	526	46	16792	6.93999008	0.03027671	4.33E-04	0.01879744
GO:0048666~neuron development	10	1.36612022	1.05E-05	526	46	16792	6.93999008	0.03027671	4.33E-04	0.01879744
GO:0042493~response to drug	26	3.55191257	1.11E-05	526	304	16792	2.7303382	0.03215222	4.54E-04	0.01998097
GO:0021904~dorsal/ventral neural tube patterning	6	0.81967213	1.16E-05	526	11	16792	17.413066	0.03346178	4.66E-04	0.02080871
GO:0003323~type B pancreatic cell development	6	0.81967213	1.16E-05	526	11	16792	17.413066	0.03346178	4.66E-04	0.02080871
GO:0048536~spleen development	9	1.2295082	1.21E-05	526	36	16792	7.98098859	0.03491327	4.80E-04	0.02172748
GO:0048557~embryonic digestive tract morphogenesis	7	0.95628415	1.22E-05	526	18	16792	12.4148711	0.03529046	4.79E-04	0.02196645
GO:0021978~telencephalon regionalization	5	0.68306011	1.35E-05	526	6	16792	26.6032953	0.03882909	5.21E-04	0.02421296
GO:0007219~Notch signaling pathway	15	2.04918033	1.41E-05	526	115	16792	4.16399405	0.04061073	5.38E-04	0.02534715
GO:0001822~kidney development	13	1.77595628	1.43E-05	526	86	16792	4.82571403	0.04124346	5.40E-04	0.02575045
GO:0033077~T cell differentiation in thymus	8	1.09289617	1.45E-05	526	27	16792	9.45894944	0.04158493	5.38E-04	0.02596821
GO:0048863~stem cell differentiation	8	1.09289617	1.45E-05	526	27	16792	9.45894944	0.04158493	5.38E-04	0.02596821
GO:0045597~positive regulation of cell differentiation	9	1.2295082	1.50E-05	526	37	16792	7.7652862	0.04315179	5.51E-04	0.02696841
GO:0045444~fat cell differentiation	12	1.63934426	1.55E-05	526	73	16792	5.24777332	0.04444285	5.61E-04	0.02779379
GO:0007267~cell-cell signaling	23	3.1420765	1.67E-05	526	254	16792	2.89075177	0.04797481	5.99E-04	0.03005744
GO:0090190~positive regulation of branching involved in ureteric bud morphogenesis	7	0.95628415	1.74E-05	526	19	16792	11.7614569	0.04984284	6.16E-04	0.03125804
GO:0021772~olfactory bulb development	7	0.95628415	1.74E-05	526	19	16792	11.7614569	0.04984284	6.16E-04	0.03125804
GO:0035116~embryonic hindlimb morphogenesis	8	1.09289617	1.88E-05	526	28	16792	9.12112982	0.05362446	6.56E-04	0.03369574
GO:0007389~pattern specification process	8	1.09289617	1.88E-05	526	28	16792	9.12112982	0.05362446	6.56E-04	0.03369574
GO:0045668~negative regulation of osteoblast differentiation	9	1.2295082	2.26E-05	526	39	16792	7.36706639	0.06436667	7.82E-04	0.04067351
GO:0002062~chondrocyte differentiation	9	1.2295082	2.26E-05	526	39	16792	7.36706639	0.06436667	7.82E-04	0.04067351
GO:0030324~lung development	12	1.63934426	2.29E-05	526	76	16792	5.04062437	0.06502817	7.82E-04	0.0411058
GO:0009611~response to wounding	11	1.50273224	2.36E-05	526	63	16792	5.57402378	0.06700872	7.97E-04	0.0424019
GO:0060412~ventricular septum morphogenesis	8	1.09289617	2.41E-05	526	29	16792	8.8066081	0.06826599	8.03E-04	0.04322609
GO:0030512~negative regulation of transforming growth factor beta receptor signaling pathway	11	1.50273224	2.72E-05	526	64	16792	5.48692966	0.07682507	8.98E-04	0.04886649
GO:0071773~cellular response to BMP stimulus	8	1.09289617	3.06E-05	526	30	16792	8.5130545	0.08584954	9.97E-04	0.05487014
GO:0021520~spinal cord motor neuron cell fate specification	5	0.68306011	3.07E-05	526	7	16792	22.8028246	0.08617694	9.90E-04	0.05508905
GO:0021796~cerebral cortex regionalization	5	0.68306011	3.07E-05	526	7	16792	22.8028246	0.08617694	9.90E-04	0.05508905
GO:0048146~positive regulation of fibroblast proliferation	10	1.36612022	4.05E-05	526	54	16792	5.9118434	0.11205873	0.00129101	0.07264607
GO:0045595~regulation of cell differentiation	8	1.09289617	4.78E-05	526	32	16792	7.98098859	0.1311405	0.0015104	0.08591912
GO:0007224~smoothened signaling pathway	11	1.50273224	5.31E-05	526	69	16792	5.08932606	0.14447817	0.00165866	0.0953698
GO:2000543~positive regulation of gastrulation	5	0.68306011	5.98E-05	526	8	16792	19.9524715	0.16119588	0.00184858	0.10742447
GO:0035137~hindlimb morphogenesis	5	0.68306011	5.98E-05	526	8	16792	19.9524715	0.16119588	0.00184858	0.10742447
GO:0043401~steroid hormone mediated signaling pathway	10	1.36612022	6.30E-05	526	57	16792	5.60069375	0.16891383	0.00192545	0.11307046

Term	Count	%	PValue	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0007507~heart development	18	2.45901639	6.32E-05	526	183	16792	3.14006109	0.16946742	0.00191248	0.11347743
GO:0010468~regulation of gene expression	13	1.77595628	6.63E-05	526	100	16792	4.15011407	0.17695866	0.00198526	0.11901131
GO:2000679~positive regulation of transcription regulatory region DNA binding	6	0.81967213	6.79E-05	526	15	16792	12.7695817	0.18075371	0.00201182	0.1218339
GO:0060441~epithelial tube branching involved in lung morphogenesis	6	0.81967213	6.79E-05	526	15	16792	12.7695817	0.18075371	0.00201182	0.1218339
GO:0045787~positive regulation of cell cycle	8	1.09289617	7.24E-05	526	34	16792	7.51151868	0.19171089	0.00212609	0.13005689
GO:0003151~outflow tract morphogenesis	9	1.2295082	7.93E-05	526	46	16792	6.24599107	0.20772927	0.00230281	0.14227971
GO:0051216~cartilage development	10	1.36612022	8.32E-05	526	59	16792	5.41083972	0.21690778	0.00239424	0.14939453
GO:0001889~liver development	11	1.50273224	9.76E-05	526	74	16792	4.74545268	0.24929661	0.00278006	0.17518082
GO:0003309~type B pancreatic cell differentiation	5	0.68306011	1.05E-04	526	9	16792	17.7355302	0.26551389	0.00296276	0.18851066
GO:0021913~regulation of transcription from RNA polymerase II promoter involved in ventral spinal cord interneuron specification	4	0.54644809	1.19E-04	526	4	16792	31.9239544	0.29450913	0.00331698	0.21308935
GO:0021521~ventral spinal cord interneuron specification	4	0.54644809	1.19E-04	526	4	16792	31.9239544	0.29450913	0.00331698	0.21308935
GO:0021522~spinal cord motor neuron differentiation	6	0.81967213	1.33E-04	526	17	16792	11.267278	0.32298268	0.00367303	0.23822303
GO:0060349~bone morphogenesis	7	0.95628415	1.53E-04	526	27	16792	8.27658076	0.36291586	0.00420472	0.2753017
GO:0001756~somitogenesis	8	1.09289617	1.81E-04	526	39	16792	6.54850346	0.41233988	0.00491019	0.32453127
GO:0030178~negative regulation of Wnt signaling pathway	9	1.2295082	1.94E-04	526	52	16792	5.5252998	0.43438482	0.00521426	0.34783192
GO:0043010~camera-type eye development	8	1.09289617	2.13E-04	526	40	16792	6.38479087	0.46592425	0.00568575	0.38278736
GO:0031290~retinal ganglion cell axon guidance	6	0.81967213	2.37E-04	526	19	16792	10.0812487	0.50141803	0.00625054	0.42466788
GO:0023019~signal transduction involved in regulation of gene expression	6	0.81967213	2.37E-04	526	19	16792	10.0812487	0.50141803	0.00625054	0.42466788
GO:0048839~inner ear development	8	1.09289617	2.51E-04	526	41	16792	6.22906427	0.52108244	0.0065519	0.44916546
GO:0060272~embryonic skeletal joint morphogenesis	5	0.68306011	2.62E-04	526	11	16792	14.5108884	0.53645191	0.00678084	0.46901888
GO:0001654~eye development	7	0.95628415	2.84E-04	526	30	16792	7.44892269	0.56623062	0.00729991	0.5094198
GO:0021510~spinal cord development	7	0.95628415	2.84E-04	526	30	16792	7.44892269	0.56623062	0.00729991	0.5094198
GO:0021797~forebrain anterior/posterior pattern specification	4	0.54644809	2.90E-04	526	5	16792	25.5391635	0.57340506	0.00738063	0.51956539
GO:0051891~positive regulation of cardioblast differentiation	4	0.54644809	2.90E-04	526	5	16792	25.5391635	0.57340506	0.00738063	0.51956539
GO:0001541~ovarian follicle development	8	1.09289617	2.93E-04	526	42	16792	6.08075321	0.5768292	0.0073862	0.52446751
GO:0001570~vasculogenesis	9	1.2295082	3.28E-04	526	56	16792	5.13063552	0.61875567	0.00820813	0.58791017
GO:0050680~negative regulation of epithelial cell proliferation	9	1.2295082	3.28E-04	526	56	16792	5.13063552	0.61875567	0.00820813	0.58791017
GO:0043525~positive regulation of neuron apoptotic process	8	1.09289617	3.40E-04	526	43	16792	5.93934035	0.63208331	0.00843792	0.60953825
GO:0007417~central nervous system development	13	1.77595628	3.79E-04	526	120	16792	3.45842839	0.6712676	0.00930527	0.6779536
GO:0021846~cell proliferation in forebrain	5	0.68306011	3.83E-04	526	12	16792	13.3016477	0.67531823	0.00933045	0.68548319
GO:0048844~artery morphogenesis	6	0.81967213	3.94E-04	526	21	16792	9.12112982	0.68546257	0.00951357	0.70475767
GO:0030514~negative regulation of BMP signaling pathway	8	1.09289617	4.54E-04	526	45	16792	5.67536967	0.73666369	0.01087748	0.8125732
GO:0071364~cellular response to epidermal growth factor stimulus	7	0.95628415	4.90E-04	526	33	16792	6.7717479	0.76286693	0.01163209	0.87612007
GO:0010719~negative regulation of epithelial to mesenchymal transition	6	0.81967213	4.96E-04	526	22	16792	8.70653301	0.7674701	0.01169508	0.88800068

Term	Count	%	PValue	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0043065~positive regulation of apoptotic process	22	3.00546448	5.13E-04	526	300	16792	2.34108999	0.77884278	0.01199849	0.91838566
GO:0043433~negative regulation of sequence-specific DNA binding transcription factor activity	9	1.2295082	5.30E-04	526	60	16792	4.78859316	0.78933915	0.01228507	0.94784032
GO:0048712~negative regulation of astrocyte differentiation	5	0.68306011	5.39E-04	526	13	16792	12.278444	0.79504932	0.01240265	0.96448283
GO:0061312~BMP signaling pathway involved in heart development	4	0.54644809	5.66E-04	526	6	16792	21.2826362	0.81070759	0.01291943	1.01259947
GO:0048286~lung alveolus development	7	0.95628415	5.79E-04	526	34	16792	6.57257884	0.81767758	0.01310698	1.03530453
GO:0001947~heart looping	9	1.2295082	5.94E-04	526	61	16792	4.71009163	0.82525967	0.01332925	1.06100464
GO:0009653~anatomical structure morphogenesis	11	1.50273224	5.94E-04	526	92	16792	3.81699454	0.82528419	0.01322923	1.06108957
GO:0090090~negative regulation of canonical Wnt signaling pathway	15	2.04918033	6.11E-04	526	163	16792	2.93778721	0.83399073	0.01351176	1.0920094
GO:0014912~negative regulation of smooth muscle cell migration	5	0.68306011	7.36E-04	526	14	16792	11.4014123	0.88520601	0.01614358	1.314867
GO:0061053~somite development	5	0.68306011	7.36E-04	526	14	16792	11.4014123	0.88520601	0.01614358	1.314867
GO:0030336~negative regulation of cell migration	11	1.50273224	7.65E-04	526	95	16792	3.69645787	0.89458128	0.0166495	1.36626484
GO:0048469~cell maturation	7	0.95628415	7.96E-04	526	36	16792	6.20743557	0.90350784	0.01717155	1.41961272
GO:0007605~sensory perception of sound	13	1.77595628	9.56E-04	526	133	16792	3.12038652	0.93981685	0.02045239	1.70375874
GO:1904628~cellular response to phorbol 13-acetate 12-myristate	4	0.54644809	9.68E-04	526	7	16792	18.2422596	0.94191072	0.02055782	1.72504031
GO:0014807~regulation of somitogenesis	4	0.54644809	9.68E-04	526	7	16792	18.2422596	0.94191072	0.02055782	1.72504031
GO:0060290~transdifferentiation	4	0.54644809	9.68E-04	526	7	16792	18.2422596	0.94191072	0.02055782	1.72504031
GO:0003337~mesenchymal to epithelial transition involved in metanephros morphogenesis	4	0.54644809	9.68E-04	526	7	16792	18.2422596	0.94191072	0.02055782	1.72504031
GO:0021516~dorsal spinal cord development	4	0.54644809	9.68E-04	526	7	16792	18.2422596	0.94191072	0.02055782	1.72504031
GO:2000049~positive regulation of cell-cell adhesion mediated by cadherin	4	0.54644809	9.68E-04	526	7	16792	18.2422596	0.94191072	0.02055782	1.72504031
GO:0021798~forebrain dorsal/ventral pattern formation	4	0.54644809	9.68E-04	526	7	16792	18.2422596	0.94191072	0.02055782	1.72504031
GO:0035050~embryonic heart tube development	5	0.68306011	9.80E-04	526	15	16792	10.6413181	0.94384148	0.0206503	1.74535089
GO:0009888~tissue development	5	0.68306011	9.80E-04	526	15	16792	10.6413181	0.94384148	0.0206503	1.74535089
GO:0008585~female gonad development	5	0.68306011	9.80E-04	526	15	16792	10.6413181	0.94384148	0.0206503	1.74535089
GO:0048714~positive regulation of oligodendrocyte differentiation	5	0.68306011	9.80E-04	526	15	16792	10.6413181	0.94384148	0.0206503	1.74535089
GO:0006367~transcription initiation from RNA polymerase II promoter	14	1.91256831	9.84E-04	526	152	16792	2.94036422	0.9445403	0.0205915	1.75287361
GO:0071333~cellular response to glucose stimulus	8	1.09289617	0.00110902	526	52	16792	4.9113776	0.96161604	0.0230175	1.97371018
GO:0042593~glucose homeostasis	11	1.50273224	0.00123251	526	101	16792	3.47686632	0.97330714	0.02537021	2.19119569
GO:0070848~response to growth factor	5	0.68306011	0.00127412	526	16	16792	9.97623574	0.97638227	0.0260337	2.26437068
GO:2000179~positive regulation of neural precursor cell proliferation	5	0.68306011	0.00127412	526	16	16792	9.97623574	0.97638227	0.0260337	2.26437068
GO:0055007~cardiac muscle cell differentiation	6	0.81967213	0.0013377	526	27	16792	7.09421208	0.98041137	0.0271274	2.37609048
GO:0048468~cell development	7	0.95628415	0.00141078	526	40	16792	5.58669202	0.98420088	0.02839329	2.50434468
GO:0030855~epithelial cell differentiation	9	1.2295082	0.00148966	526	70	16792	4.10450842	0.9874733	0.02975451	2.64260857
GO:0043392~negative regulation of DNA binding	6	0.81967213	0.00158717	526	28	16792	6.84084737	0.99059769	0.03145892	2.81325675
GO:0003281~ventricular septum development	6	0.81967213	0.00158717	526	28	16792	6.84084737	0.99059769	0.03145892	2.81325675
GO:0021537~telencephalon development	5	0.68306011	0.00162531	526	17	16792	9.38939834	0.99159604	0.03198777	2.87994093

Term	Count	%	PValue	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0030902~hindbrain development	5	0.68306011	0.00162531	526	17	16792	9.38939834	0.99159604	0.03198777	2.87994093
GO:0007601~visual perception	16	2.18579235	0.00167234	526	201	16792	2.5412103	0.99268214	0.03268001	2.96208728
GO:0021987~cerebral cortex development	8	1.09289617	0.00172705	526	56	16792	4.56056491	0.99377042	0.03350924	3.05757564
GO:0009791~post-embryonic development	9	1.2295082	0.00195533	526	73	16792	3.93582999	0.99681839	0.03761028	3.45505323
GO:0060037~pharyngeal system development	5	0.68306011	0.00203851	526	18	16792	8.8677651	0.99750948	0.03892586	3.5995157
GO:0060041~retina development in camera-type eye	8	1.09289617	0.00212237	526	58	16792	4.40330405	0.99805433	0.04023496	3.74493488
GO:0060325~face morphogenesis	6	0.81967213	0.00218616	526	30	16792	6.38479087	0.99838752	0.04115518	3.85542335
GO:0044344~cellular response to fibroblast growth factor stimulus	6	0.81967213	0.00218616	526	30	16792	6.38479087	0.99838752	0.04115518	3.85542335
GO:0007623~circadian rhythm	9	1.2295082	0.00232463	526	75	16792	3.83087452	0.99892744	0.04342939	4.09481967
GO:0010977~negative regulation of neuron projection development	7	0.95628415	0.00233537	526	44	16792	5.07881092	0.99896084	0.0433506	4.11336966
GO:0030325~adrenal gland development	5	0.68306011	0.00251895	526	19	16792	8.40104062	0.99939484	0.04638964	4.42985488
GO:0030509~BMP signaling pathway	9	1.2295082	0.00252869	526	76	16792	3.78046828	0.99941195	0.04627534	4.44661455
GO:0006342~chromatin silencing	7	0.95628415	0.0026248	526	45	16792	4.96594846	0.99955694	0.04769716	4.61188226
GO:0001843~neural tube closure	9	1.2295082	0.0027465	526	77	16792	3.73137129	0.99969043	0.04954998	4.82076977

C.4 H1299 EZH2 ChIP-seq Quality Control

Both replicates of the H1299 EZH2 ChIP-seq experiment were processed through ENCODE's Transcription Factor ChIP-seq Processing Pipeline v2 (https://www.encodeproject.org/chip-seq/transcription_factor/#tfbs). Quality control metrics collected as part of the pipeline are displayed below.

C.2.1 QC report

	general
Report generated at	2021-08-29 17:55:39
Title	H1299 EZH2 replicates 1 and 2 (paired end)
Description	This is the input JSON file for running the chip seq 2 pipeline for H1299 EZH2 ChIP-seq replicates 1 and 2 processing (paired-end).
Pipeline version	v1.9.0
Pipeline type	tf
Genome	hg19
Aligner	bowtie2
Sequencing endedness	{'rep1': {'paired_end': True}, 'rep2': {'paired_end': True}, 'ctl1': {'paired_end': True}, 'ctl2': {'paired_end': True}}
Peak caller	spp

C.2.2 Alignment quality metrics

C.2.2.1 SAMstat (raw unfiltered BAM)

	rep1	rep2	ctl1	ctl2
Total Reads	114328506	60673884	95067898	57144396
Total Reads (QC-failed)	0	0	0	0
Duplicate Reads	0	0	0	0
Duplicate Reads (QC-failed)	0	0	0	0
Mapped Reads	110608685	58062631	92687910	55878027
Mapped Reads (QC-failed)	0	0	0	0
% Mapped Reads	96.7	95.7	97.5	97.8
Paired Reads	114328506	60673884	95067898	57144396
Paired Reads (QC-failed)	0	0	0	0
Read1	57164253	30336942	47533949	28572198
Read1 (QC-failed)	0	0	0	0
Read2	57164253	30336942	47533949	28572198
Read2 (QC-failed)	0	0	0	0
Properly Paired Reads	107944960	57257966	90698732	55136034
Properly Paired Reads (QC-failed)	0	0	0	0
% Properly Paired Reads	94.39999999999999	94.39999999999999	95.39999999999999	96.5
With itself	109693058	57612054	91904426	55491670
With itself (QC-failed)	0	0	0	0
Singletons	915627	450577	783484	386357
Singletons (QC-failed)	0	0	0	0
% Singleton	0.8	0.7000000000000001	0.8	0.7000000000000001
Diff. Chroms	1280144	140957	846617	158041
Diff. Chroms (QC-failed)	0	0	0	0

C.2.2.2 Marking duplicates (filtered BAM)

	rep1	rep2	ctl1	ctl2
Unpaired Reads	0	0	0	0
Paired Reads	48220100	25639554	40401957	24782003
Unmapped Reads	0	0	0	0
Unpaired Duplicate Reads	0	0	0	0
Paired Duplicate Reads	4435655	549478	646652	496694
Paired Optical Duplicate Reads	11829	5112	8565	4479
% Duplicate Reads	9.1988	2.1431	1.6004999999999998	2.0042999999999997

Filtered out (samtools view -F 1804):

- read unmapped (0x4)
- mate unmapped (0x8, for paired-end)
- not primary alignment (0x100)
- read fails platform/vendor quality checks (0x200)
- read is PCR or optical duplicate (0x400)

C.2.2.3 SAMstat (filtered/deduped BAM) – filtered and duplicates removed.

	rep1	rep2	ctl1	ctl2
Total Reads	87568890	50180152	79510610	48570618
Total Reads (QC-failed)	0	0	0	0
Duplicate Reads	0	0	0	0
Duplicate Reads (QC-failed)	0	0	0	0
Mapped Reads	87568890	50180152	79510610	48570618
Mapped Reads (QC-failed)	0	0	0	0
% Mapped Reads	100.0	100.0	100.0	100.0
Paired Reads	87568890	50180152	79510610	48570618
Paired Reads (QC-failed)	0	0	0	0
Read1	43784445	25090076	39755305	24285309
Read1 (QC-failed)	0	0	0	0
Read2	43784445	25090076	39755305	24285309
Read2 (QC-failed)	0	0	0	0
Properly Paired Reads	87568890	50180152	79510610	48570618
Properly Paired Reads (QC-failed)	0	0	0	0
% Properly Paired Reads	100.0	100.0	100.0	100.0
With itself	87568890	50180152	79510610	48570618
With itself (QC-failed)	0	0	0	0
Singletons	0	0	0	0
Singletons (QC-failed)	0	0	0	0
% Singleton	0.0	0.0	0.0	0.0
Diff. Chroms	0	0	0	0
Diff. Chroms (QC-failed)	0	0	0	0

C.2.3 Library complexity quality metrics

C.2.3.1 Library complexity (filtered non-mito BAM)

	rep1	rep2	ctl1	ctl2
Total Fragments	48219578	25638794	40400851	24780385
Distinct Fragments	43785558	25089677	39754420	24284120
Positions with Two Read	3692428	525412	617520	477052
NRF = Distinct/Total	0.908045	0.978583	0.984	0.979973
PBC1 = OneRead/Distinct	0.907635	0.978627	0.984175	0.979992
PBC2 = OneRead/TwoRead	10.762919	46.731753	63.358795	49.886048

Mitochondrial reads are filtered out by default. The non-redundant fraction (NRF) is the fraction of non-redundant mapped reads in a dataset; it is the ratio between the number of positions in the genome that uniquely mapped reads map to and the total number of uniquely mappable reads. The NRF should be > 0.8. The PBC1 is the ratio of genomic locations with EXACTLY one read pair over the genomic locations with AT LEAST one read pair. PBC1 is the primary measure, and the PBC1 should be close to 1. Provisionally 0-0.5 is severe bottlenecking, 0.5-0.8 is moderate bottlenecking, 0.8-0.9 is mild bottlenecking, and 0.9-1.0 is no bottlenecking. The PBC2 is the ratio of genomic locations with EXACTLY one read pair over the genomic locations with EXACTLY two read pairs. The PBC2 should be significantly greater than 1. See more details at the ENCODE portal standard for ChIP-Seq pipeline

NRF (non redundant fraction)
PBC1 (PCR Bottleneck coefficient 1)
PBC2 (PCR Bottleneck coefficient 2)
PBC1 is the primary measure. Provisionally

- 0-0.5 is severe bottlenecking
- 0.5-0.8 is moderate bottlenecking
- 0.8-0.9 is mild bottlenecking
- 0.9-1.0 is no bottlenecking

C.2.4 Replication quality metrics

C.2.4.1 Reproducibility QC and peak detection statistics

	overlap	idr
Nt	39052	2182
N1	36057	2068
N2	27091	1557
Np	45791	2793
N optimal	45791	2793
N conservative	39052	2182
Optimal Set	pooled-pr1_vs_pooled-pr2	pooled-pr1_vs_pooled-pr2
Conservative Set	rep1_vs_rep2	rep1_vs_rep2
Rescue Ratio	1.1725647854143193	1.2800183318056828
Self Consistency Ratio	1.330958620944225	1.3281952472703917
Reproducibility Test	pass	pass

Reproducibility QC

- N1: Replicate 1 self-consistent peaks (comparing two pseudoreplicates generated by subsampling Rep1 reads)
- N2: Replicate 2 self-consistent peaks (comparing two pseudoreplicates generated by subsampling Rep2 reads)
- Ni: Replicate i self-consistent peaks (comparing two pseudoreplicates generated by subsampling RepX reads)
- Nt: True Replicate consistent peaks (comparing true replicates Rep1 vs Rep2)
- Np: Pooled-pseudoreplicate consistent peaks (comparing two pseudoreplicates generated by subsampling pooled reads from Rep1 and Rep2)
- Self-consistency Ratio: max(N1,N2) / min (N1,N2)
- Rescue Ratio: max(Np,Nt) / min (Np,Nt)
- Reproducibility Test: If Self-consistency Ratio >2 AND Rescue Ratio > 2, then 'Fail' else 'Pass'

C.2.4.2 Number of raw peaks

	rep1	rep2
Number of peaks	298549	298882

Top 300000 raw peaks from spp with FDR 0.01

C.2.5 Peak calling statistics

C.2.5.1 Peak region size

	rep1	rep2	idr_opt	overlap_opt
Min size	50.0	50.0	50.0	50.0
25 percentile	196.0	184.0	196.0	196.0
50 percentile (median)	196.0	184.0	196.0	196.0
75 percentile	196.0	184.0	196.0	196.0
Max size	196.0	297.0	263.0	263.0
Mean	195.98046886775705	183.9851881344477	192.80379520229144	195.80129283046887

C.2.6 Enrichment / signal-to-noise ratio

C.2.6.1 Strand cross-correlation measures (trimmed/filtered SE BAM)

	rep1	rep2
Number of Subsampled Reads	15000000	15000000
Estimated Fragment Length	150	145
Cross-correlation at Estimated Fragment Length	0.172155003226747	0.175733485251598
Phantom Peak	75	80
Cross-correlation at Phantom Peak	0.169514	0.1729907
Argmin of Cross-correlation	1500	1500
Minimum of Cross-correlation	0.1675704	0.1702225
NSC (Normalized Strand Cross-correlation coeff.)	1.02736	1.032375
RSC (Relative Strand Cross-correlation coeff.)	2.358821	1.990836

Performed on subsampled (15000000) reads mapped from FASTQs that are trimmed to 50. Such FASTQ trimming and subsampling reads are for cross-corrleation analysis only. Untrimmed FASTQs are used for all the other analyses.

NOTE1: For SE datasets, reads from replicates are randomly subsampled to 15000000.
NOTE2: For PE datasets, the first end (R1) of each read-pair is selected and trimmed to 50 the reads are then randomly subsampled to 15000000.

- Normalized strand cross-correlation coefficient (NSC) = col9 in outFile
- Relative strand cross-correlation coefficient (RSC) = col10 in outFile
- Estimated fragment length = col3 in outFile, take the top value

C.2.6.2 Jensen-Shannon distance (filtered/deduped BAM)

	rep1	rep2
AUC	0.2896385953244884	0.272085555718559
Synthetic AUC	0.49563515119393253	0.49424194427225926
X-intercept	0.10657126115658516	0.12841147025943872
Synthetic X-intercept	0.0	9.286791911007705e-259
Elbow Point	0.5766887115256549	0.5881057753264233
Synthetic Elbow Point	0.5040827076298817	0.5020554514560518
JS Distance	0.02561459628141144	0.041522170701275694
Synthetic JS Distance	0.2600826273471597	0.2753929383199148
% Genome Enriched	33.40886047344118	34.21939400969121
Diff. Enrichment	9.303494300486815	13.228667209598072
CHANCE Divergence	0.08003216706275648	0.11463951459578296

C.2.7 Peak enrichment

C.2.7.1 Fraction of reads in peaks (FRIP)

	rep1	rep2	rep1-pr1	rep2-pr1	rep1-pr2	rep2-pr2	pooled	pooled-pr1	pooled-pr2
Fraction of Reads in Peaks	0.08974746625199885	0.10171908207850786	0.10307208180731578	0.1229391254135699	0.10309453741150625	0.12280201941197787	0.0832439546113141	0.09259601104745235	0.09264238792517175

FRiP for overlap peaks

	rep1_vs_rep2	rep1-pr1_vs_rep1-pr2	rep2-pr1_vs_rep2-pr2	pooled-pr1_vs_pooled-pr2
Fraction of Reads in Peaks	0.015929816775059676	0.01613061442254207	0.014538537069397478	0.01846240063143234

FRiP for IDR peaks

	rep1_vs_rep2	rep1-pr1_vs_rep1-pr2	rep2-pr1_vs_rep2-pr2	pooled-pr1_vs_pooled-pr2
Fraction of Reads in Peaks	0.0023474791207622335	0.0023627112322652487	0.00221695223428471	0.002865958225684067

For spp raw peaks:

- repX: Peak from true replicate X
- repX-prY: Peak from Yth pseudoreplicates from replicate X
- pooled: Peak from pooled true replicates (pool of rep1, rep2, ...)
- pooled-pr1: Peak from 1st pooled pseudo replicate (pool of rep1-pr1, rep2-pr1, ...)
- pooled-pr2: Peak from 2nd pooled pseudo replicate (pool of rep1-pr2, rep2-pr2, ...)

For overlap/IDR peaks:

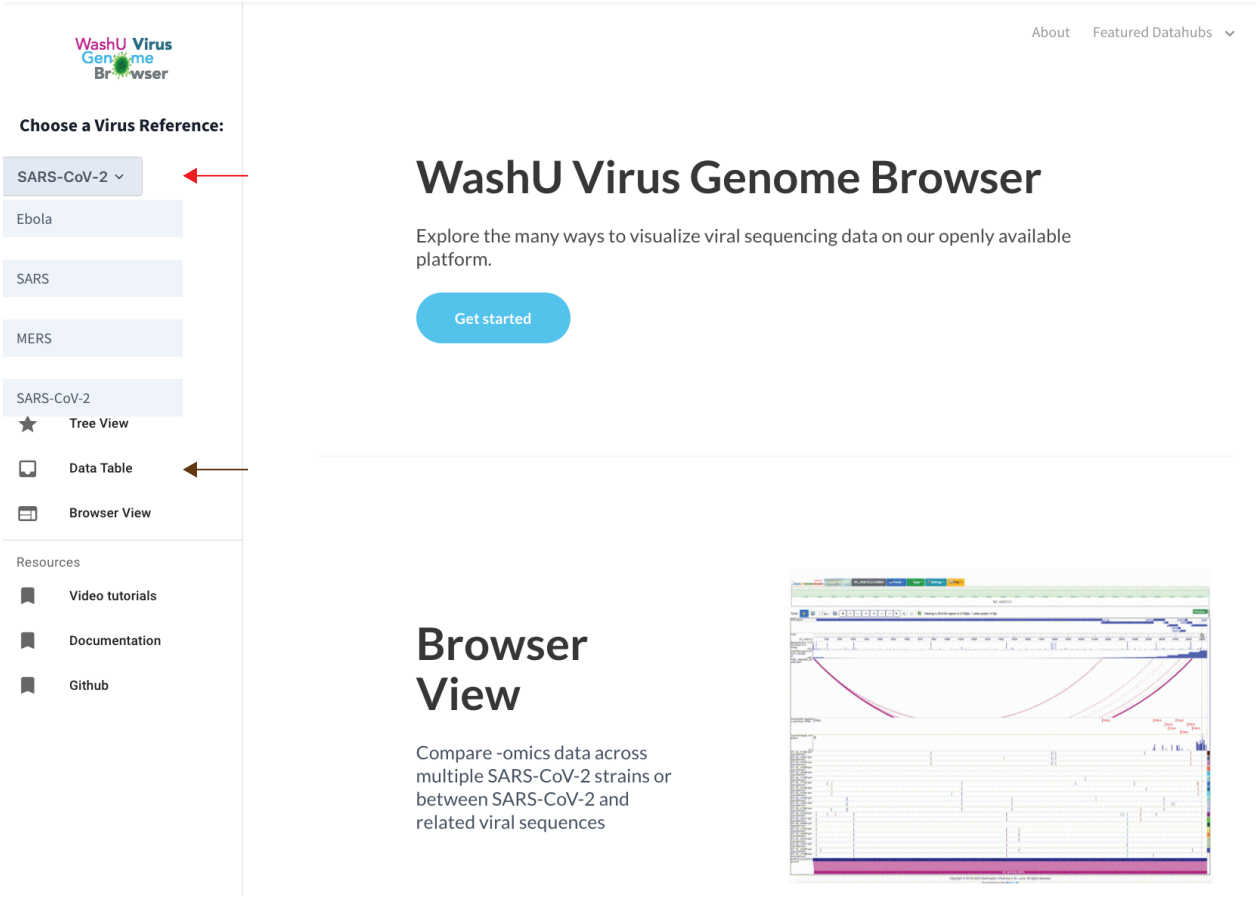
- repX_vs_repY: Comparing two peaks from true replicates X and Y
- repX-pr1_vs_repX-pr2: Comparing two peaks from both pseudoreplicates from replicate X
- pooled-pr1_vs_pooled-pr2: Comparing two peaks from 1st and 2nd pooled pseudo replicates

Appendix D: Chapter 7 extended data

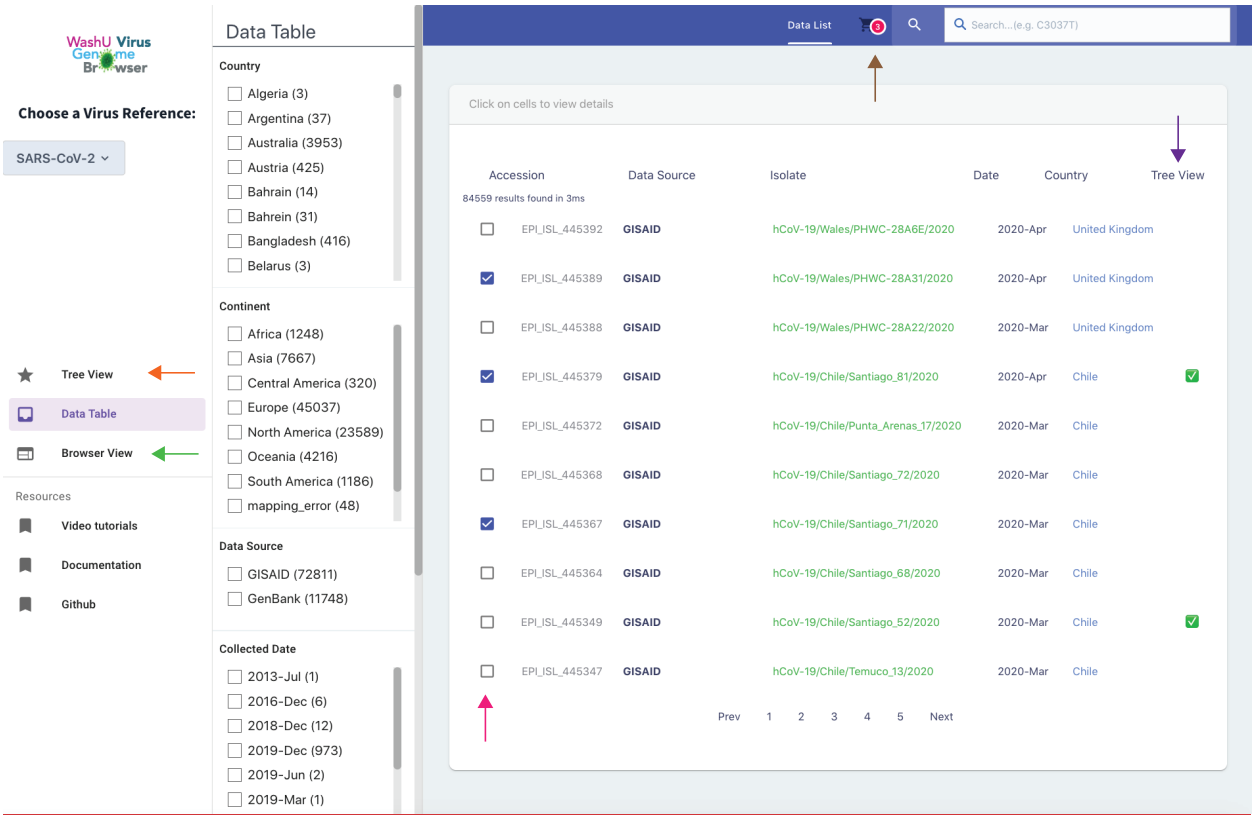
D.1 Navigating the WashU Virus Genome Browser

D.1.1 Virus Browser Home Page and Data Table

In addition to SARS-CoV-2, the WashU Virus Genome Browser hosts hundreds to thousands of genomic sequence pairwise alignments to the reference for 3 additional virus species: severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and Ebola. When users first navigate to the browser landing page (http://virusgateway.wustl.edu/), they are provided with a drop-down menu of the 4 hosted viruses (red arrow below), to select which virus they would like to view as their reference.



Upon selecting a reference species and clicking “Data Table” (brown arrow, above), the user is then taken to a table populated with searchable and filterable metadata for all available strains of the species, to easily identify strains they would like to view.



In the data table, the user can apply filters (which include “Country”, “Continent”, “Data Source”, “Collected Date”, “Filter on Treeview availability?”, and “Clade”), to hone in on strains of interest. Upon identifying strains the user would like to examine further, he or she can select the strains of interest, as demonstrated by the checked boxes (pink arrow). Selected strains are then added to the user’s cart (brown arrow). If the selected strains are available for viewing in the phlogenetic tree view (for SARS-CoV-2, this includes all Nextstrain strains [1], and is one of the filtering options and metadata columns (purple arrow)), the user can select “Tree View” (orange arrow) to see how their selected strains relate to all other available strains [Figure 6.5, Appendix D.1]. Alternatively, the user can select the “Browser View” (green arrow) and view SNV tracks of all strains on the browser in the context of additional -omics data [Appendix D.1, Appendix D.2].

D.1.2 Genomic Track View

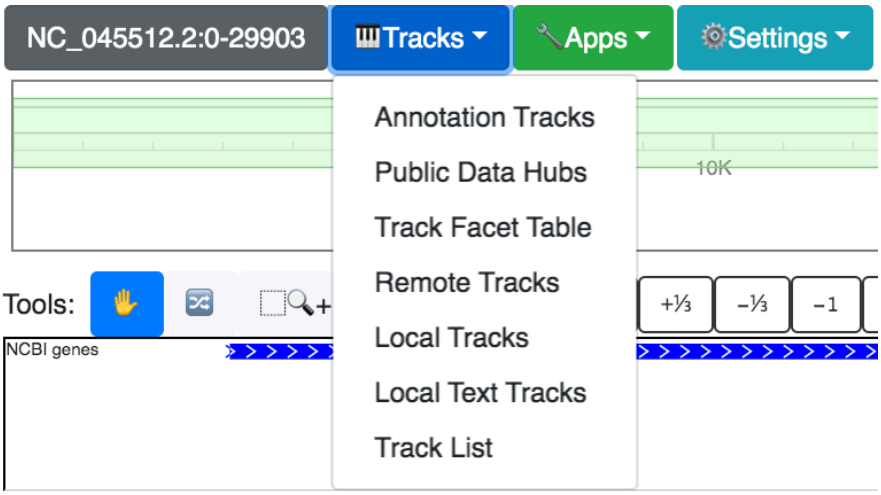
When users select the “Bowser View” (green arrow above), they are taken to our standard browser view layout, adapted from the WashU Epigenome Browser (https://epigenomegateway.wustl.edu/) [2-5]. The WashU Virus Genome Browser maintains the functionality of the Epigenome Browser while also incorporating new features particularly useful for probing virus genomics data.

To the right of the WashU Virus Genome Browser logo is a utility bar (circled above in red), similar to the one provided in the WashU Epigenome Browser. From left to right, this bar provides a platform for navigating the genome to particular regions of interest (gray), provides a drop-down menu for selecting data tracks to be viewed on the browser (blue), provides several applications for viewing, sharing, or saving region views and tracks of interest (green), and provides several customizable options in "Settings", such as the ability to highlight a region of interest and change the track legend width, among others (light blue). Detailed information regarding the use of the region locator can be found within the Epigenome Browser tutorial (https://eg.readthedocs.io/en/latest/usage.html#the-top-menu).

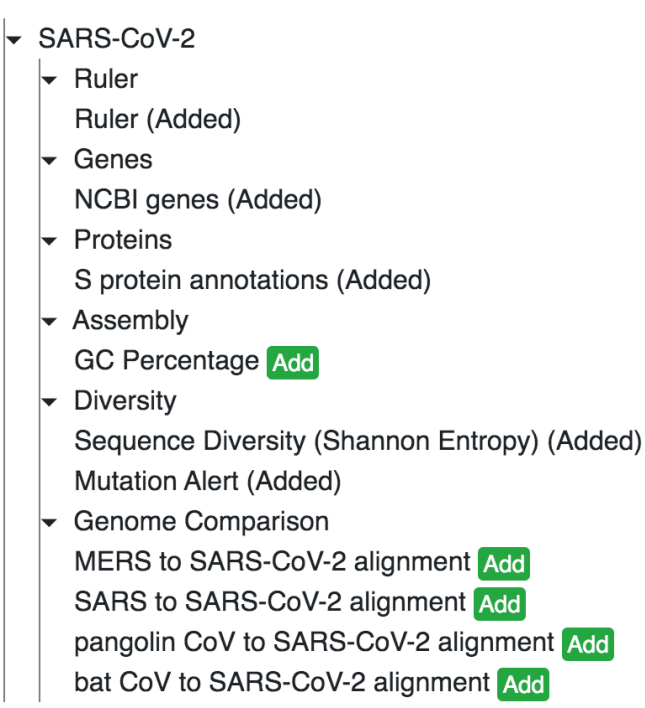
Details regarding the default tracks loaded into view for SARS-CoV-2, which include a sequence diversity track, a mutation frequency track, RNA expression, putative SARS immune epitopes, transcription regulatory sequence locations, recombinations, RNA modifications, and SNV tracks of any strains in the user's cart, are discussed further below.

D.1.2.1 Tracks

The "Tracks" tab lists several functionalities, from loading existing tracks hosted by the browser to user-specified uploads.

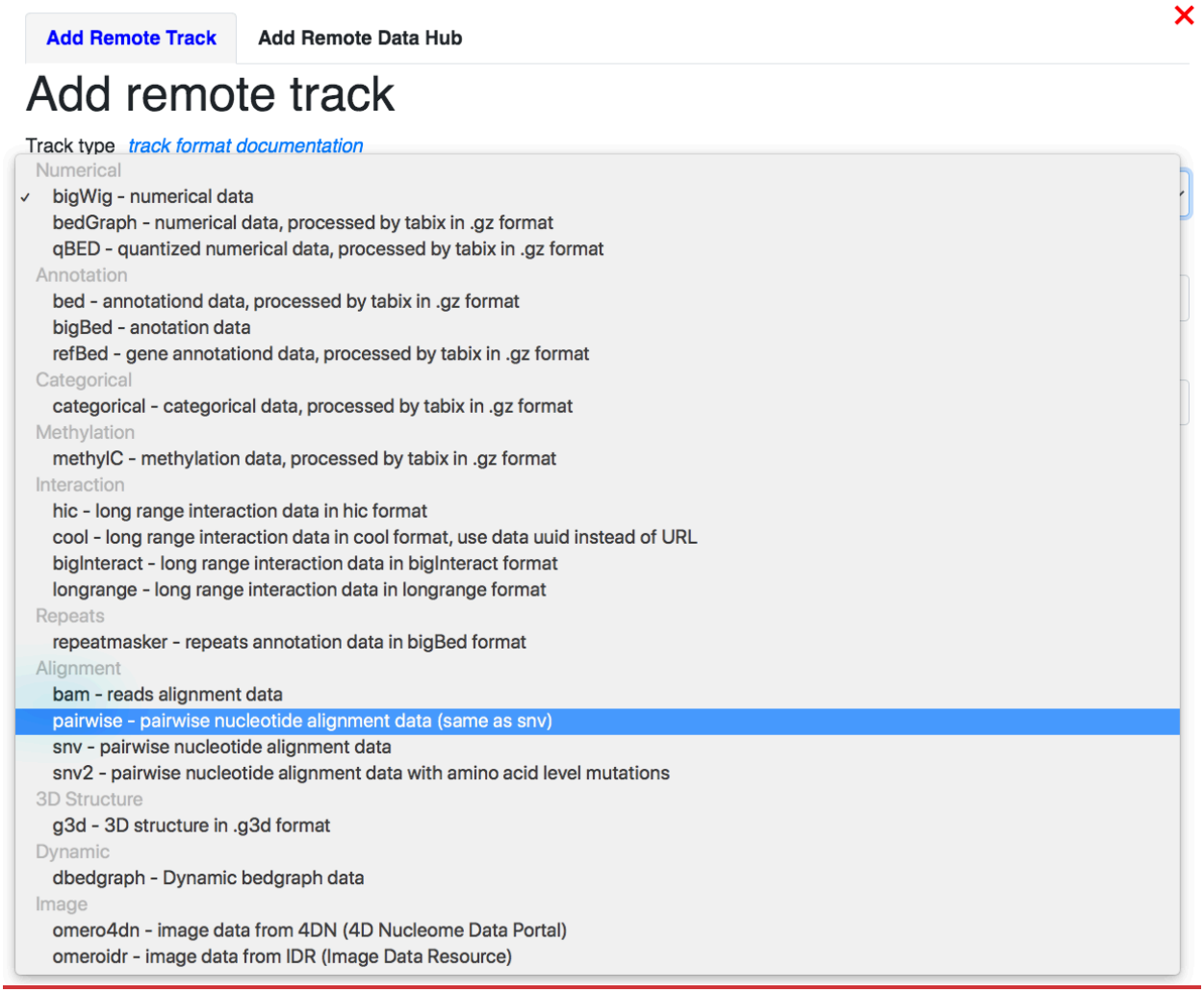


From this drop-down menu, if the user selects “Annotation Tracks”, an expandable menu of pre-existing data tracks appears for the user to select from. For SARS-CoV-2, these tracks include gene annotations, genome comparison tracks, a GC density track, as well as a sequence diversity track and a mutation alert track, all of which can be loaded into the current browser session. Of particular clinical importance, the “Sequence Diversity (Shannon Entropy)” track comprehensively displays the divergence across all GISAID [7] strains at each genomic position. Additionally, the “Mutation Alert” track displays the number of GISAID strains with a mutation at each position. Together, these tracks provide an overview of accumulating mutations and their abundance, aiding in monitoring diagnostic primers for expected effectiveness as the virus evolves [Figure 6.1]. Because of this utility, both the “Sequence Diversity (Shannon Entropy)” track and the “Mutation Alert” track are loaded into view by default.



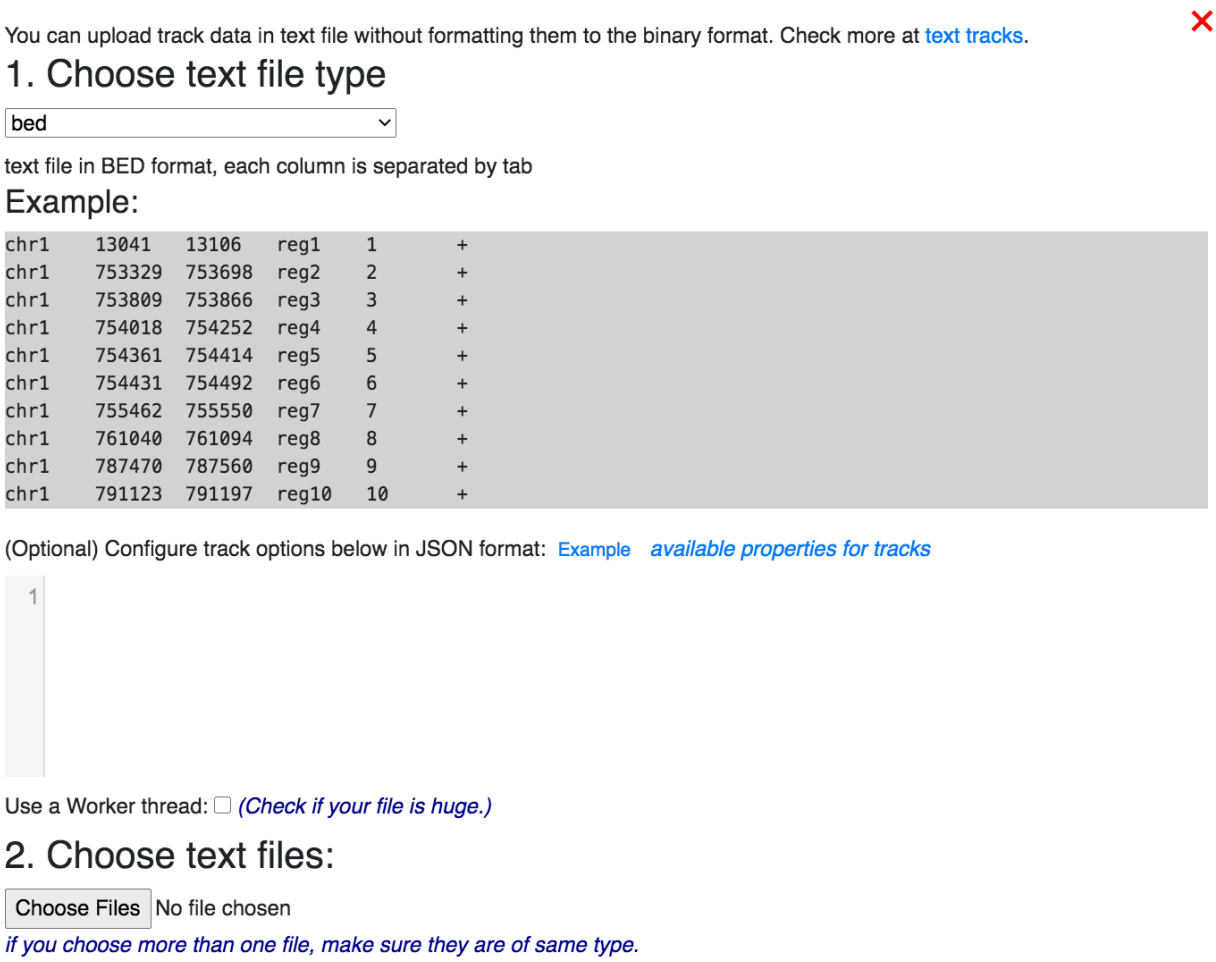
The “Remote Tracks” selection allows for user-upload of individual tracks or data hubs from a hosted url. Upon selection, the user is prompted to select whether he or she would like to upload

an individual track (default tab) or upload a data hub (right-hand tab). If uploading an individual track, the user can then select the drop-down arrow under “Track type” to view all tracks supported by the WashU Virus Browser and select the track type matching their data.

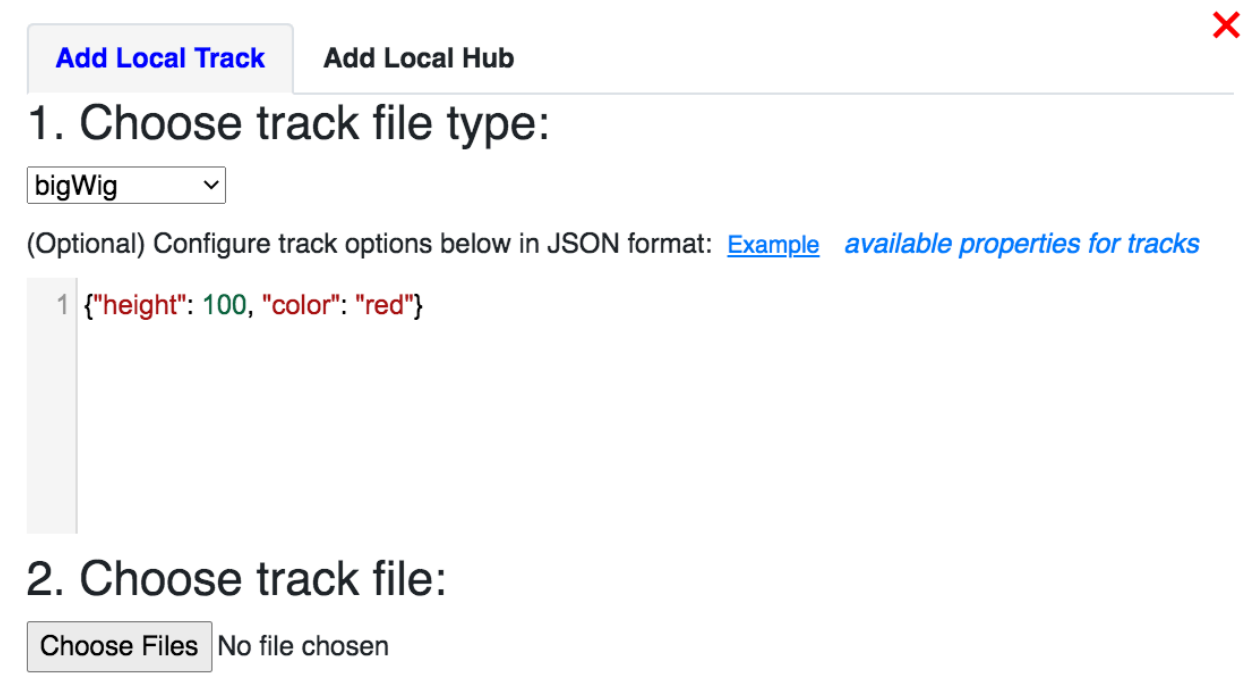


A comprehensive list and description of these tracks can be found here: https://eg.readthedocs.io/en/latest/tracks.html.

To upload a track (in text format, such as .bed.txt, .bedgraph.txt, or .longrange.txt, etc.) directly from one’s computer, the user would select “Tracks” > “Local Text Tracks”. Upon selection, the user can select from the drop-down menu the file type that matches their data. In the default view of this pop-up window, the text file format is “bed” and an example of the text file format is shown below. Optionally, the user can also configure track options, such as metadata, track height, track color, etc., by filling in the text box below.



To upload a track (in binary format, such as .bigwig, .hic, .g3d, .bedgraph.gz(.tbi), .bed.gz, etc.) directly from one's computer, the user would select "Tracks" > "Local Tracks". Here, the user can choose whether to upload a track (default tab) or data hub (right-handed tab) from his/her computer directly. After selecting "Add Local Tracks", the user can then select the track type matching their file via the drop-down menu. As with uploading a track using through "Local Text Tracks", users can optionally specify in the accompanying text box display preferences for their added track, as demonstrated below by selecting "Example".



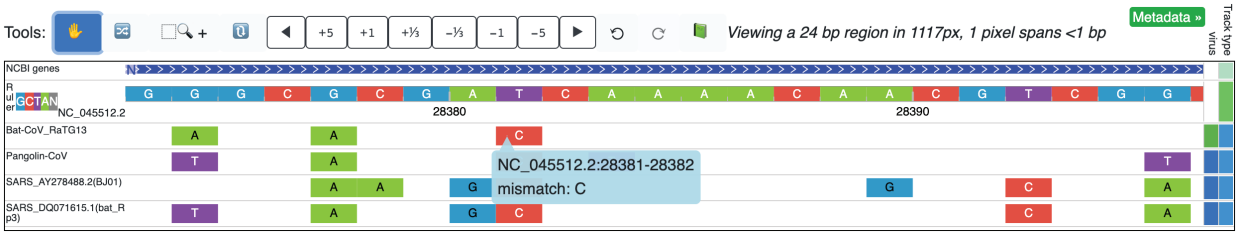
All tracks loaded onto the browser can be easily managed by selecting “Tracks” > “Track List”, as demonstrated below.

Displayed tracks

Label	Track type	Remove
NCBI genes	geneannotation	
Ruler	ruler	
Sequence diversity (Shannon Entropy)	bedgraph	
Mutation Alert	qbed	
Viral RNA expression (nanopore)	bigwig	
Putative SARS Immune Epitopes	bed	
Transcription regulatory sequences (…	categorical	
TRS-L-dependent recombination	longrange	
Viral RNA Modifications	dbedgraph	
EPI_ISL_416425 pairwise alignment	pairwise	
EPI_ISL_402130 pairwise alignment	pairwise	
EPI_ISL_402128 pairwise alignment	pairwise	

Page 1 of 1

In addition to pre-existing track types hosted on the WashU Epigenome Browser, the WashU Virus Genome Browser also introduces the user to a new additional track type, called a "SNV" track, which displays alignment results in "pairwise" format, and its extension – the "SNV2" track – which additionally displays amino acid-level mutations [Appendix D.2]. Both the SNV and SNV2 track types display genomic variations the strain has from the chosen reference (three such SNV tracks are pictured below). If the variant is a mismatch, the track will display the deviated nucleotide following the same color scheme as the reference. If a user selects a particular sequence variation, a pop-up window will show the details of the variation. A very detailed description and tutorial on the SNV/SNV2 tracks are located in Appendix D.2.



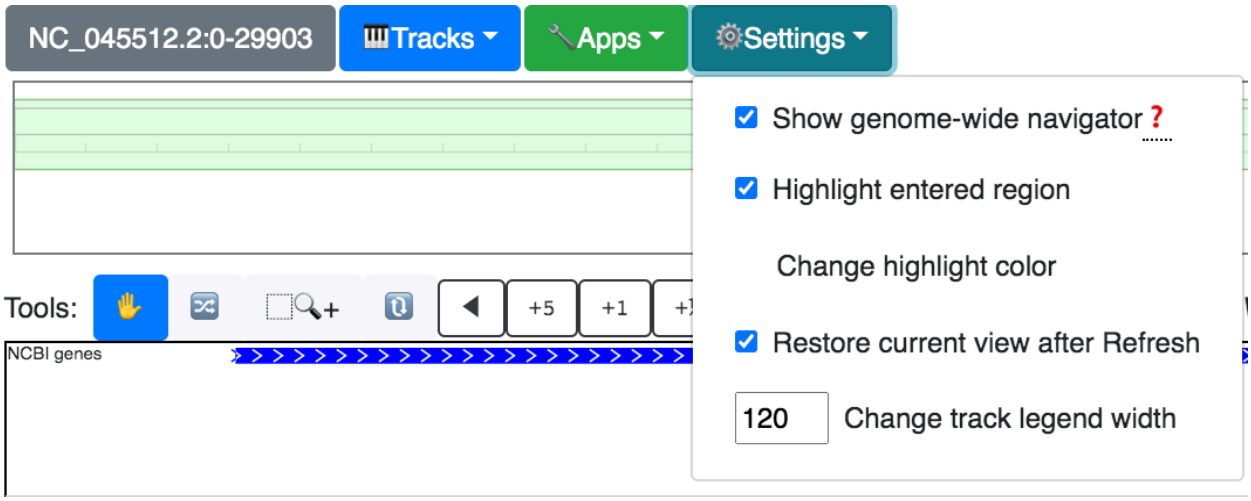
In addition to annotation tracks and user-uploaded tracks, the browser hosts sequencing data of various types organized according to studies, pre-prints, or data providers, in an effort to rapidly integrate data. Detailed information pertaining to navigating existing data hubs is provided in Appendix D.3, and existing data hubs available on the browser as of August 13, 2020 are summarized in Appendix D.5.

D.1.2.2 Apps

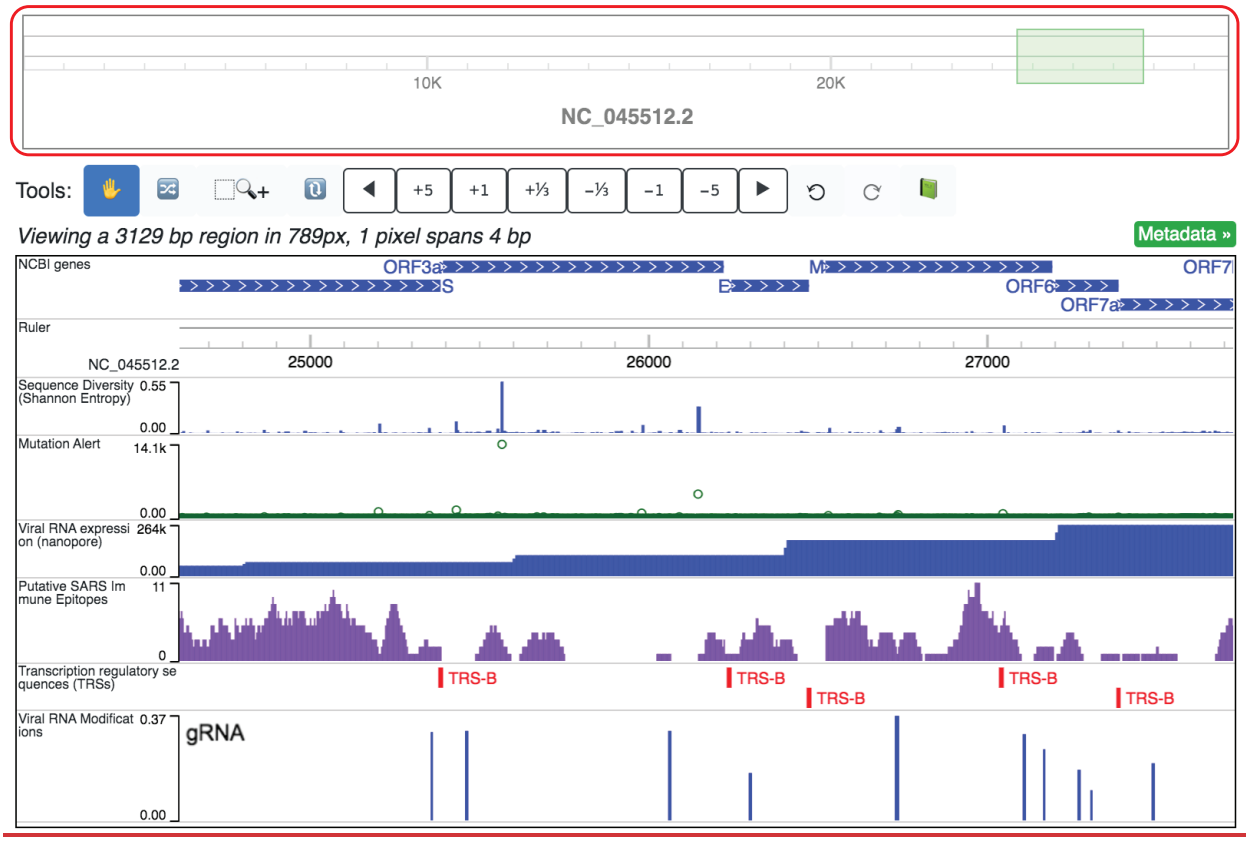
All applications available on the WashU Virus Genome Browser are also available on the WashU Epigenome Browser, and are described in detail in the Epigenome Browser tutorial (https://eg.readthedocs.io/en/latest/usage.html#apps). Of the available apps, "Region Set View" allows the user to visualize several distant genomic regions in the same viewing window. Selecting "Session" allows the user to save their current browser status, generating a session ID that can be shared with collaborators and allowing for easily resuming at a later time. "Fetch Sequence" allows the user to quickly obtain the reference sequence spanning the current view in a FASTA format. Selecting "Screenshot" allows the user to generate publication-quality SVGs or PDFs of the current frame of view, with the option to highlight genomic regions of interest. Additional apps include "Gene Plot", "Scatter Plot", and "Go Live", which are explained in detail in the Epigenome Browser tutorial.

D.1.2.3 Settings

Several browser settings have customizable options which the user may define. When selecting the "Settings" drop-down menu, several functions are provided as shown below (and as described in the Epigenome Browser tutorial: (https://eg.readthedocs.io/en/latest/usage.html#settings):



The first option "Show genome-wide navigator" is a toggle option, allowing the user to hide or show the complete genome layout at the top of the browser (circled in red below), highlighting in green the section of the genome currently being viewed in the browser.

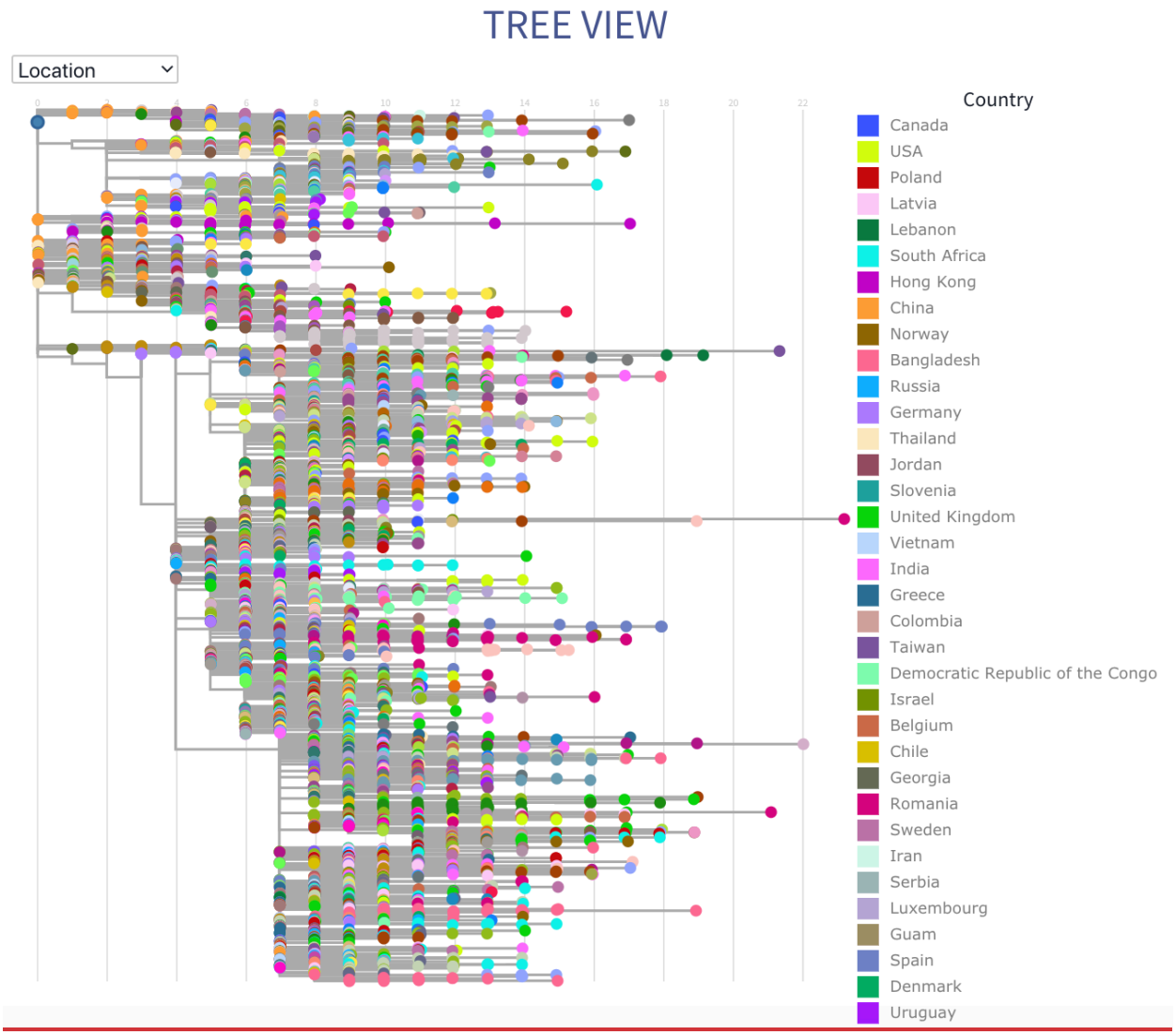


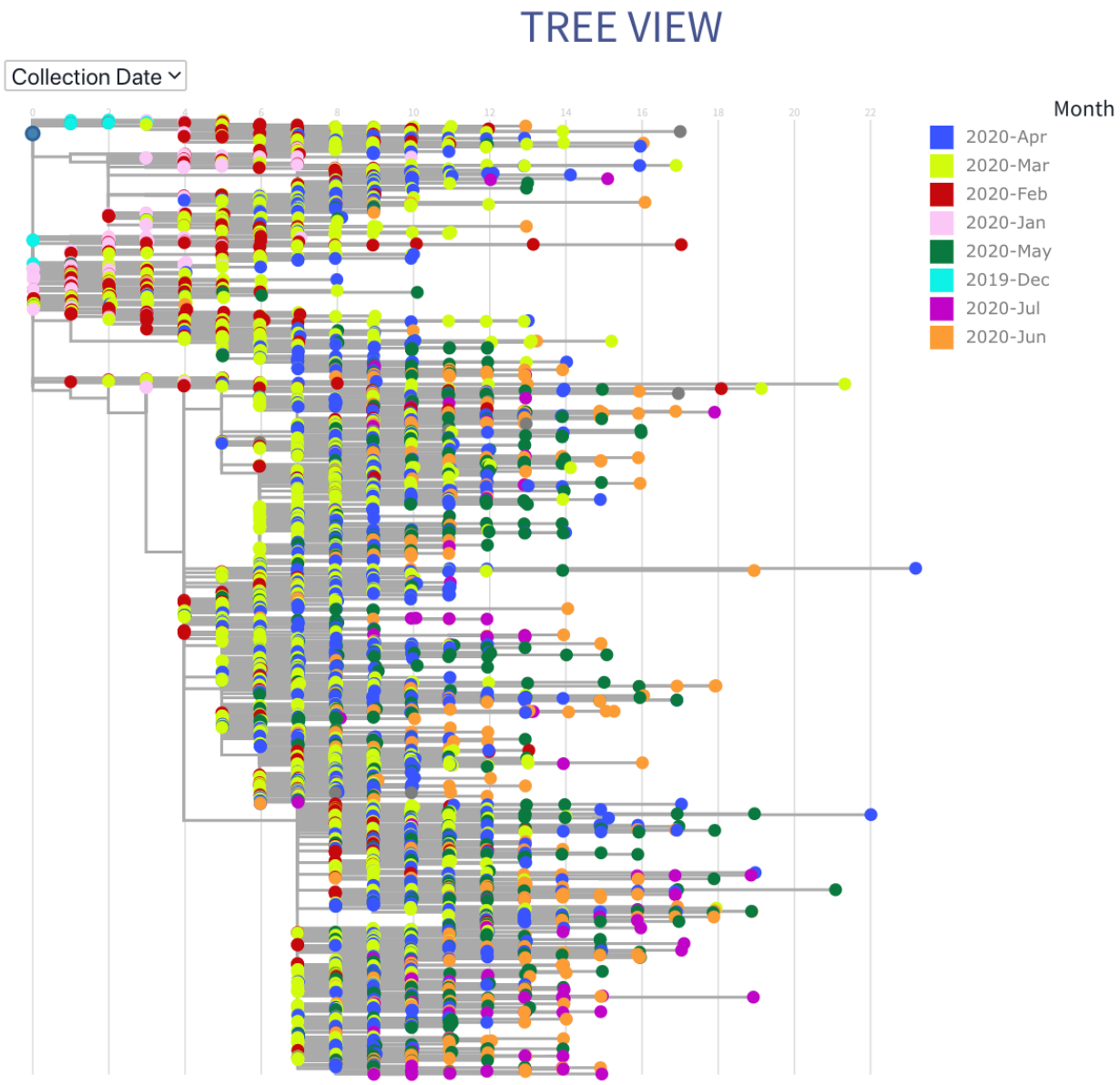
D.1.3 Phylogenetic Tree View

When users select the "Tree View" (orange arrow in the second figure of Appendix D.1), they are taken to a tree visual representation of the strains from the selected reference species. If the selected reference is SARS-CoV-2, the browser will load in a phylogenetic tree parsed from Nextstrain (http://data.Nextstrain.org/ncov_global.json) [1], and therefore includes only strains available from Nextstrain. If the user selects any of the remaining three viruses (SARS-CoV, MERS-CoV, or Ebola), they are directed to a tree consisting of all available strains hosted by NCBI (https://www.ncbi.nlm.nih.gov/nuccore) [8] [Appendix D.5], generated as described in Appendix D.4.

D.1.3.1 Coloring Tree by Metadata

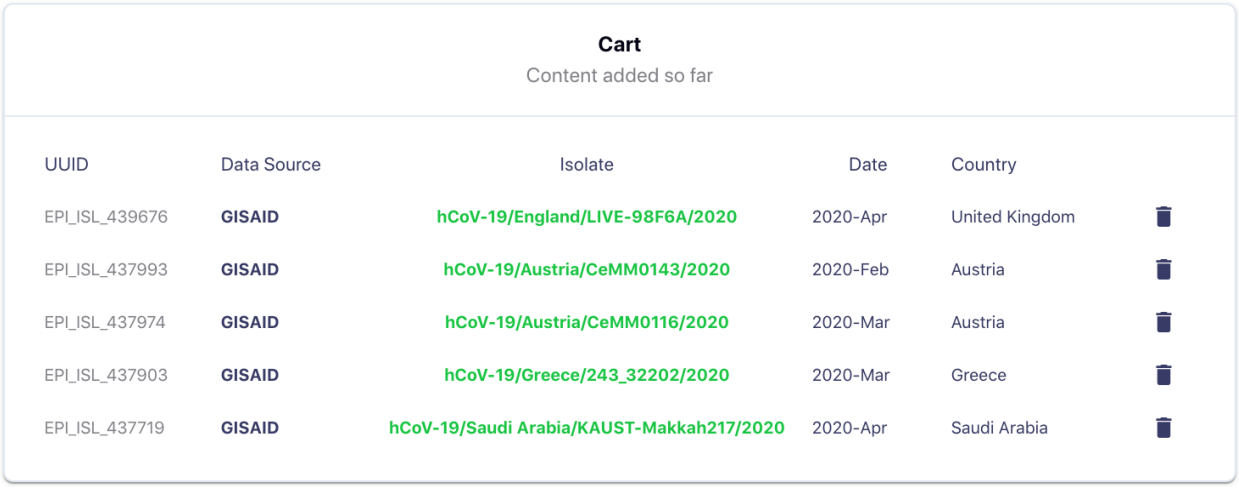
In any of the four trees available, the user is able to organize the tree by color by selecting either "Clade" (as is the default and shown in Figure 6.5), "Location", or "Collection date" (the latter two shown below for SARS-CoV-2).


TREE VIEW
Location
Country
Canada
USA
Poland
Latvia
Lebanon
South Africa
Hong Kong
China
Norway
Bangladesh
Russia
Germany
Thailand
Jordan
Slovenia
United Kingdom
Vietnam
India
Greece
Colombia
Taiwan
Democratic Republic of the Congo
Israel
Belgium
Chile
Georgia
Romania
Sweden
Iran
Serbia
Luxembourg
Guam
Spain
Denmark
Uruguay

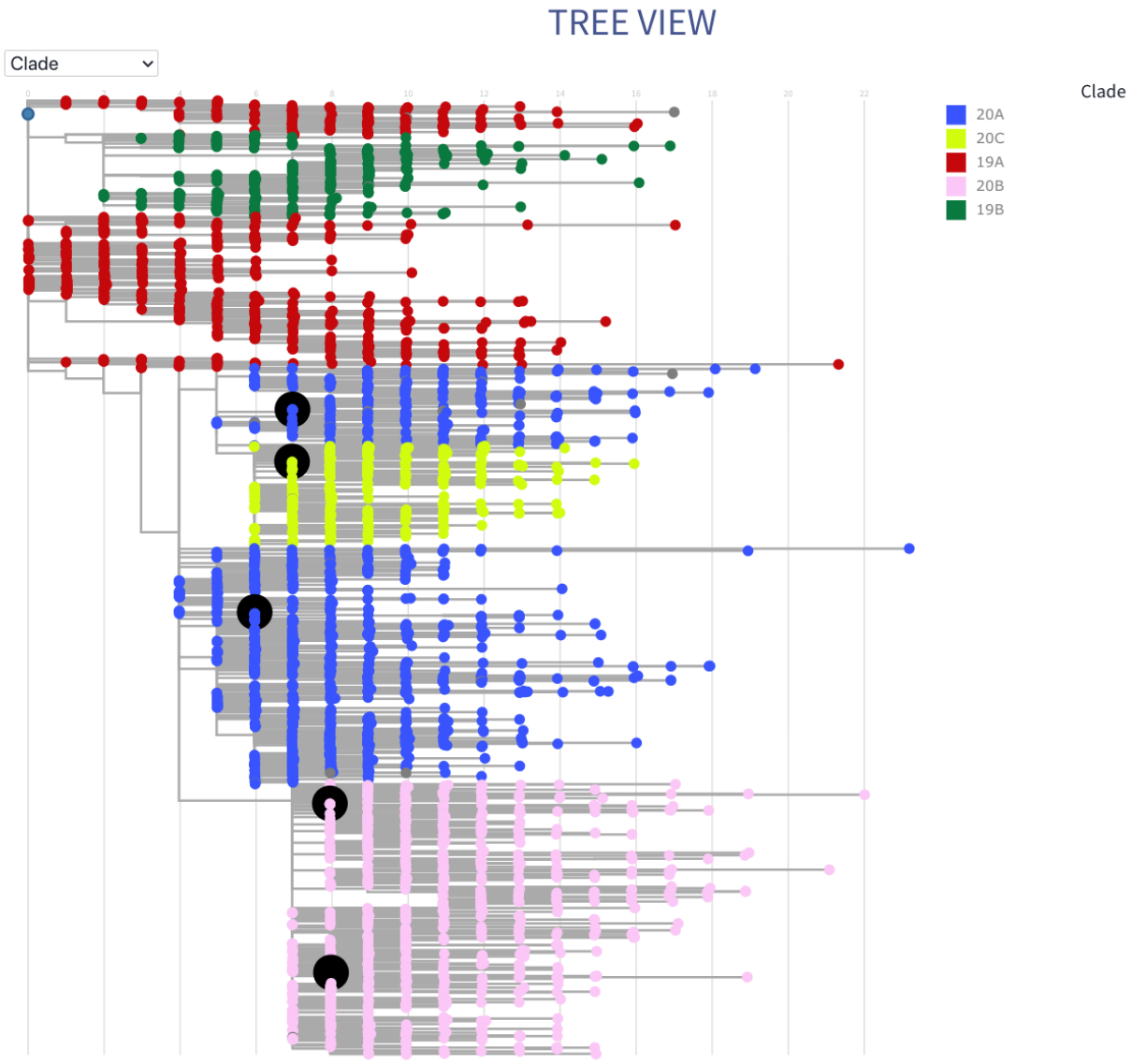


D.1.3.2 Identifying Strains of Interest

If users pre-select strains and add them to their cart from the data table, they can see where their strains of interest fall within the tree. Please keep in mind that for SARS-CoV-2, only strains housed in Nextstrain will be available for viewing in the tree. As an example below, the following tracks have been added to cart.



Upon loading in the tree view, all selected strains are enlarged and colored black, as shown below.



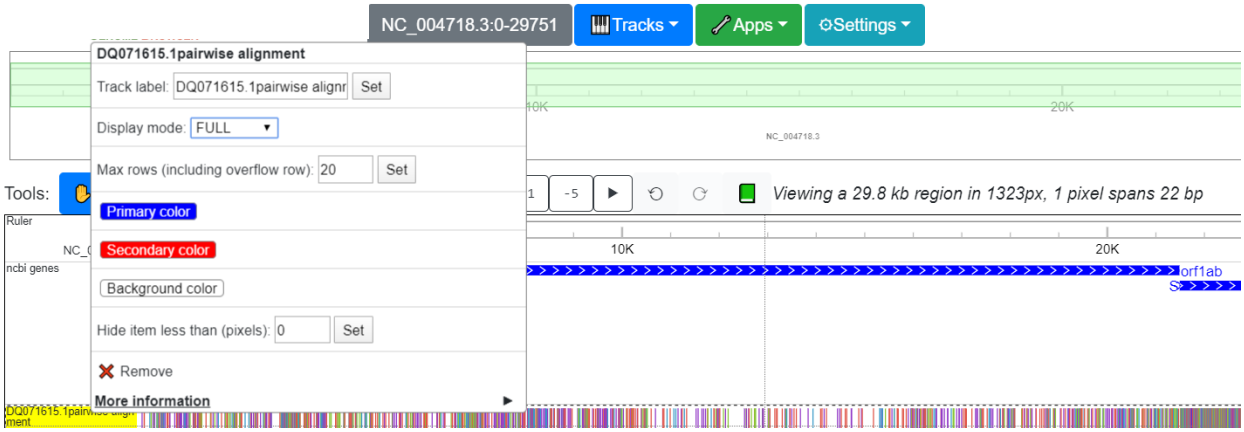
D.2 Displaying variations from the reference via SNV/SNV2 tracks

D.2.1 The SNV Track

D.2.1.1 Overview of the SNV track

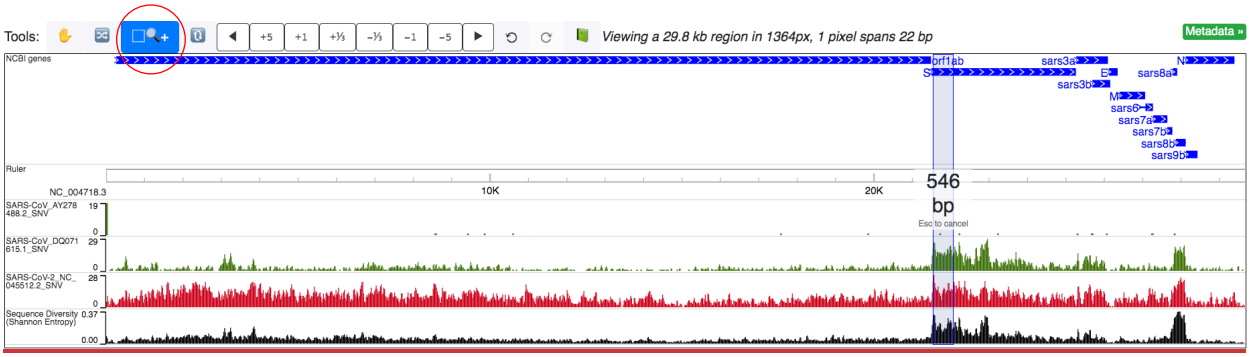
The SNV track is a new track type added to the WashU Virus Genome Browser for visualizing sequence variations from the reference. The track supports two display modes: a “density” mode for a “zoomed-out” view, and a “full” mode for a “zoomed-in” view. The density mode displays the density of variation, suitable for a large, genomic view, whereas the full mode follows a color code, detailing information of each variation, suitable for viewing an individual locus. To switch

between density mode and full mode, right click on the track and use the “display mode” drop-down menu.



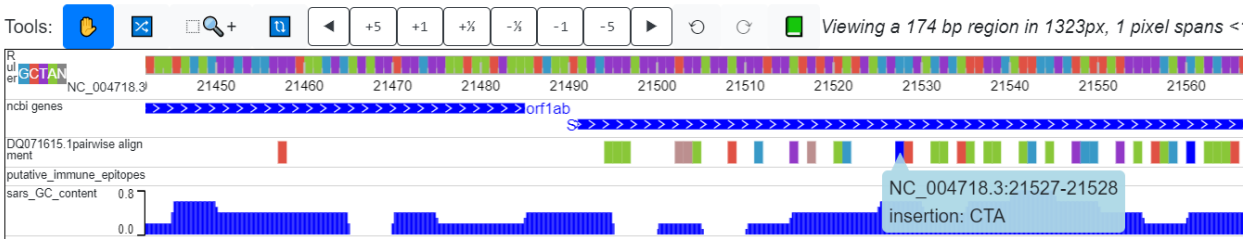
D.2.1.2 Density mode and the “zoomed-out” view

The density mode was implemented because when viewing the entire genome, individual variations can be difficult to see. Instead, the density mode depicts the frequency of mutations across the genome, averaging over genomic intervals, as illustrated in the figure below. Two SARS-CoV strains (AY278488.2 (“BJ01”); DQ71615.1 (“bat rp3”)) and one SARS-CoV-2 strain (the reference strain, NC_045512.2) are aligned to the SARS-CoV reference genome. Sequence variation displayed in density mode shows that the divergence between the SARS-CoV-2 reference genome (red, below) and the SARS-CoV reference genome is higher than the divergence between the two additional SARS-CoV strains (green, below) and the SARS-CoV reference genome. For AY278488.2, the variation from reference is mainly confined to the beginning of the genome, while the remainder of the genome is relatively consistent with the reference. However, for DQ071615.1 (bat-derived), the 5’ end of gene S displays high variation from the reference genome. Likewise, the SARS-CoV Sequence Diversity (Shannon Entropy) track (black, below) shows that the SARS-CoV genome is highly diverse across different strains at gene S. Once a region of interest is identified, the standard magnification tool (circled in red) of the browser can be used to quickly zoom into the region.



D.2.1.3 Full mode and the “zoomed-in” view

When zoomed-in to the nucleotide level and displayed in “Full” mode, a color-coded track indicating all variation from the reference will be shown. The “Full” mode is further detailed below.



D.2.1.3.1 Nomenclature

Reference: The “reference” is the sequence corresponding to the viral species selected by the user. It is the one completely color-coded according to nucleotides and is shown as the “ruler” at the top of the browser view. The references hosted on the browser for the 4 virus species (SARS-CoV-2, SARS-CoV, MERS-CoV, and Ebola) are NCBI reference sequences.

Query: The “query” is the sequence being aligned to the “reference”.

Variation: “Variation” refers to events where the nucleotide of the query at a certain position is different from the reference. It can be a mismatch, insertion, or deletion.

D.2.1.3.2 Color code

Mismatches:

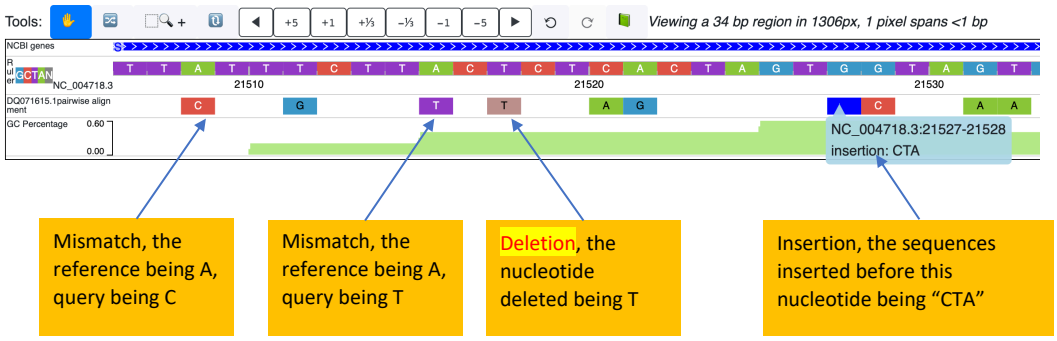
1. A mismatch from the reference is observed at a specific nucleotide, the query being “A”: green (#89C738)
2. The query being “T”: purple (#9238C7)
3. The query being “C”: orange (#E05144)
4. The query being “G”: light blue (#3899C7)
5. The query being “N”: grey52 (#858585)

Deletions:

If a deletion is present in the query (a gap for the query in the pairwise alignment), the nucleotide will be colored “rosy brown” (#BC8F8F).

Insertions:

The reference on the browser is always ungapped. If an insertion into the query happens (which signifies a gap for the reference in a pairwise alignment), the nucleotide proceeding the insertion will be colored “blue”. Details of inserted sequences will be revealed if you click on the nucleotide colored blue.

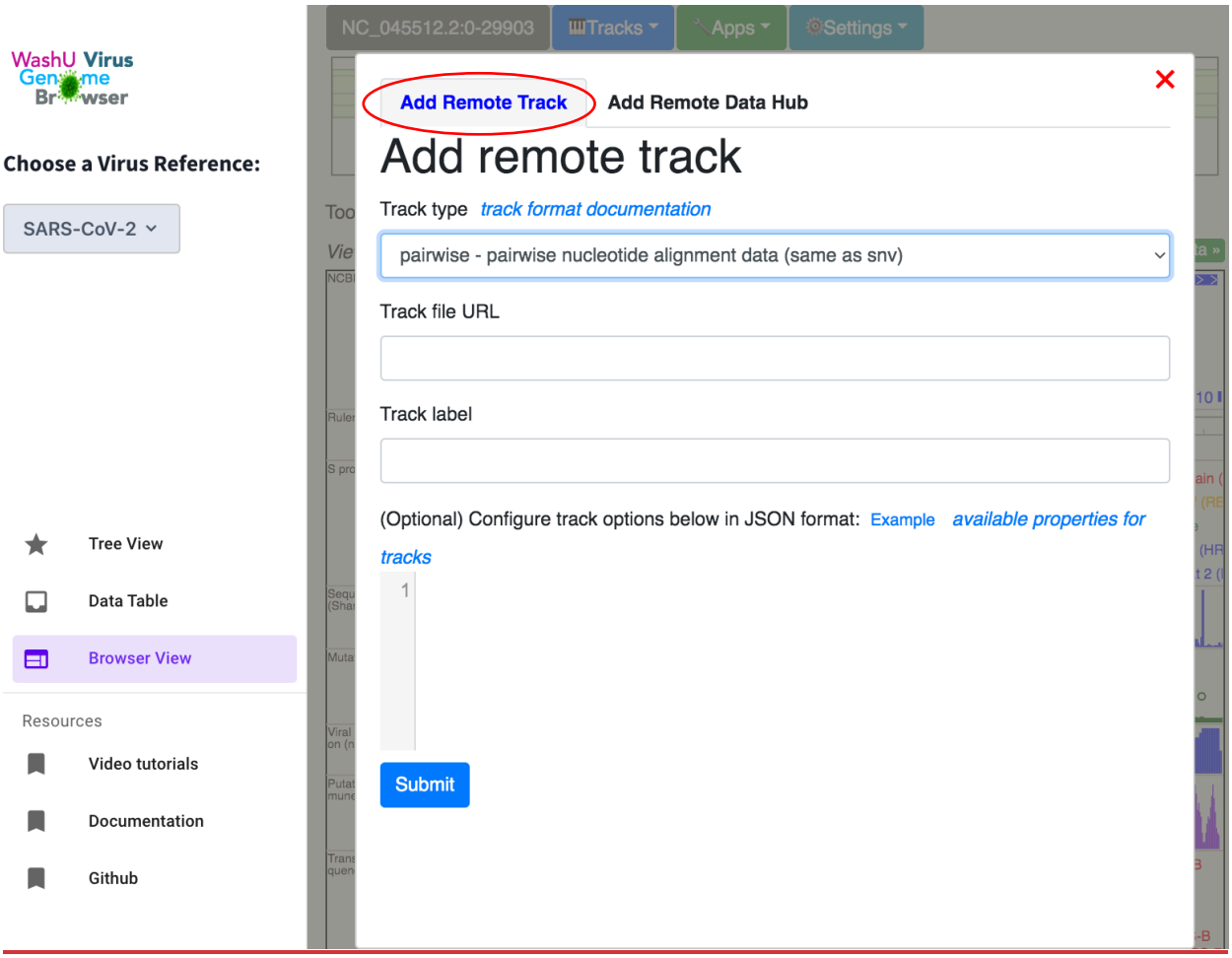


D.2.1.4 Behind the SNV track: the "pairwise" format

Alignment results can differ significantly when different aligners and different parameters are used. However, most aligners return alignments in the same format: FASTA (or markx3 in the case of EMBL aligners). Therefore, the WashU Virus Genome Browser offers scripts for the user to convert their own alignment results into "pairwise"-formatted files, which can be directly uploaded and displayed on the browser. Scripts can be found on our GitHub page: https://github.com/debugpoint136/WashU-Virus-Genome-Browser.

D.2.1.4.1 Uploading interface

In the browser view, select "Tracks" and then "Remote Tracks". This will lead the user to an upload interface. Once there, select "pairwise" as the track type and enter the track's URL.



D.2.1.4.2 The pairwise format

The pairwise format is an extension of the .bed format, where the 4th column contains variations from the reference. Specifically:

Column 1: Name of the reference to which the query is aligned
Column 2: The start position in the reference
Column 3: The end position in the reference
Column 4: Variation type and details

For columns 2 and 3, since the pairwise format is per-nucleotide, “end” usually equals “start” + 1 (the only exception is a deletion, and when several consecutive nucleotides are deleted, they can be merged into 1 line). Also, the browser is 0-based, which means that if there is a mismatch at the first nucleotide, column 2 would be “0” while column 3 would be “1”.

For column 4, the format is “variation_type:detail”. Variation types are: “insertion”, “deletion”, and “mismatch”. For “insertion”, “detail” is the sequence inserted before this nucleotide. For “deletion”, the “detail” is the nucleotide of the reference at this position that was deleted. For “mismatch”, “detail” is the nucleotide of the query.

Please note that matches between the query and the reference are not coded in the pairwise format.

```
NC_004718.3 96      97      mismatch: T
NC_004718.3 140     141     mismatch: C
NC_004718.3 142     143     mismatch: G
NC_004718.3 258     259     mismatch: A
…       …       …       ...
NC_004718.3 3089    3090    insertion: GG
NC_004718.3 3093    3094    insertion: CTCA
NC_004718.3 21527   21528   insertion: CTA
NC_004718.3 21560   21561   insertion: C
…
NC_004718.3 3059    3062    deletion: AGC
NC_004718.3 3223    3224    deletion: A
```

Please note that the files need to be sorted, zipped, and indexed before uploading, just as you would need to do for any .bed files. To prepare the files, please use the following commands:

```
sort -k1,1 -k2,2n snvfile > snvfile.sort
bgzip snvfile.sort
tabix -p bed snvfile.sort.gz
```

D.2.1.4.3 Generating the pairwise format

We offer a script ("publicConvertMarkx3.py") to generate a pairwise-formatted file from any pairwise alignment result in markx3 or FASTA format. The script calls another script "convert_tsv_to_bed_and_cat.sh", both of which are located on our GitHub page: https://github.com/debugpoint136/WashU-Virus-Genome-Browser. The requirements for conversion are as follows:

1. The reference should be the first sequence.
2. Only two sequences should be present in the file (as is the case for pairwise alignment).

An example FASTA-formatted pairwise alignment result is shown below:

```
>test_reference ..
ATGAGTCTCTCTGATAAGGACAAGGCTGCTGTGAAAGCCCTATGG------A

>test_query..
CTG--TCTC-CTG---CCGACAAGACCAACGTCAAGGCCGCCTGGGGTAAGA
```

The output files will be automatically zipped and ready for upload on the WashU Virus Genome Browser.

D.2.1.4.4 Batch alignment from FASTA to pairwise format

We offer another script ("publicAlignment.py", located here: https://github.com/debugpoint136/WashU-Virus-Genome-Browser/blob/master/scripts/snv/publicAlignment.py) that can perform pairwise alignments in batch (using the EMBL aligner "stretcher" [9]) and directly generate files in pairwise format that can be uploaded to the browser as SNV tracks.

D.2.1.4.5 Batch upload as json files

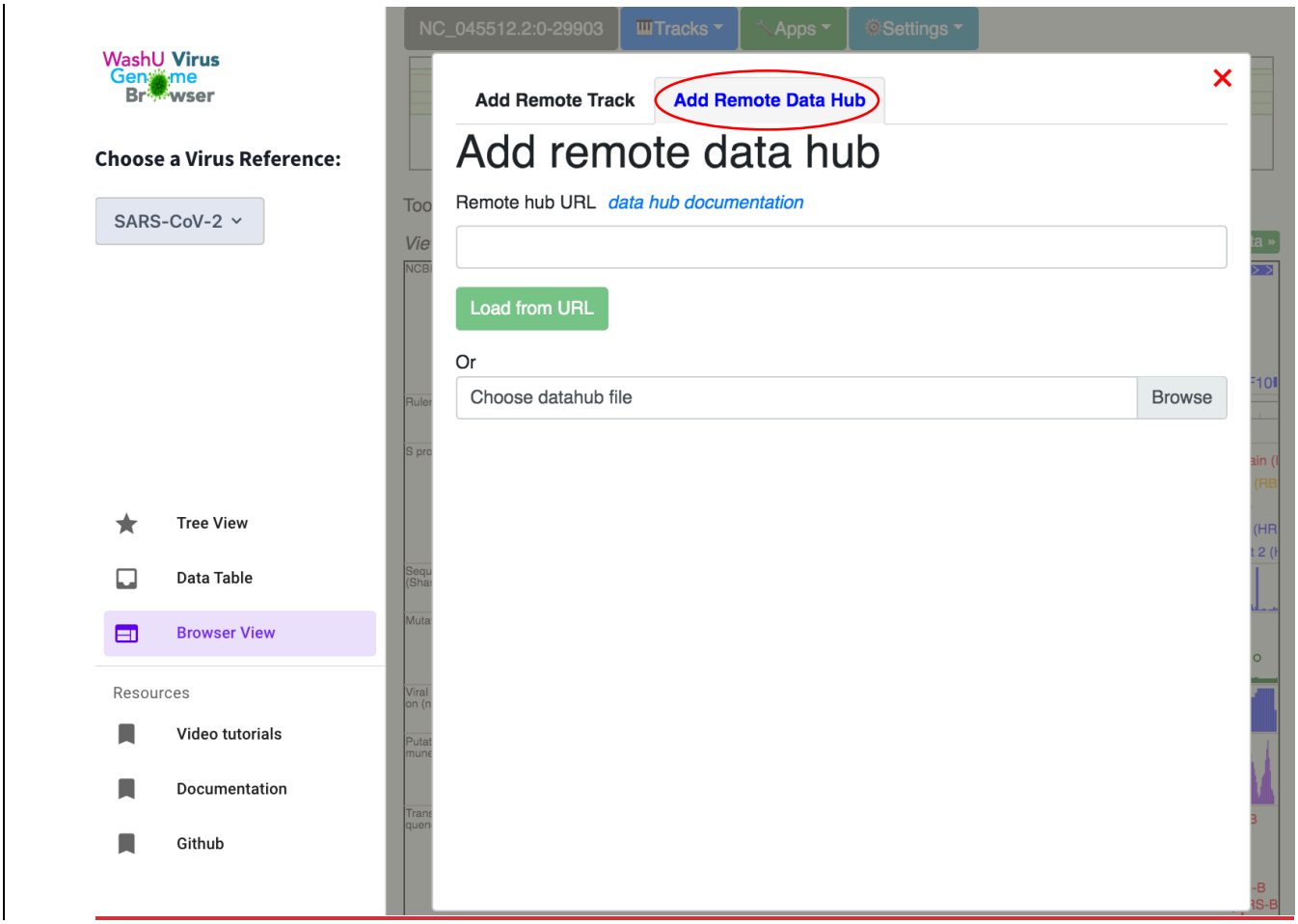
We offer another script ("publicJsonGen.py", located here: https://github.com/debugpoint136/WashU-Virus-Genome-Browser/blob/master/scripts/snv/publicJsonGen.py) that takes in a tab-delimited text file (.tsv file) listing the web location and track type of individual files, and outputs a .json file that can be used to upload multiple tracks in batch.

A .tsv file should have a format similar to that shown in the table below in order to be converted.

name	url	track_type	virus
SARS_AY278488.2_SNV	https://your.url.to.file1	pairwise	SARS
SARS_DQ071615.1_SNV	https://your.url.to.file2	pairwise	SARS
SARS_AY278488.2_SNV	https://your.url.to.file3	pairwise	SARS

D.2.1.4.6 Upload a json-formatted data hub

To upload a json-formatted data hub, in the browser view, select "Tracks" and then select either "Remote Tracks" or "Local Tracks" (depending on whether the json file is stored remotely or locally). This will lead the user to an upload interface. Once there, click the "Add Remote Data Hub" tab, then upload the json file. See Appendix D.1 for more details.



D.2.2 The SNV2 Track

D.2.2.1 Functionality of SNV2 tracks

The SNV2 track is an extension of the SNV track, formatted with extreme flexibility. It is equipped with the ability to show amino acid-level mutations, as demonstrated in the figure below. It also supports everything offered with the SNV track, including color-coded mutations, zoomed-in and zoomed-out view, and detailed information upon selection of specific mutations.

The SNV2 format is essentially a combination of the categorical track format and bed track format, producing the equivalent of color-coded bed tracks, which can be used for a variety of purposes. We provide scripts (https://github.com/debugpoint136/WashU-Virus-Genome-Browser/tree/master/scripts/snv2/) to generate SNV2 tracks for displaying potential amino acid mutations, but we encourage users to customize them for their specific needs.

D.2.2.2 Defining the format of SNV2 tracks

SNV2 tracks can have as many columns as necessary. The first 3 columns encode the genome positions, and the 4th column encodes the specific variation type (as outlined above in Appendix D.2.1.4.2, excluding details which are listed in additional columns). All columns beginning with the 4th column will be shown as text upon mouse selection. The mapping between categories and colors is user-specified in the accompanying input json.

column	detail
1	Reference virus name (such as NC_045512.2 for SARS-CoV-2)
2	Start position of the reference genome, 0-based, inclusive
3	Stop position of the reference genome, 0-based, not inclusive
4	Variation type, controls color code, will also show as text in pop-up window upon

	selection
5, 6, 7…	Text columns, will show as text in pop-up window upon selection

The 4th column is mapped to colors through specifying “segementColors” in the “options” part of the data hub json file. A detailed tutorial with examples on how json files are formatted is available here: https://epigenomegateway.readthedocs.io/en/latest/datahub.html?highlight=json. However, if SNV2 tracks are being used to display amino acid mutations, uploading them as a json file is optional because we have a default color-code mapping scheme:

```
"options": {
   "segmentColors": {
      "un_sequenced": "Linen",
      "noncoding_insertion": "LightGrey",
      "noncoding_deletion": "LightGrey",
      "noncoding_mismatch": "LightGrey",
      "silent": "DimGrey",
      "frameshift": "FireBrick",
      "missense": "CornflowerBlue",
      "AA_deletion": "CornflowerBlue",
      "AA_insertion": "CornflowerBlue",
      "N_mask": "Linen",
      "deletion_mask": "Linen"
   }
}
```

D.2.2.3 Using SNV2 tracks

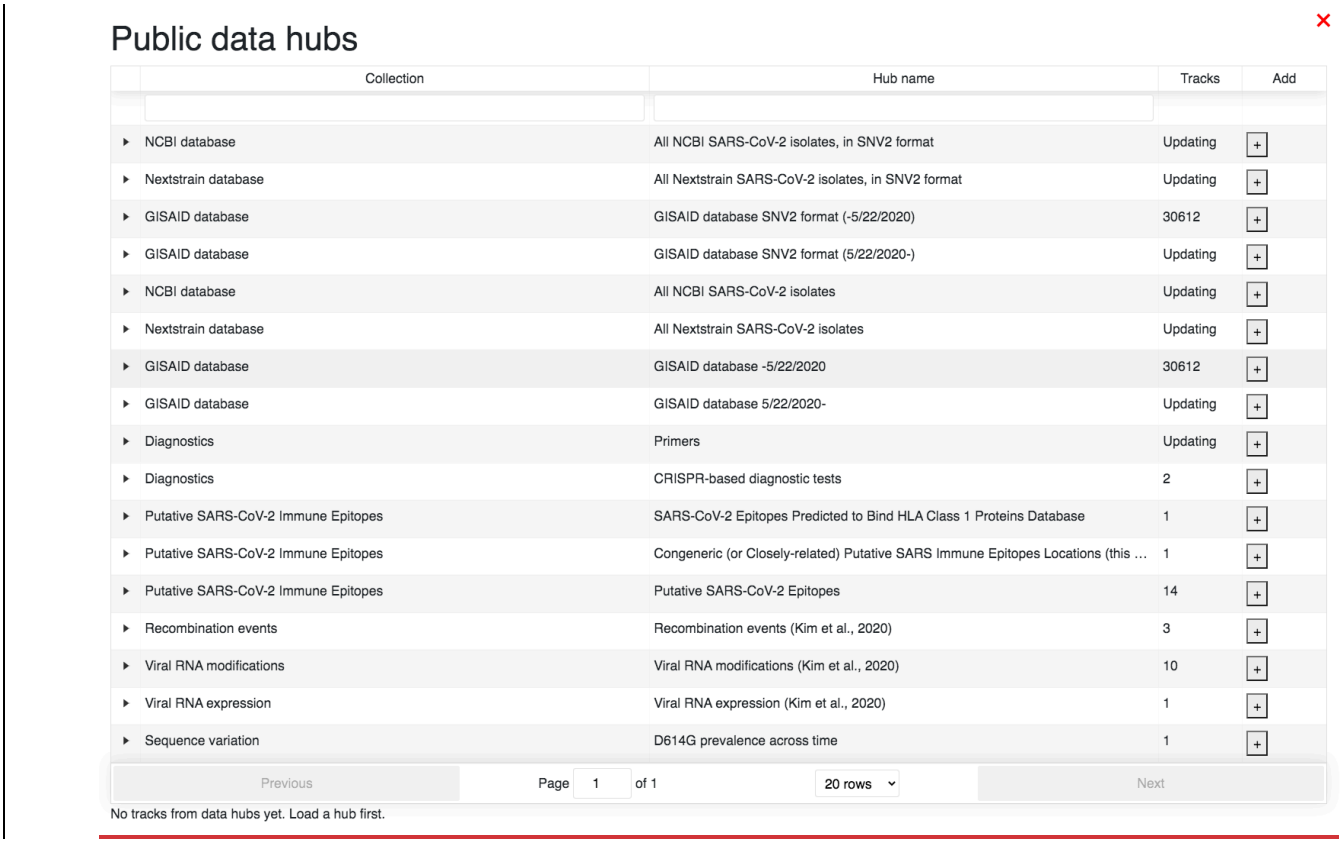
Please visit our documentation page, https://virusgateway.readthedocs.io/en/latest/snv2.html, for a quick demonstration of using SNV2 tracks. A set of scripts for generating SNV2 tracks to visualize amino acid-level mutations on the WashU Virus Genome Browser is available at https://github.com/debugpoint136/WashU-Virus-Genome-Browser/tree/master/scripts/snv2/.

D.3 Public data hubs

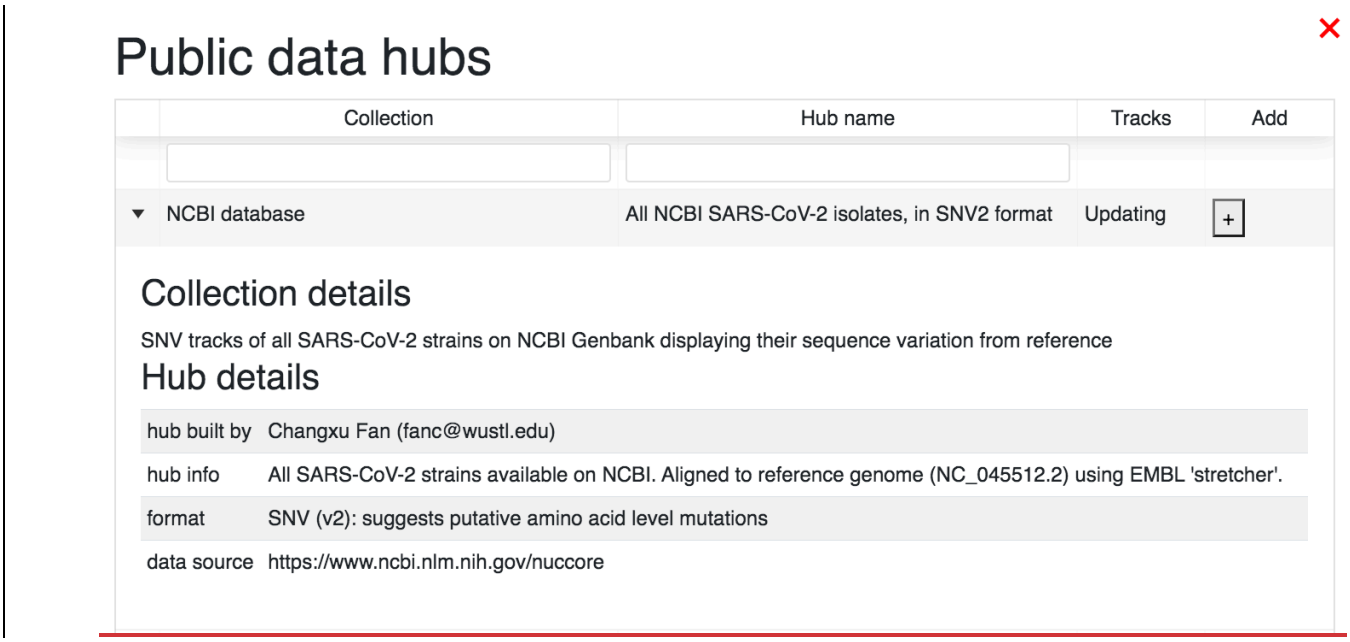
As individual research groups continue to rapidly study SARS-CoV-2, there has been a surge in available SARS-CoV-2 -omics data, ranging from thousands of sequenced strains, to host immune responses, and viral expression and modifications. As this data is becoming available, we are integrating relevant findings onto our browser in the form of public data hubs for efficient upload of several related tracks for easy visual comparisons. Here, we will introduce how to load a specific data hub of interest and also briefly introduce all existing public data hubs (as of August 13, 2020).

D.3.1 Loading in a public data hub

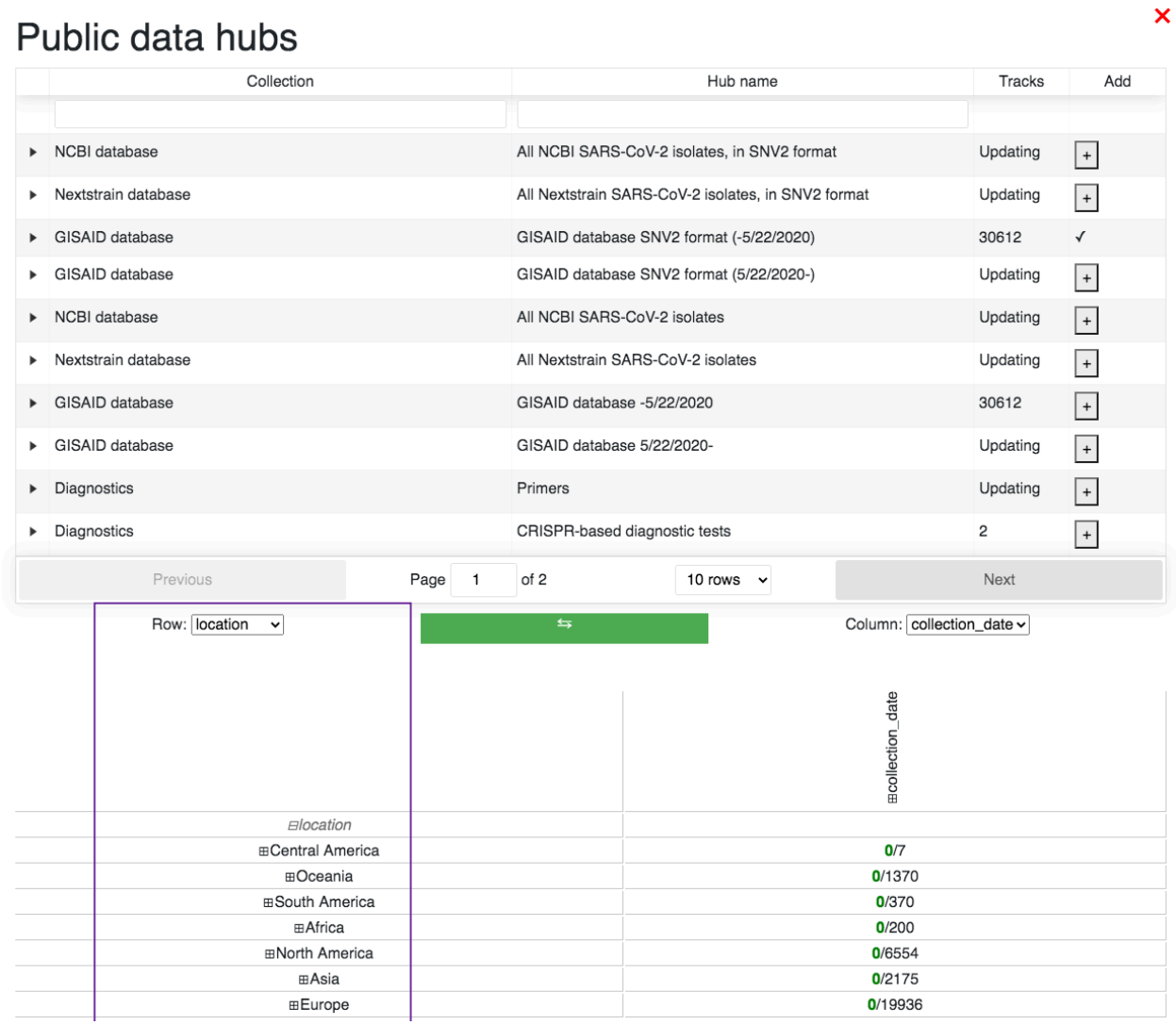
From the main genomic track-based browser view, information regarding existing data hubs can be viewed by selecting “Tracks” > “Public Data Hubs”. This will populate a table with information regarding data hub collection names, hub names, and numbers of tracks. Data hubs that are continuously updated display the term “Updating” instead of the number of tracks.



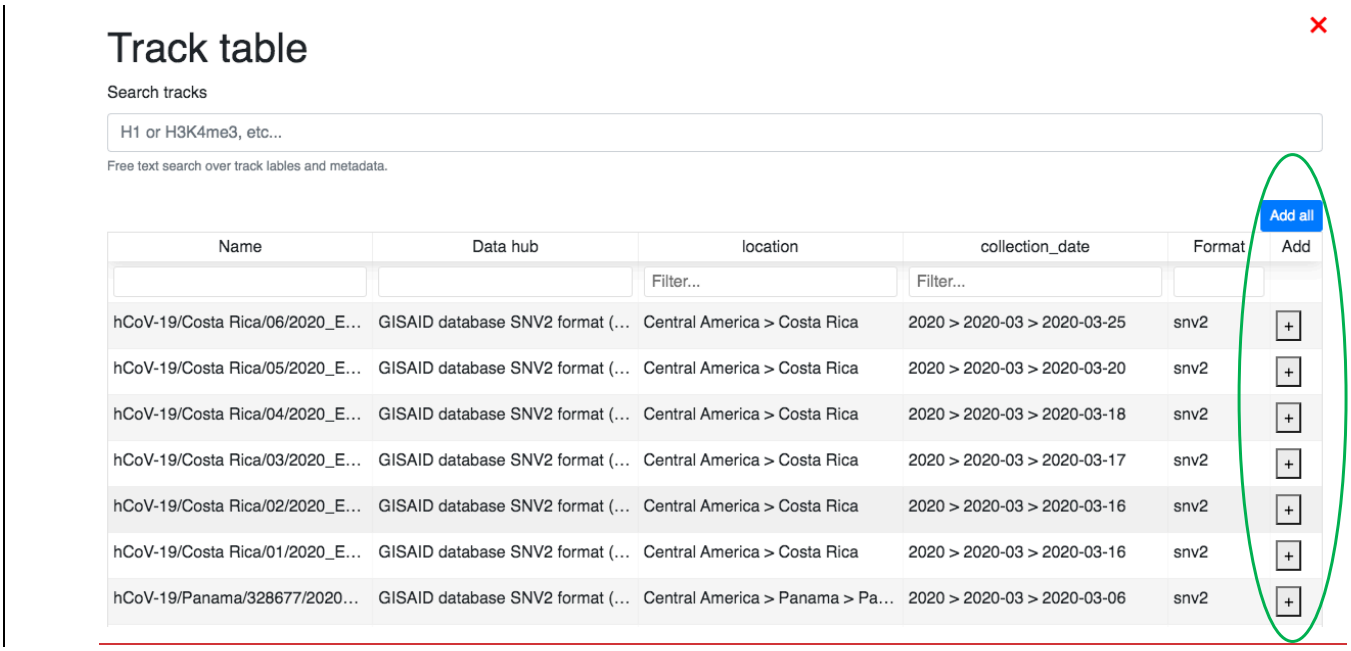
Users can get more information regarding a data hub of interest, such as the data source and track formatting, by selecting the collection, as demonstrated below.



To load tracks from a given data hub into the browser, the user can select the "+" button under the "Add" column. This will populate a sortable metadata table where the user can select specifically which tracks to add from the data hub. In the example below, data are being sorted by location, as highlighted in the purple box.



After selecting a metadata term of interest, such as “Central America” above, the user can add the tracks to their browser view by selecting the “+” to the right of the track(s), as shown in green below.



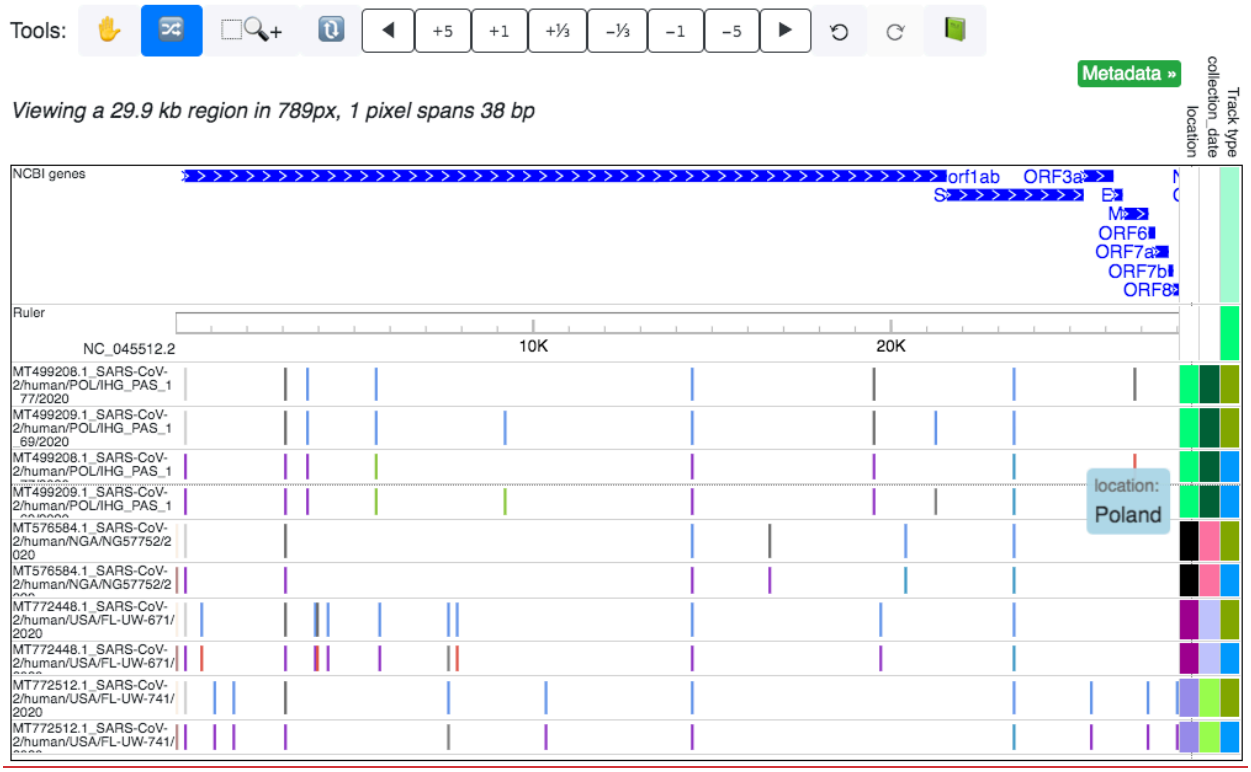
Data tracks from all existing data hubs or user-uploaded tracks can be managed in a similar way by selecting “Tracks” and then “Track Facet Table”.

D.3.2 Introducing currently available public data hubs

Here, we introduce all available public data hubs as of August 13, 2020, organized by Collection. Please note that this selection will continue to expand as new studies are made available regarding SARS-CoV-2 and related host genomic data.

D.3.2.1 NCBI database

As of August 13, 2020, 15,215 SARS-CoV-2 sequences have been retrieved from NCBI (https://www.ncbi.nlm.nih.gov/nuccore) [8]. Pairwise alignments between all downloaded strains and the reference have been added to the NCBI database public data hub collection and are available for viewing on the WashU Virus Genome Browser in the form of SNV tracks (data hub name: “All NCBI SARS-CoV-2 isolates”) and SNV2 tracks (data hub name: “All NCBI SARS-CoV-2 isolates, in SNV2 format”) [Appendix D.2]. All strains can be added to the browser at once (this is not recommended due to the large number of strains), or the user can sort the strains by location and collection date to pre-filter for specific strains of interest. Metadata information regarding location and collection date can also be displayed on the right side of the tracks once loaded into the browser view, by selecting “Metadata” and specific terms of interest, as shown below.



D.3.2.2 Nextstrain database

As of August 13, 2020, 4,263 SARS-CoV-2 sequences have been collected from Nextstrain [1]. Pairwise alignments of all strains to the reference have been parsed from the Nextstrain json (http://data.nextstrain.org/ncov_global.json) and added to the Nextstrain database public data hub collection and are available for viewing on the browser in the form of SNV tracks (data hub name: "All Nextstrain SARS-CoV-2 isolates") and SNV2 tracks (data hub name: "All Nextstrain SARS-CoV-2 isolates, in SNV2 format") [Appendix D.2]. As described above, strains of interest can be pre-selected based on specific metadata terms.

D.3.2.3 GISAID database

As of August 13, 2020, 80,850 SARS-CoV-2 sequences have been retrieved from GISAID [7]. Pairwise alignments between all downloaded strains and the reference sequence have been added to the GISAID database public data hub collection and are available for viewing on the browser in the form of SNV tracks (data hub names: "GISAID database -5/22/2020" and "GISAID database 5/22/2020-") and SNV2 tracks (data hub names: "GISAID database SNV2 format (-5/22/2020)" and "GISAID database SNV2 format (5/22/2020-)") [Appendix D.2]. As described above, strains can be selected for upload by pre-filtering based on specific metadata terms of interest.

D.3.2.4 Diagnostics

The Diagnostics collection currently houses two separate data hubs, both of which are described below. Data hubs encompassed in this collection contain relevant annotations pertaining to diagnostic testing.

D.3.2.4.1 Primers

As of August 13, 2020, the locations of primers for SARS-CoV-2 testing have been made available for the USA (3 sets of Centers for Disease Control and Prevention (CDC) primers), China, Hong Kong, France, Germany, Japan, and Thailand. All primer sequence locations are made available in the Primers database public data hub, and can be loaded in all at once or based on country of interest. An example highlighting mutations within China CDC primers is outlined in Figure 6.1.

D.3.2.4.2 CRISPR-based diagnostic tests

The CRISPR-based diagnostic tests data hub consists of two tracks: a SHERLOCK diagnostic test track displaying primer and guide RNA sequence locations used in the CRISPR-Cas13a-based SHERLOCK assay for detecting SARS-CoV-2 (https://www.broadinstitute.org/files/publications/special/COVID-19%20detection%20(updated).pdf), and a DETECTR diagnostic test track displaying the primer and guide RNA sequence locations used in the CRISPR-Cas12-based DETECTR assay for detecting SARS-CoV-2 [10]. An example highlighting the location of a DETECTR E guide-RNA is shown below, revealing low sequence diversity at this location.

D.3.2.5 Putative SARS-CoV-2 immune epitopes

The collection "Putative SARS-CoV-2 Immune Epitopes" currently consists of three data hubs: "SARS-CoV-2 Epitopes Predicted to Bind HLA Class 1 Proteins", "Congeneric (or Closely-related) Putative SARS Immune Epitopes", and "Putative SARS-CoV-2 Epitopes". All three data hubs feature the locations of immune epitopes, some of which were identified in SARS-CoV and maintain sequence conservation in SARS-CoV-2.

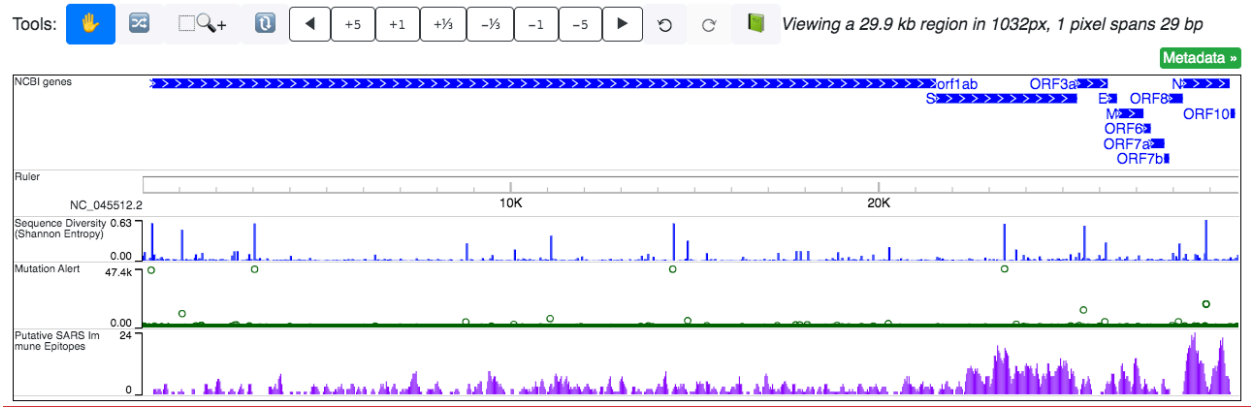
D.3.2.5.1 SARS-CoV-2 epitopes predicted to bind HLA class 1 proteins

Predicted SARS-CoV-2 epitopes likely to bind class 1 MHC proteins were made available in the pre-print Campbell, et al., 2020 [11]. Locations of the predicted sequences within the SARS-CoV-2 genome were identified (those with 100% sequence similarity and on the positive strand [Appendix D.4]) and their locations (x-axis) as well as the number of unique strain IDs reporting the peptide (y-axis) are shown when loading in the track "Antigenic_peptide_predictions".

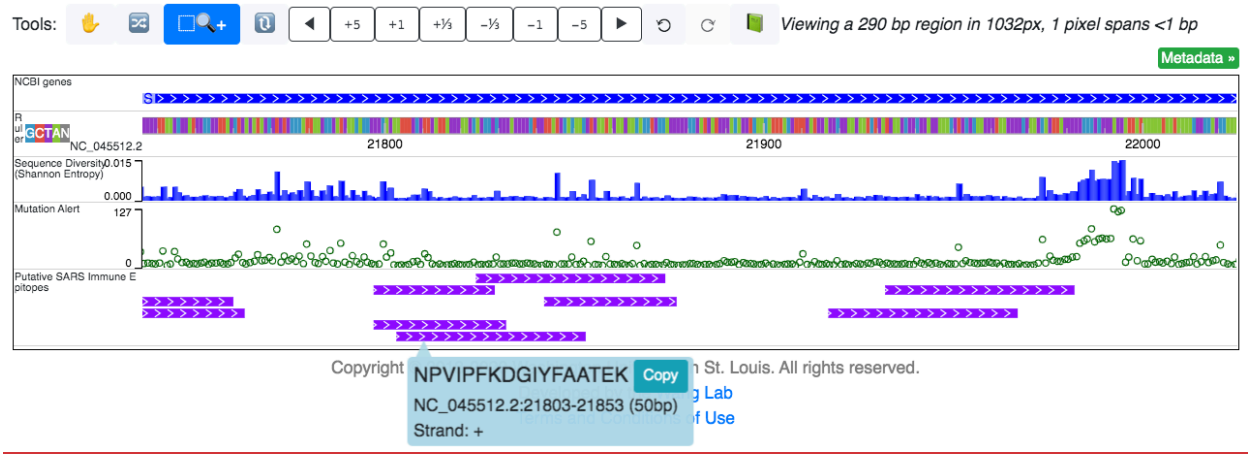
D.3.2.5.2 Congeneric (or closely-related) putative SARS immune epitopes

Linear immune epitopes identified in SARS-CoV cataloged in the Immune Epitope Database and Analysis Resource (IEDB) [12] that retain 100% sequence identity in SARS-CoV-2 are

displayed in a single track, which can be set to either “Density” mode to view the abundance of epitopes over large portions of the genome (purple track below):



Or can be set to “Full” mode to visualize individual epitopes, whose sequences are displayed upon selection.



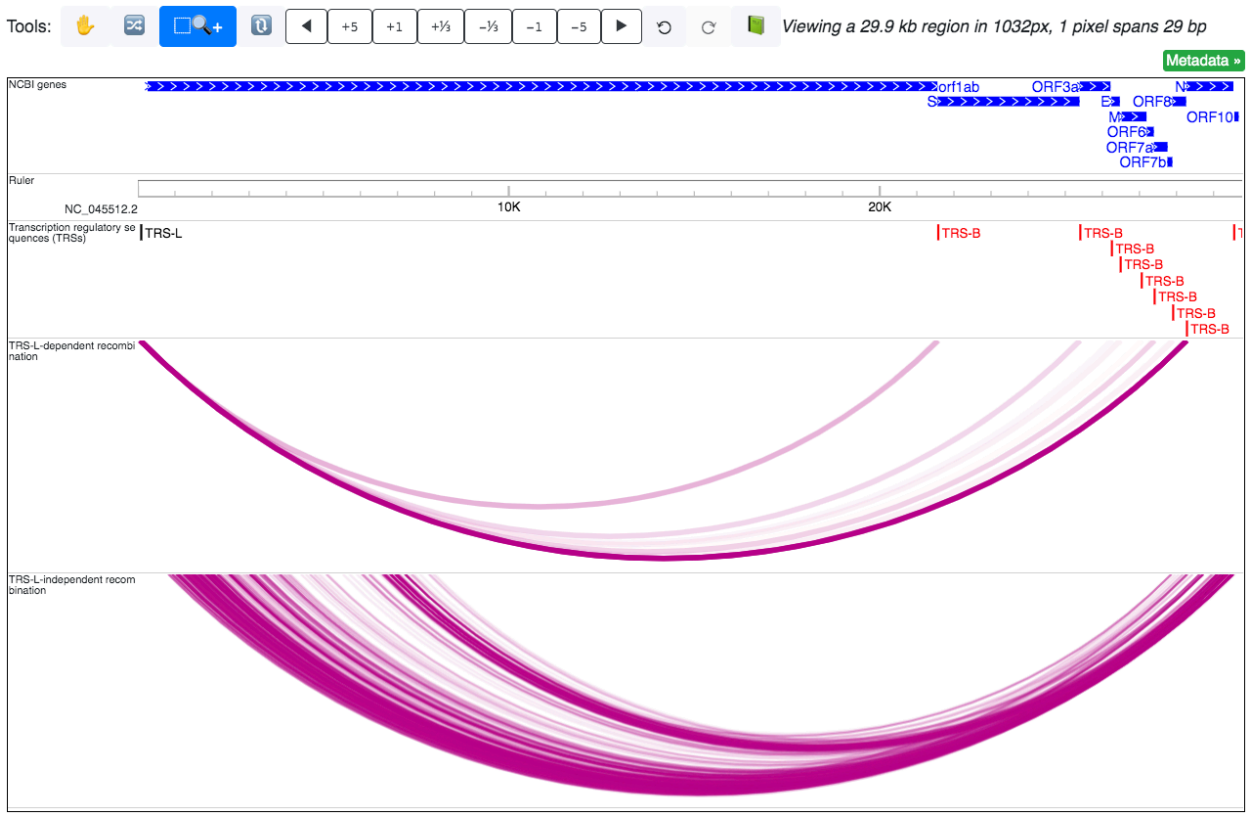
D.3.2.5.3 Putative SARS-CoV-2 epitopes

This data hub hosts several (14) different tracks, pertaining to an assortment of different studies (see Appendix D.5), and includes tracks displaying CD8 epitopes restricted to HLA-A*02:01 [13], B cell immune epitope predictions [14], CD4 T-cell immune epitope predictions [14], CD8 T-cell immune epitope predictions [14], putative epitopes for CD8+ T cells with widespread HLA binding properties [15], and N-terminal SARS-CoV-2 putative MHC-II epitopes [16].

D.3.2.6 Recombination events

Recombination events in the SARS-CoV-2 transcriptome were detected by junction-spanning RNA-seq reads generated by Kim, et al., 2020 [17], and comprise three tracks collectively making up the Recombination events data hub. Of the three included tracks, two are longrange interaction tracks, displaying transcription regulatory sequences (TRS)-L-dependent recombination events and TRS-L-independent recombination events, respectively. Locations of

predicted recombination sites are also available as an additional track. The TRS-L-dependent recombination events track and predicted recombination sites track are loaded in the default SARS-CoV-2 browser view. All three are shown below.

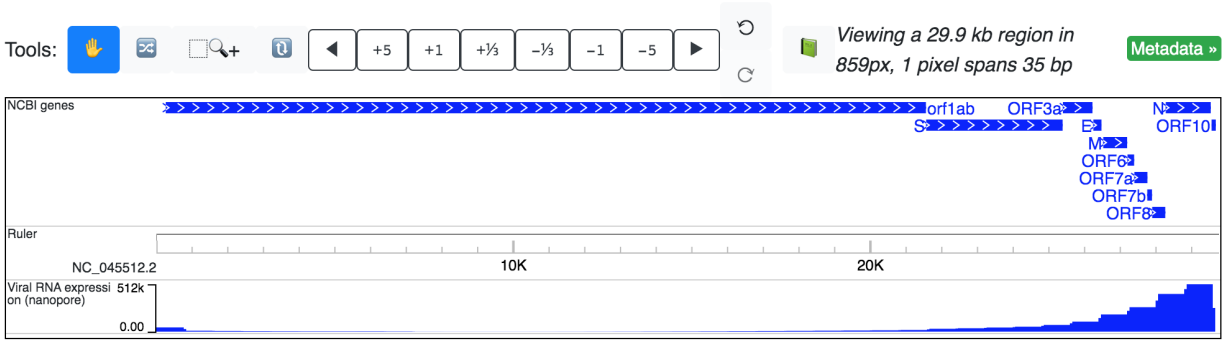


D.3.2.7 Viral RNA modifications

RNA modifications detected using Nanopore direct sequencing are reported in Kim et al., 2020 [17], and comprise the 11 tracks available in the Viral RNA modifications data hub. Modifications include: gRNA, S, 3a, E, M, 6, 7a, 7b, 8, and N. Each modification has a static track that can be loaded individually. In addition, a dynamic track is available (and also loaded in the default SARS-CoV-2 browser view) which rotates through displaying the modification signal across the genome for each modification.

D.3.2.8 Viral RNA expression

Nanopore expression data was collected for SARS-CoV-2-infected Vero cells, and reported in Kim et al., 2020 [17]. In our data hub "Viral RNA expression", we have added a bigwig file which displays the raw nanopore read counts at each genomic position. This track, shown below, is also one of the tracks displayed by default for SARS-CoV-2.



D.3.2.9 Sequence variation

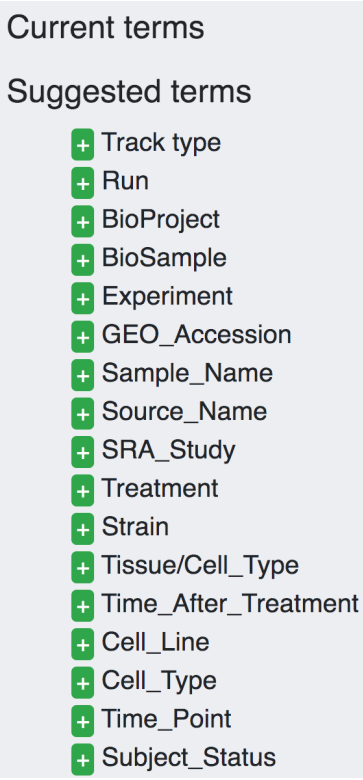
The Sequence variation collection houses one data hub: "D614G prevalence across time". This data hub contains a single demo track, which displays the percentage of SARS-CoV-2 strains per week (from 12/24/2019 to 08/06/2020) with the D164G mutation. The data within this track was retrieved from GISAID [1].

D.3.2.10 SARS-CoV-2 host transcriptional responses database

In addition to viral genomics pairwise alignments hosted on the browser, the WashU Virus Browser offers a unique view of host transcriptional responses to SARS-CoV-2 infection through partnership with the WashU Epigenome Browser. When navigating to the WashU Virus Genome Browser landing page, the user can opt to view data hubs containing host responses by selecting the link "Host transcriptional responses to SARS-CoV-2" under the "Featured Datahubs" drop-

down menu [Figure 6.6]. Selecting this link redirects the user to the hg38 genome hosted in the WashU Epigenome Browser, as shown below.

As demonstrated above, 12 RNA-seq tracks are pre-loaded into view from the pre-print Blanco-Melo, et al., 2020 [6]. However, the user can choose to look at additional tracks available within the data hub by selecting “Tracks” > “Public Data Hubs” > “SARS-CoV-2 Host Transcriptional Responses (Blanco-Melo, et al. 2020) Data Hub”. The data hub houses 195 RNA-seq tracks which can be either directly loaded into view, or can be pre-filtered based on several metadata terms. Once desired tracks are loaded into view, associated metadata can be displayed and includes the options shown below.



D.4 Extended Methods

D.4.1 Reference sequences and additional strains

The genomic sequences of all viral strains of SARS-CoV, MERS-CoV and Ebola were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/nucleotide/) [8]. For SARS-CoV-2, the NCBI database was downloaded from https://www.ncbi.nlm.nih.gov/nucleotide/, the GISAID database [7] was from GISAID.org and the Nextstrain database [1] was parsed from http://data.nextstrain.org/ncov_global.json. Pairwise comparisons between the reference and each available sequence as of August 13, 2020 for SARS-CoV-2, MERS-CoV, SARS-CoV, and Ebola are available (n=>80,000, 551, 332, and 1574, respectively) [Appendix D.5]. The database is also updated at least once a week.

The reference genomic sequence of the selected virus (SARS-CoV-2: NC_045512.2; MERS-CoV: NC_019843.3; SARS-CoV: NC_004718.3, Ebola: KM034562.1) is automatically displayed as a color-coded track when opening the “Browser View” through home page.

D.4.2 Sequence alignment and tree generation

With the exception of the Nextstrain database for SARS-CoV-2, where the alignment was parsed from http://data.nextstrain.org/ncov_global.json, all genome alignments were performed using the pairwise alignment tool stretcher [9] with parameters “-gapopen 16 -gapextend 4”. The Phylogenetic tree for SARS-CoV-2 was parsed from http://data.nextstrain.org/ncov_global.json. To generate the phylogenetic trees for the other 3 viruses, we used the MAFFT program, employing the fast option to align individual strains of each viral genome to its reference [18,

19]. Approximately-maximum-likelihood phylogenetic trees were built using FastTree with the GTR substitution model [20, 21].

D.4.3 Comparison of SARS-CoV-2 to bat RaTG13 and Pangolin-CoV using next-generation sequencing reads

We downloaded all pangolin short reads from a metagenomics dataset [22] and obtained the bat reads from the Rhinolophus affinis fecal swab RNA-seq sample (SRR11085797) [23]. We classified all reads using Kraken2 and extracted reads classified as coronavirus (NCBI taxID: 2499399) [24]. We then aligned the coronavirus reads to the SARS-CoV-2 reference genome using BWA-MEM to generate bam files (one for pangolin and one for bat) [25, 26]. Finally, we sorted and indexed the alignment bam files and displayed them on the WashU Virus Genome Browser.

D.4.4 Annotation data tracks

D.4.4.1 Gene annotations

Genic annotations of reference genomes were downloaded as GFF3 files from NCBI and converted to refBed format for viewing on the browser. This track is automatically displayed when opening the browser, and can be re-added to view if removed by selecting: “Tracks” > “Annotation Tracks” > “SARS-CoV-2” > “Genes”.

D.4.4.2 Genome comparison track

We adopted the genome comparison track from the WashU Epigenome Browser [2]. Any pairwise alignment result in markx3 or FASTA format can be converted using our publicly accessible script “aligned_fa_2_genomealign.py” (available at https://github.com/debugpoint136/WashU-Virus-Genome-Browser) and displayed as a genome comparison track on the WashU Virus Genome Browser. The browser currently contains 4 such comparison tracks: MERS to SARS-CoV-2, SARS to SARS-CoV-2, pangolin CoV to SARS-CoV-2, and bat CoV to SARS-CoV-2 alignments, which can be accessed by selecting: “Tracks” > “Annotation Tracks” > “SARS-CoV-2” > “Genome Comparison”.

D.4.4.3 GC density track

GC density tracks were created for each reference genome, displaying the percentage of G (guanine) and C (cytosine) bases in 5-bp windows. This track can be accessed by selecting: “Tracks” > “Annotation Tracks” > “SARS-CoV-2” > “Assembly”.

D.4.4.4 Sequence diversity track and mutation alert track

We calculated the percentage of each of the 4 nucleotides at each position in the genome across all strains for a given virus species (SARS-CoV-2 strains downloaded from GISAID [7]). We then calculated Shannon entropy for each position along the genome using the percentages of each of the 4 nucleotides, and generated the sequence diversity track (bedgraph format). A high Shannon entropy at a position signifies that the 4 possible nucleotides are equally likely across all strains of this virus, and thus the position is likely divergent. A low Shannon entropy at a position means that the identity of the nucleotide at this position is highly conserved across all strains. The entropy() function of the R package “entropy” was used for calculations.

The mutation alert track was generated in a similar fashion. We used the qBED track format [27] to display the number of available strains containing a mutation for each nucleotide along the reference genome, along with the accession numbers of these strains (upon nucleotide selection).

Both the sequence diversity track and the mutation alert track are loaded by default and can be re-added if removed by selecting: “Tracks” > “Annotation Tracks” > “SARS-CoV-2” > “Diversity”.

D.4.5 Specialty tracks and public data hub tracks

D.4.5.1 SNV/SNV2 track

We developed the SNV track type to display sequence variation of individual strains relative to their reference. Variations from the reference genome, including mismatches, insertions, and deletions, are displayed with different colors. Details of variation will be shown upon selection. When viewing large regions, such as the whole genome, it is not possible to display all individual variation events. Therefore, the frequency of variation events is also displayable in a “density mode”, where a high value over a region signifies multiple sequence variation events within the region. Each housed strain can be directly loaded as a SNV track, which uses the “pairwise” format. Any user-defined alignment result can be converted to this format through the scripts we offer, as detailed in Appendix D.2. We have additionally developed the “SNV2” track type, which is specified with high flexibility to indicate potential amino acid mutations in addition to DNA mutations. Full details regarding the SNV2 track are also located in Appendix D.2.

D.4.5.2 Congeneric (or closely-related) immune epitope locations

We wrote a text processing utility to import antibody-binding epitopes curated by the Immune Epitope Database and Analysis Resource (IEDB) for MERS-CoV and SARS-CoV [12]. Subsequently, we used tblastn [28] to align linear epitopes to the SARS-CoV-2 reference genome (Taxonomy ID: 2697049; NCBI:txid2697049). We found 955 out of 2,817 linear epitopes identified in SARS-CoV had at least 1 “hit” in the SARS-CoV-2 genome, 320 of which had amino acid sequences identical to predicted translated products from the SARS-CoV-2 reference [Appendix D.2]. Three epitopes had 2 “hits” each. However, the secondary hit is on the negative strand with very low percent identity (37.5% to 53.8%) to the SARS-CoV-2 genome and were hence filtered out as SARS-CoV-2 is a (+) ssRNA virus. Similarly, we found 1 hit out of 38 linear epitopes identified in MERS-CoV. We also provide scripts (available at https://github.com/debugpoint136/WashU-Virus-Genome-Browser) that can be used to obtain a quick overview of the similarity of linear epitopes identified in other viruses in databases such as IEDB [12]. These tracks can provide researchers preliminary data to support exploratory analyses pertaining to the immunogenicity of SARS-CoV-2—an actively explored vertical of SARS-CoV-2 research.

D.4.5.3 Putative SARS-CoV-2 epitopes

We annotate various immune epitopes across the SARS-CoV-2 genome by displaying them as bars along the genome browser. The x-axis of these “BED” tracks display the epitopes’ DNA sequence location and upon clicking the bar one can see the actual peptide sequence. Along with the BED files, we also generated qBED tracks wherein the y-axis displays the score associated with the immune epitopes.

For Nerli et al. [13], the score is calculated as (0.5 * (((NetMHCPan affinity - Average NetMHCPan affinity) / range of NetMHCPan affinities) + ((Rosetta binding energy - Average Rosetta binding energy) / range of Rosetta binding energies)) + 1) * 500).

For Grifoni et al. [14], the score reflects the epitope score calculated using software packages: B cells = BebiPred 2.0, CD4 = IEDB Tepitool, CD8 = NetMHCpan4.0EL.

For Poran et al. [15], the score is calculated as (USA_coverage+EUR_coverage+API_coverage) *1000/3.

D.4.5.4 Recombination events and transcription regulatory sequences (TRSs)
Recombination events within the SARS-CoV-2 genome were annotated by [17]. Counts of RNA-seq reads spanning junctions between non-contiguous regions of the SARS-CoV-2 genome were used to generate longrange interaction tracks for TRS-L-dependent and TRS-L-independent recombination events. Coordinates of predicted recombination sites (also called transcription regulatory sequences; TRSs) were retrieved from WU, et al. [29].

D.4.5.5 Predicted SARS-CoV-2 epitope binding sites to class 1 HLA proteins track in qbed format
Predicted SARS-CoV-2 peptide sequences were annotated in Campbell, et al. [11], highlighting regions where viral protein is predicted to bind class 1 HLA proteins. Predicted translated sequences were compared with the reference SARS-CoV-2 genome (NC_044512.2, using tblastn [28]), and regions of the SARS-CoV-2 sequence with a perfect match with a predicted epitope sequence were identified (excluding one perfect match on the reverse strand due to low likelihood of translation). The x-axis of the track displays the epitope DNA sequence location, and the y-axis displays the number of unique strain IDs in which the epitope was reported in Campbell et al. [11].

D.4.5.6 Primer tracks
All primer tracks were generated according to the document: https://www.who.int/docs/default-source/coronaviruse/whoinhouseassays.pdf?sfvrsn=de3a76aa_2, in BED format.

D.4.5.7 Tracks for CRISPR-based detection assays
Tracks were generated for two CRISPR-based assays that have been used for detecting SARS-CoV-2: the SHERLOCK assay [https://www.broadinstitute.org/files/publications/special/COVID-19%20detection%20(updated).pdf; accessed on 05-08-2020] and the DETECTR assay [13]. The genomic coordinates of the primer sequences and guide RNA sequences associated with each assay were identified by aligning the sequences to the SARS-CoV-2 genome using BLAST [30].

D.4.5.8 Resources to convert user-defined tracks for browser display
In addition to our housed data tracks, we also offer a script ("publicConvertMarkx3.py") to convert any markx3 or FASTA-formatted alignment into "pairwise" format directly displayable on the browser, and a script ("publicJsonGen.py") to generate a json file for uploading multiple data files together for display (available at https://github.com/debugpoint136/WashU-Virus-Genome-Browser). We also offer a script ("publicAlignment.py") to perform alignment in batch

with EMBL “stretcher” aligner [9], and generate files in “pairwise” format directly displayable on the browser. Finally, to display protein level mutations, we also offer a set of scripts (https://github.com/debugpoint136/WashU-Virus-Genome-Browser/tree/master/scripts/snv2/) that can be used to generate “SNV2” tracks.

D.4.5.9 Long read transcriptome track

Nanopore reads derived from SARS-CoV-2 were obtained from Kim et al. [17]. We aligned nanopore reads to SARS-CoV-2 following the method from Kim et al. At last, we converted the long read BAM file to a BigWig file using deeptools with default parameters [31].

D.4.5.10 RNA modifications track

RNA modification sites within SARS-CoV-2 transcripts were annotated by [17]. The fraction of modified RNA-seq reads at each modification site within each transcript was retrieved and used to generate a dynamic bedgraph track, which is displayed in the genome browser view for SARS-CoV-2 by default.

D.4.5.11 D614G mutation track

Metadata was retrieved from all SARS-CoV-2 isolates in the GISAID database [1] (accessed on 08/19/2020). Isolates were annotated based on whether they contained nucleotide changes that correspond to a D614G amino acid mutation within the S protein. Collection date metadata of isolates with and without the mutation were used to determine the fraction of isolates per collection week that contained the mutation.

D.5 Public data hubs available on the WashU Virus Genome Browser as of 08/13/2020

Collection Name	Hub Name	Number of Tracks	Description	Data Source / Reference
NCBI database	All NCBI SARS-CoV-2 isolates	15215	SNV tracks of SARS-CoV-2 strains in NCBI Genbank displaying their sequence variation from reference	https://www.ncbi.nlm.nih.gov/nuccore, [8]
	All NCBI SARS-CoV-2 isolates, in SNV2 format	15215	SNV2 tracks of SARS-CoV-2 strains in NCBI Genbank displaying their sequence variation from reference	
Nextstrain database	All Nextstrain SARS-CoV-2 isolates	4263	SNV tracks of SARS-CoV-2 strains in Nextstrain, displaying their sequence variation from reference	http://data.nextstrain.org/ncov_global.json, [1]
	All Nextstrain SARS-CoV-2 isolates, in SNV2 format	4263	SNV2 tracks of SARS-CoV-2 strains in Nextstrain, displaying their sequence variation from reference	
GISAID database	GISAID database -5/22/2020	30612	SNV tracks of SARS-CoV-2 strains in GISAID, displaying their sequence variation from reference	GISAID.org, [7]
	GISAID database 5/22/2020-	50238	SNV tracks of SARS-CoV-2 strains in GISAID, displaying their sequence variation from reference	
	GISAID database SNV2 format (-5/22/2020)	30612	SNV2 tracks of SARS-CoV-2 strains in GISAID, displaying their sequence variation from reference	
	GISAID database SNV2 format (5/22/2020-)	50238	SNV2 tracks of SARS-CoV-2 strains in GISAID, displaying their sequence variation from reference	
Diagnostics	Primers	9	CDC and WHO primers for detecting SARS-CoV-2	https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/laboratory-guidance
	CRISPR-based diagnostic tests	2	Primers, gRNAs, etc. for various diagnostic tests	https://www.broadinstitute.org/files/publications/special/COVID-19%20detection%20(updated).pdf, [10]
Putative SARS-CoV-2 Immune Epitopes	Putative SARS-CoV-2 Epitopes	14	Collection of predicted and putative immune epitopes identified in SARS-CoV and SARS-CoV-2 across various studies using a variety of different computational and experimental techniques. Please refer to individual manuscripts and methods for further details.	[11, 13-16, 32]
	SARS-CoV-2 Epitopes Predicted to Bind HLA Class 1 Proteins	1	Collection of SARS-CoV epitope predictions with 100% sequence identity in SARS-CoV-2 that are predicted to bind to class 1 HLA proteins	[11]
	Congeneric (or Closely-related) Putative SARS Immune Epitopes	1	Linear immune epitopes identified in SARS-CoV-1 cataloged in the Immune Epitope Database and Analysis Resource (IEDB) [DOI: 10.1093/nar/gky1006] that retain 100% sequence identity in SARS-CoV-2	Immune Epitope Database and Analysis Resource (IEDB) [12]
Sequence Variation	D614G prevalence across time	1	Dynamic tracks for using the browser to visualize and study sequence variation and diversity across strains over time	[32]
Recombination events	Recombiination events	3	Recombination events detected by junction-spanning RNA-seq	[17, 29]
Viral RNA modifications	Viral RNA modifications	10	RNA modifications detected using Nanopre direct RNA sequence	[17]
Viral RNA expression	Viral RNA expression	1	Viral RNA expression measured by Nanopore	[17]
SARS-CoV-2 Host Transcriptional Responses Database	SARS-CoV-2 Host Transcriptional Responses	195	Human transcriptional responses to SARS-CoV-2 and other respiratory infections	[6]

Please note that data hubs are continuously being added.

D.5 High-confidence linear immune epitopes identical to predicted SARS-CoV-2 translated peptides

Linear antibody-binding epitopes curated by the Immune Epitope Database and Analysis Resource (IEDB) identified in SARS-CoV with amino acid sequences identical to predicted translated products from the SARS-CoV-2 reference.

Chr	Start	Stop	EpitopeID	Identity	Nucleotide length (bp)	AA length	Epitope sequence	-log(eValue)	Link
NC_045512.2	25066	25192	28513	1	126	42	ISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYI	54.2112	https://www.iedb.org/epitope/28513
NC_045512.2	24997	25093	10113	1	96	32	DSFKEELDKYFKNHTSPDVDLGDISGINASVV	42.4599	https://www.iedb.org/epitope/10113
NC_045512.2	25105	25189	558417	1	84	28	EIDRLNEVAKNLNESLIDLQELGKYEQY	34.9348	https://www.iedb.org/epitope/558417
NC_045512.2	25105	25189	12426	1	84	28	EIDRLNEVAKNLNESLIDLQELGKYEQY + ACET(E1)	33.4302	https://www.iedb.org/epitope/12426
NC_045512.2	28750	28816	51485	1	66	22	QLPQGTTLPKGFYAEGSRGGSQ	23.674	https://www.iedb.org/epitope/51485
NC_045512.2	25321	25381	6476	1	60	20	CKFDEDDSEPVLKGVKLHYT	22.0665	https://www.iedb.org/epitope/6476
NC_045512.2	29080	29140	71468	1	60	20	VTQAFGRRGPEQTQGNFGDQ	21.5196	https://www.iedb.org/epitope/71468
NC_045512.2	29074	29131	75235	1	57	19	YNVTQAFGRRGPEQTQGNF	20.8088	https://www.iedb.org/epitope/75235
NC_045512.2	29209	29263	1295	1	54	18	AFFGMSRIGMEVTPSGTW	20.1299	https://www.iedb.org/epitope/1295
NC_045512.2	24988	25039	47341	1	51	17	PELDSFKEELDKYFKNH	20.1135	https://www.iedb.org/epitope/47341
NC_045512.2	26990	27044	21996	1	54	18	GRCDIKDLPKEITVATSR	20.0503	https://www.iedb.org/epitope/21996
NC_045512.2	24658	24709	69513	1	51	17	VLGQSKRVDFCGKGYHL	19.9671	https://www.iedb.org/epitope/69513
NC_045512.2	28402	28456	21065	1	54	18	GLPNNTASWFTALTQHGK	19.6795	https://www.iedb.org/epitope/21065
NC_045512.2	25009	25060	11740	1	51	17	EELDKYFKNHTSPDVDL	19.6655	https://www.iedb.org/epitope/11740
NC_045512.2	25066	25126	28512	1	60	20	ISGINASVVNIQKEIDRLNE	19.5919	https://www.iedb.org/epitope/28512
NC_045512.2	29146	29197	28371	1	51	17	IRQGTDYKHWPQIAQFA	19.4204	https://www.iedb.org/epitope/28371
NC_045512.2	29212	29263	15814	1	51	17	FFGMSRIGMEVTPSGTW	19.2531	https://www.iedb.org/epitope/15814
NC_045512.2	24967	25018	67220	1	51	17	TVYDPLQPELDSFKEEL	18.9689	https://www.iedb.org/epitope/67220
NC_045512.2	25018	25066	9007	1	48	16	DKYFKNHTSPDVDLGD	18.9348	https://www.iedb.org/epitope/9007
NC_045512.2	16731	16776	62435	1	45	15	SWEVGKPRPPLNRNY	18.7164	https://www.iedb.org/epitope/62435
NC_045512.2	29326	29380	37611	1	54	18	LLNKHIDAYKTFPPTEPK	18.6899	https://www.iedb.org/epitope/37611
NC_045512.2	10513	10558	73526	1	45	15	YDCVSFCYMHHMELP	18.6388	https://www.iedb.org/epitope/73526
NC_045512.2	25030	25081	32508	1	51	17	KNHTSPDVDLGDISGIN	18.6364	https://www.iedb.org/epitope/32508
NC_045512.2	28750	28804	51484	1	54	18	QLPQGTTLPKGFYAEGSR	18.3624	https://www.iedb.org/epitope/51484
NC_045512.2	24265	24316	41177	1	51	17	MAYRFNGIGVTQNVLYE	18.0986	https://www.iedb.org/epitope/41177
NC_045512.2	24250	24301	47479	1	51	17	PFAMQMAYRFNGIGVTQ	18.056	https://www.iedb.org/epitope/47479
NC_045512.2	19464	19509	53297	1	45	15	RCNLGGAVCRHHANE	18.002	https://www.iedb.org/epitope/53297
NC_045512.2	28511	28568	153445	1	57	19	QMTKLATTEELPDEFVVVTA	17.926	https://www.iedb.org/epitope/153445
NC_045512.2	28384	28435	48131	1	51	17	PKQRRPQGLPNNTASWFT	17.8164	https://www.iedb.org/epitope/48131
NC_045512.2	29167	29218	31166	1	51	17	KHWPQIAQFAPSASAFF	17.6787	https://www.iedb.org/epitope/31166
NC_045512.2	24097	24148	4129	1	51	17	ARDLICAQKFNGLTVLP	17.5124	https://www.iedb.org/epitope/4129
NC_045512.2	22540	22591	70719	1	51	17	VRFPNITNLCPFGEVFN	17.4237	https://www.iedb.org/epitope/70719
NC_045512.2	24010	24061	16183	1	51	17	FIEDLLFNKVTLADAGF	17.2087	https://www.iedb.org/epitope/16183
NC_045512.2	24232	24283	18515	1	51	17	GAALQIPFAMQMAYRFN	17.1969	https://www.iedb.org/epitope/18515
NC_045512.2	29185	29239	3958	1	54	18	AQFAPSASAFFGMSRIGM	17.1736	https://www.iedb.org/epitope/3958
NC_045512.2	25018	25063	9006	1	45	15	DKYFKNHTSPDVDLG	17.0146	https://www.iedb.org/epitope/9006
NC_045512.2	11713	11758	65836	1	45	15	TQEFRYMNSQGLLPP	16.9856	https://www.iedb.org/epitope/65836
NC_045512.2	20673	20718	51846	1	45	15	QPGVAMPNLYKMQRM	16.7605	https://www.iedb.org/epitope/51846
NC_045512.2	29335	29395	31116	1	60	20	KHIDAYKTFPPTEPKKDKKK	16.6908	https://www.iedb.org/epitope/31116
NC_045512.2	20691	20736	48674	1	45	15	PNLYKMQRMLLEKCD	16.6628	https://www.iedb.org/epitope/48674
NC_045512.2	17223	17268	7072	1	45	15	CSRIIPARARVECFD	16.6339	https://www.iedb.org/epitope/7072
NC_045512.2	18147	18192	20789	1	45	15	GLCVDIPGIPKDMTY	16.6239	https://www.iedb.org/epitope/20789
NC_045512.2	29329	29380	38249	1	51	17	LNKHIDAYKTFPPTEPK	16.6108	https://www.iedb.org/epitope/38249
NC_045512.2	24634	24685	462	1	51	17	AATKMSECVLGQSKRVD	16.5707	https://www.iedb.org/epitope/462
NC_045512.2	18486	18531	27933	1	45	15	IPLMYKGLPWNVVRI	16.5644	https://www.iedb.org/epitope/27933
NC_045512.2	10666	10711	35022	1	45	15	LAWLYAAVINGDRWF	16.5351	https://www.iedb.org/epitope/35022
NC_045512.2	28774	28828	48067	1	54	18	PKGFYAEGSRGGSQASSR	16.4287	https://www.iedb.org/epitope/48067
NC_045512.2	25015	25060	35205	1	45	15	LDKYFKNHTSPDVDL	16.4273	https://www.iedb.org/epitope/35205
NC_045512.2	29323	29368	27183	1	45	15	ILLNKHIDAYKTFPP	16.4084	https://www.iedb.org/epitope/27183
NC_045512.2	13470	13515	67847	1	45	15	VCGVSAARLTPCGTG	16.2712	https://www.iedb.org/epitope/67847
NC_045512.2	20766	20811	48069	1	45	15	PKGIMMNVAKYTQLC	16.262	https://www.iedb.org/epitope/48069
NC_045512.2	24265	24313	100300	1	48	16	MAYRFNGIGVTQNVLY	16.2551	https://www.iedb.org/epitope/100300
NC_045512.2	24616	24667	53202	1	51	17	RASANLAATKMSECVLG	16.18	https://www.iedb.org/epitope/53202
NC_045512.2	25102	25153	30435	1	51	17	KEIDRLNEVAKNLNESL	16.18	https://www.iedb.org/epitope/30435
NC_045512.2	4378	4423	62482	1	45	15	SWNLREMLAHAEETR	16.1444	https://www.iedb.org/epitope/62482
NC_045512.2	18621	18666	41897	1	45	15	MKYFVKIGPERTCCL	16.1342	https://www.iedb.org/epitope/41897
NC_045512.2	10663	10708	69393	1	45	15	VLAWLYAAVINGDRW	16.1181	https://www.iedb.org/epitope/69393
NC_045512.2	19023	19068	36365	1	45	15	LHDIGNPKAIKCVPQ	16.1081	https://www.iedb.org/epitope/36365
NC_045512.2	15189	15234	18805	1	45	15	GATVVIGTSKFYGGW	16.0789	https://www.iedb.org/epitope/18805
NC_045512.2	14472	14517	23353	1	45	15	GYHFRELGVVHNQDV	16.0504	https://www.iedb.org/epitope/23353
NC_045512.2	19521	19566	35122	1	45	15	LDAYNMMISAGFSLW	16.0228	https://www.iedb.org/epitope/35122
NC_045512.2	17418	17463	25210	1	45	15	HYVYIGDPAQLPAPR	16.0137	https://www.iedb.org/epitope/25210
NC_045512.2	24214	24259	23293	1	45	15	GWTFGAGAALQIPFA	15.9697	https://www.iedb.org/epitope/23293
NC_045512.2	17646	17691	6234	1	45	15	CFKMFYKGVITHDVS	15.9611	https://www.iedb.org/epitope/6234

Chr	Start	Stop	EpitopeID	Identity	Nucleotide length (bp)	AA length	Epitope sequence	-log(eValue)	Link
NC_045512.2	18174	18219	48042	1	45	15	PKDMTYRRLISMMGF	15.887	https://www.iedb.org/epitope/48042
NC_045512.2	19845	19890	22936	1	45	15	GVDIAANTVIWDYKR	15.887	https://www.iedb.org/epitope/22936
NC_045512.2	28585	28630	60243	1	45	15	SPRWYFYYLGTGPEA	15.887	https://www.iedb.org/epitope/60243
NC_045512.2	18573	18618	10063	1	45	15	DRVVFVLWAHGFELT	15.8791	https://www.iedb.org/epitope/10063
NC_045512.2	15864	15909	47811	1	45	15	PHEFCSQHTMLVKQG	15.8635	https://www.iedb.org/epitope/47811
NC_045512.2	25123	25174	14626	1	51	17	EVAKNLNESLIDLQELG	15.8635	https://www.iedb.org/epitope/14626
NC_045512.2	7492	7537	42108	1	45	15	MMCYKRNRATRVECT	15.8635	https://www.iedb.org/epitope/42108
NC_045512.2	20880	20925	8878	1	45	15	DKGVAPGTAVLRQWL	15.8405	https://www.iedb.org/epitope/8878
NC_045512.2	27014	27065	48052	1	51	17	PKEITVATSRTLSYYKL	15.8254	https://www.iedb.org/epitope/48052
NC_045512.2	13758	13803	8362	1	45	15	DGDMVPHISRQRLTK	15.7604	https://www.iedb.org/epitope/8362
NC_045512.2	11914	11959	61140	1	45	15	SSKLWAQCVQLHNDI	15.6928	https://www.iedb.org/epitope/61140
NC_045512.2	18180	18225	9413	1	45	15	DMTYRRLISMMGFKM	15.6928	https://www.iedb.org/epitope/9413
NC_045512.2	14388	14433	17868	1	45	15	FSTVFPPTSFGPLVR	15.6544	https://www.iedb.org/epitope/17868
NC_045512.2	24904	24955	29108	1	51	17	ITTDNTFVSGNCDVVIG	15.6234	https://www.iedb.org/epitope/29108
NC_045512.2	24232	24280	100048	1	48	16	GAALQIPFAMQMAYRF	15.6173	https://www.iedb.org/epitope/100048
NC_045512.2	10600	10645	21624	1	45	15	GPFVDRQTAQAAGTD	15.5471	https://www.iedb.org/epitope/21624
NC_045512.2	20808	20853	6904	1	45	15	CQYLNTLTLAVPYNM	15.5193	https://www.iedb.org/epitope/6904
NC_045512.2	24499	24550	27357	1	51	17	ILSRLDKVEAEVQIDRL	15.4922	https://www.iedb.org/epitope/27357
NC_045512.2	17046	17091	34023	1	45	15	KVGMQKYSTLQGPPG	15.4868	https://www.iedb.org/epitope/34023
NC_045512.2	18735	18780	15475	1	45	15	FDYVYNPFMIDVQQW	15.3434	https://www.iedb.org/epitope/15475
NC_045512.2	24523	24574	11038	1	51	17	EAEVQIDRLITGRLQSL	15.2895	https://www.iedb.org/epitope/11038
NC_045512.2	24244	24289	51112	1	45	15	QIPFAMQMAYRFNGI	15.2722	https://www.iedb.org/epitope/51112
NC_045512.2	29188	29239	50741	1	51	17	QFAPSASAFFGMSRIGM	15.2679	https://www.iedb.org/epitope/50741
NC_045512.2	10783	10828	50469	1	45	15	QDHVDILGPLSAQTG	15.2637	https://www.iedb.org/epitope/50469
NC_045512.2	16077	16122	68484	1	45	15	VFHLYLQYIRKLHDE	15.2637	https://www.iedb.org/epitope/68484
NC_045512.2	20661	20706	60394	1	45	15	SQAWQPGVAMPNLYK	15.2552	https://www.iedb.org/epitope/60394
NC_045512.2	25054	25105	9094	1	51	17	DLGDISGINASVVNIQK	15.251	https://www.iedb.org/epitope/9094
NC_045512.2	29320	29365	69035	1	45	15	VILLNKHIDAYKTFP	15.2261	https://www.iedb.org/epitope/69035
NC_045512.2	10597	10642	74066	1	45	15	YGPFVDRQTAQAAGT	15.2058	https://www.iedb.org/epitope/74066
NC_045512.2	12898	12943	47038	1	45	15	PCRFVTDTPKGPKVK	15.1248	https://www.iedb.org/epitope/47038
NC_045512.2	20805	20850	35092	1	45	15	LCQYLNTLTLAVPYN	15.0956	https://www.iedb.org/epitope/35092
NC_045512.2	10351	10396	34479	1	45	15	KYKFVRIQPGQTFSV	15.0673	https://www.iedb.org/epitope/34479
NC_045512.2	28405	28450	38599	1	45	15	LPNNTASWFTALTQH	15.0568	https://www.iedb.org/epitope/38599
NC_045512.2	5050	5095	49820	1	45	15	PTYLDGADVTKIKPH	15.0568	https://www.iedb.org/epitope/49820
NC_045512.2	28390	28435	55683	1	45	15	RRPQGLPNNTASWFT	15.0095	https://www.iedb.org/epitope/55683
NC_045512.2	17487	17532	13646	1	45	15	EPEYFNSVCRLMKTI	14.9803	https://www.iedb.org/epitope/13646
NC_045512.2	14715	14760	30926	1	45	15	KGFFKEGSSVELKHF	14.9707	https://www.iedb.org/epitope/30926
NC_045512.2	25336	25381	7868	1	45	15	DDSEPVLKGVKLHYT	14.9394	https://www.iedb.org/epitope/7868
NC_045512.2	18510	18555	50044	1	45	15	PWNVVRIKIVQMLSD	14.8826	https://www.iedb.org/epitope/50044
NC_045512.2	24544	24595	54599	1	51	17	RLITGRLQSLQTYVTQQ	14.8768	https://www.iedb.org/epitope/54599
NC_045512.2	16020	16065	59002	1	45	15	SLAIDAYPLTKHPNQ	14.8234	https://www.iedb.org/epitope/59002
NC_045512.2	20697	20742	40820	1	45	15	LYKMQRMLLEKCDLQ	14.8234	https://www.iedb.org/epitope/40820
NC_045512.2	16191	16236	47282	1	45	15	PEFYEAMYTPHTVLQ	14.7752	https://www.iedb.org/epitope/47282
NC_045512.2	11209	11254	73872	1	45	15	YFNMVYMPASWVMRI	14.7444	https://www.iedb.org/epitope/73872
NC_045512.2	28588	28630	49278	1	42	14	PRWYFYYLGTGPEAS	14.7243	https://www.iedb.org/epitope/49278
NC_045512.2	17649	17694	16409	1	45	15	FKMFYKGVITHDVSS	14.7047	https://www.iedb.org/epitope/16409
NC_045512.2	13288	13333	15320	1	45	15	FCDLKGKYVQIPTTC	14.6572	https://www.iedb.org/epitope/15320
NC_045512.2	10360	10405	18273	1	45	15	FVRIQPGQTFSVLAC	14.6365	https://www.iedb.org/epitope/18273
NC_045512.2	29230	29272	26273	1	42	14	IGMEVTPSGTWLTYH	14.6274	https://www.iedb.org/epitope/26273
NC_045512.2	13941	13986	45314	1	45	15	NPDILRVYANLGERV	14.6185	https://www.iedb.org/epitope/45314
NC_045512.2	21387	21432	35815	1	45	15	LFDMSKFPLKLRGTA	14.5986	https://www.iedb.org/epitope/35815
NC_045512.2	29221	29266	42648	1	45	15	MSRIGMEVTPSGTWL	14.5877	https://www.iedb.org/epitope/42648
NC_045512.2	15972	16017	7804	1	45	15	DDIVKTDGTLMIERF	14.5684	https://www.iedb.org/epitope/7804
NC_045512.2	24592	24646	100428	1	54	18	QLIRAAEIRASANLAATK	14.5663	https://www.iedb.org/epitope/100428
NC_045512.2	28748	28793	75144	1	45	15	YNFLKEQHCQKASTQ	14.5207	https://www.iedb.org/epitope/75144
NC_045512.2	19431	19476	76386	1	45	15	YVPLKSATCITRCNL	14.4908	https://www.iedb.org/epitope/76386
NC_045512.2	7933	7978	4832	1	45	15	ASVYYSQLMCQPILL	14.4504	https://www.iedb.org/epitope/4832
NC_045512.2	16203	16248	11142	1	45	15	EAMYTPHTVLQAVGA	14.441	https://www.iedb.org/epitope/11142
NC_045512.2	19287	19332	22417	1	45	15	GSLYVNKHAFHTPAF	14.4317	https://www.iedb.org/epitope/22417
NC_045512.2	20910	20955	39176	1	45	15	LRQWLPTGTLLVDSD	14.4225	https://www.iedb.org/epitope/39176
NC_045512.2	14892	14937	19692	1	45	15	GGCINANQVIVNNLD	14.4115	https://www.iedb.org/epitope/19692
NC_045512.2	13300	13345	30995	1	45	15	KGKYVQIPTTCANDP	14.3637	https://www.iedb.org/epitope/30995
NC_045512.2	24568	24619	59425	1	51	17	SLQTYVTQQLIRAAEIR	14.3482	https://www.iedb.org/epitope/59425
NC_045512.2	24430	24481	50311	1	51	17	QALNTLVKQLSSNFGAI	14.3381	https://www.iedb.org/epitope/50311
NC_045512.2	14550	14595	37937	1	45	15	LLVYAADPAMHAASG	14.2952	https://www.iedb.org/epitope/37937
NC_045512.2	24268	24313	5908	1	45	15	AYRFNGIGVTQNVLY	14.2952	https://www.iedb.org/epitope/5908
NC_045512.2	7834	7879	35243	1	45	15	LDNLRANNTKGSLPI	14.2952	https://www.iedb.org/epitope/35243
NC_045512.2	20982	21027	61722	1	45	15	STLIGDCATVHTANK	14.2855	https://www.iedb.org/epitope/61722
NC_045512.2	12424	12469	8358	1	45	15	DGCVPLNIIPLTTAA	14.2649	https://www.iedb.org/epitope/8358
NC_045512.2	13303	13348	20739	1	45	15	GKYVQIPTTCANDPV	14.2556	https://www.iedb.org/epitope/20739
NC_045512.2	11212	11257	17203	1	45	15	FNMVYMPASWVMRIM	14.2265	https://www.iedb.org/epitope/17203
NC_045512.2	14430	14475	54361	1	45	15	RKIFVDGVPFVVSTG	14.216	https://www.iedb.org/epitope/54361
NC_045512.2	11218	11263	43010	1	45	15	MVYMPASWVMRIMTW	14.1968	https://www.iedb.org/epitope/43010
NC_045512.2	10789	10834	25027	1	45	15	HVDILGPLSAQTGIA	14.1866	https://www.iedb.org/epitope/25027
NC_045512.2	20901	20946	63033	1	45	15	TAVLRQWLPTGTLLV	14.1779	https://www.iedb.org/epitope/63033
NC_045512.2	25210	25255	72717	1	45	15	WLGFIAGLIAIVMVT	14.1482	https://www.iedb.org/epitope/72717
NC_045512.2	24451	24502	33032	1	51	17	KQLSSNFGAISSVLNDI	14.1426	https://www.iedb.org/epitope/33032
NC_045512.2	17457	17502	49262	1	45	15	PRTLLTKGTLEPEYF	14.1289	https://www.iedb.org/epitope/49262
NC_045512.2	20412	20457	35381	1	45	15	LEDFIPMDSTVKNYF	14.118	https://www.iedb.org/epitope/35381

Chr	Start	Stop	EpitopeID	Identity	Nucleotide length (bp)	AA length	Epitope sequence	-log(eValue)	Link
NC_045512.2	29332	29377	44501	1	45	15	NKHIDAYKTFPPTEP	14.0795	https://www.iedb.org/epitope/44501
NC_045512.2	9124	9169	76351	1	45	15	YVLMDGSIIQFPNTY	14.0795	https://www.iedb.org/epitope/76351
NC_045512.2	9118	9163	66138	1	45	15	TRYVLMDGSIIQFPN	14.0691	https://www.iedb.org/epitope/66138
NC_045512.2	18348	18393	44839	1	45	15	NLPLQLGFSTGVNLV	14.0601	https://www.iedb.org/epitope/44839
NC_045512.2	9121	9166	56677	1	45	15	RYVLMDGSIIQFPNT	14.0399	https://www.iedb.org/epitope/56677
NC_045512.2	18855	18900	7545	1	45	15	DAIMTRCLAVHECFV	14.0201	https://www.iedb.org/epitope/7545
NC_045512.2	7945	7987	75868	1	42	14	YSQLMCQPILLLDQV	14.0006	https://www.iedb.org/epitope/75868
NC_045512.2	15417	15462	4092	1	45	15	AQVLSEMVMCGGSLY	13.9911	https://www.iedb.org/epitope/4092
NC_045512.2	20817	20862	38352	1	45	15	LNTLTLAVPYNMRVI	13.9722	https://www.iedb.org/epitope/38352
NC_045512.2	28750	28795	51483	1	45	15	QLPQGTTLPKGFYAE	13.9422	https://www.iedb.org/epitope/51483
NC_045512.2	28990	29044	52117	1	54	18	QQQGQTVTKKSAAEASKK	13.9388	https://www.iedb.org/epitope/52117
NC_045512.2	17859	17904	10974	1	45	15	DYVIFTQTTETAHSC	13.932	https://www.iedb.org/epitope/10974
NC_045512.2	16200	16245	73621	1	45	15	YEAMYTPHTVLQAVG	13.8935	https://www.iedb.org/epitope/73621
NC_045512.2	19422	19467	8614	1	45	15	DIDYVPLKSATCITR	13.8838	https://www.iedb.org/epitope/8614
NC_045512.2	25042	25087	60024	1	45	15	SPDVDLGDISGINAS	13.8838	https://www.iedb.org/epitope/60024
NC_045512.2	14049	14094	69808	1	45	15	VLTLDNQDLNGNWYD	13.8439	https://www.iedb.org/epitope/69808
NC_045512.2	24016	24061	11384	1	45	15	EDLLFNKVTLADAGF	13.8439	https://www.iedb.org/epitope/11384
NC_045512.2	10855	10900	36836	1	45	15	LKELLQNGMNGRTIL	13.8347	https://www.iedb.org/epitope/36836
NC_045512.2	15453	15498	59624	1	45	15	SLYVKPGGTSSGDAT	13.8347	https://www.iedb.org/epitope/59624
NC_045512.2	24475	24526	2092	1	51	17	AISSVLNDILSRLDKVE	13.8286	https://www.iedb.org/epitope/2092
NC_045512.2	12619	12664	60163	1	45	15	SPNLAWPLIVTALRA	13.8056	https://www.iedb.org/epitope/60163
NC_045512.2	16929	16974	70009	1	45	15	VMPLSAPTLVPQEHY	13.8056	https://www.iedb.org/epitope/70009
NC_045512.2	17652	17697	32314	1	45	15	KMFYKGVITHDVSSA	13.8056	https://www.iedb.org/epitope/32314
NC_045512.2	27453	27498	50560	1	45	15	QECVRGTTVLLKEPC	13.7957	https://www.iedb.org/epitope/50560
NC_045512.2	11203	11248	67813	1	45	15	VAYFNMVYMPASWVM	13.786	https://www.iedb.org/epitope/67813
NC_045512.2	15018	15063	35801	1	45	15	LFAYTKRNVIPTITQ	13.7572	https://www.iedb.org/epitope/35801
NC_045512.2	24409	24460	10778	1	51	17	DVVNQNAQALNTLVKQL	13.7572	https://www.iedb.org/epitope/10778
NC_045512.2	14148	14193	76601	1	45	15	YYSLLMPILTLTRAL	13.7385	https://www.iedb.org/epitope/76601
NC_045512.2	17262	17307	15413	1	45	15	FDKFKVNSTLEQYVF	13.7293	https://www.iedb.org/epitope/15413
NC_045512.2	17097	17142	33338	1	45	15	KSHFAIGLALYYPSA	13.6845	https://www.iedb.org/epitope/33338
NC_045512.2	24232	24277	18514	1	45	15	GAALQIPFAMQMAYR	13.6845	https://www.iedb.org/epitope/18514
NC_045512.2	15315	15360	42301	1	45	15	MPNMLRIMASLVLAR	13.6085	https://www.iedb.org/epitope/42301
NC_045512.2	16128	16173	20160	1	45	15	GHMLDMYSVMLTNDN	13.6085	https://www.iedb.org/epitope/20160
NC_045512.2	27035	27083	66409	1	48	16	TSRTLSYYKLGASQRV	13.6085	https://www.iedb.org/epitope/66409
NC_045512.2	17376	17421	46593	1	45	15	NYDLSVVNARLRAKH	13.6004	https://www.iedb.org/epitope/46593
NC_045512.2	13926	13971	73546	1	45	15	YDFVENPDILRVYAN	13.5531	https://www.iedb.org/epitope/73546
NC_045512.2	21402	21447	30814	1	45	15	KFPLKLRGTAVMSLK	13.5531	https://www.iedb.org/epitope/30814
NC_045512.2	16896	16941	68645	1	45	15	VGDYFVLTSHTVMPL	13.5379	https://www.iedb.org/epitope/68645
NC_045512.2	12685	12730	45072	1	45	15	NNELSPVALRQMSCA	13.5303	https://www.iedb.org/epitope/45072
NC_045512.2	17340	17385	8822	1	45	15	DIVVFDEISMATNYD	13.508	https://www.iedb.org/epitope/8822
NC_045512.2	14142	14187	10288	1	45	15	DSYYSLLMPILTLTR	13.4509	https://www.iedb.org/epitope/10288
NC_045512.2	16923	16968	24998	1	45	15	HTVMPLSAPTLVPQE	13.4439	https://www.iedb.org/epitope/24998
NC_045512.2	29200	29245	56980	1	45	15	SASAFFGMSRIGMEV	13.4439	https://www.iedb.org/epitope/56980
NC_045512.2	15045	15090	28088	1	45	15	IPTITQMNLKYAISA	13.4167	https://www.iedb.org/epitope/28088
NC_045512.2	11419	11464	74537	1	45	15	YKVYYGNALDQAISM	13.4034	https://www.iedb.org/epitope/74537
NC_045512.2	16125	16170	63847	1	45	15	TGHMLDMYSVMLTND	13.4034	https://www.iedb.org/epitope/63847
NC_045512.2	27014	27059	48051	1	45	15	PKEITVATSRTLSYY	13.3968	https://www.iedb.org/epitope/48051
NC_045512.2	17661	17706	74457	1	45	15	YKGVITHDVSSAINR	13.3773	https://www.iedb.org/epitope/74457
NC_045512.2	16221	16266	24997	1	45	15	HTVLQAVGACVLCNS	13.3644	https://www.iedb.org/epitope/24997
NC_045512.2	11686	11731	36351	1	45	15	LGVYDYLVSTQEFRY	13.3581	https://www.iedb.org/epitope/36351
NC_045512.2	12055	12100	6161	1	45	15	CEEMLDNRATLQAIA	13.3087	https://www.iedb.org/epitope/6161
NC_045512.2	15318	15363	48677	1	45	15	PNMLRIMASLVLARK	13.3087	https://www.iedb.org/epitope/48677
NC_045512.2	24538	24583	25662	1	45	15	IDRLITGRLQSLQTY	13.279	https://www.iedb.org/epitope/25662
NC_045512.2	23326	23371	47041	1	45	15	PCSFGGVSVITPGTN	13.2559	https://www.iedb.org/epitope/47041
NC_045512.2	14427	14472	70758	1	45	15	VRKIFVDGVPFVVST	13.2389	https://www.iedb.org/epitope/70758
NC_045512.2	24115	24160	3982	1	45	15	AQKFNGLTVLPPLLT	13.2389	https://www.iedb.org/epitope/3982
NC_045512.2	28765	28810	66707	1	45	15	TTLPKGFYAEGSRGG	13.2277	https://www.iedb.org/epitope/66707
NC_045512.2	18024	18069	67769	1	45	15	VATLQAENVTGLFKD	13.1896	https://www.iedb.org/epitope/67769
NC_045512.2	10003	10048	10672	1	45	15	DVLYQPPQTSITSAV	13.1789	https://www.iedb.org/epitope/10672
NC_045512.2	11812	11857	32667	1	45	15	KPCIKVATVQSKMSD	13.1789	https://www.iedb.org/epitope/32667
NC_045512.2	17343	17388	29492	1	45	15	IVVFDEISMATNYDL	13.1789	https://www.iedb.org/epitope/29492
NC_045512.2	9139	9184	22376	1	45	15	GSIIQFPNTYLEGSV	13.1789	https://www.iedb.org/epitope/22376
NC_045512.2	16971	17016	76399	1	45	15	YVRITGLYPTLNISD	13.1124	https://www.iedb.org/epitope/76399
NC_045512.2	27002	27047	26759	1	45	15	IKDLPKEITVATSRT	13.0928	https://www.iedb.org/epitope/26759
NC_045512.2	16902	16947	10887	1	45	15	DYFVLTSHTVMPLSA	13.0688	https://www.iedb.org/epitope/10887
NC_045512.2	18015	18060	55665	1	45	15	RRNVATLQAENVTGL	13.05	https://www.iedb.org/epitope/55665
NC_045512.2	16905	16950	73959	1	45	15	YFVLTSHTVMPLSAP	13.0408	https://www.iedb.org/epitope/73959
NC_045512.2	15141	15186	45723	1	45	15	NRQFHQKLLKSIAAT	13.0225	https://www.iedb.org/epitope/45723
NC_045512.2	14643	14688	65382	1	45	15	TNNVAFQTVKPGNFN	13.0046	https://www.iedb.org/epitope/65382
NC_045512.2	17439	17484	46921	1	45	15	PAQLPAPRTLLTKGT	13.0046	https://www.iedb.org/epitope/46921
NC_045512.2	24592	24643	51379	1	51	17	QLIRAAEIRASANLAAT	12.9957	https://www.iedb.org/epitope/51379
NC_045512.2	28744	28789	69721	1	45	15	VLQLPQGTTLPKGFY	12.9913	https://www.iedb.org/epitope/69721
NC_045512.2	15153	15198	24586	1	45	15	HQKLLKSIAATRGAT	12.9739	https://www.iedb.org/epitope/24586
NC_045512.2	14532	14577	54757	1	45	15	RLSFKELLVYAADPA	12.9654	https://www.iedb.org/epitope/54757
NC_045512.2	15108	15153	67621	1	45	15	VAGVSICSTMTNRQF	12.9654	https://www.iedb.org/epitope/67621
NC_045512.2	16851	16896	18936	1	45	15	GDAVVYRGTTTYKLN	12.9654	https://www.iedb.org/epitope/18936
NC_045512.2	12226	12271	33297	1	45	15	KSEFDRDAAMQRKLE	12.9526	https://www.iedb.org/epitope/33297
NC_045512.2	17031	17076	67700	1	45	15	VANYQKVGMQKYSTL	12.9235	https://www.iedb.org/epitope/67700
NC_045512.2	26678	26723	16974	1	45	15	FLWLLWPVTLACFVL	12.9235	https://www.iedb.org/epitope/16974

Chr	Start	Stop	EpitopeID	Identity	Nucleotide length (bp)	AA length	Epitope sequence	-log(eValue)	Link
NC_045512.2	15675	15720	73478	1	45	15	YAYLRKHFSMMILSD	12.9032	https://www.iedb.org/epitope/73478
NC_045512.2	16839	16884	30915	1	45	15	KGDYGDAVVYRGTTT	12.9032	https://www.iedb.org/epitope/30915
NC_045512.2	29335	29380	31115	1	45	15	KHIDAYKTFPPTEPK	12.8833	https://www.iedb.org/epitope/31115
NC_045512.2	15993	16038	22745	1	45	15	GTLMIERFVSLAIDA	12.8447	https://www.iedb.org/epitope/22745
NC_045512.2	23173	23218	7289	1	45	15	CVNFNFNGLTGTGVL	12.8372	https://www.iedb.org/epitope/7289
NC_045512.2	14139	14184	68075	1	45	15	VDSYYSLLMPILTLT	12.826	https://www.iedb.org/epitope/68075
NC_045512.2	12694	12739	39582	1	45	15	LSPVALRQMSCAAGT	12.7859	https://www.iedb.org/epitope/39582
NC_045512.2	10792	10837	67984	1	45	15	VDILGPLSAQTGIAV	12.7682	https://www.iedb.org/epitope/67984
NC_045512.2	15516	15561	68533	1	45	15	VFNICQAVTANVNAL	12.7577	https://www.iedb.org/epitope/68533
NC_045512.2	24013	24058	25754	1	45	15	IEDLLFNKVTLADAG	12.7577	https://www.iedb.org/epitope/25754
NC_045512.2	17382	17427	9289	1	45	15	DLSVVNARLRAKHYV	12.7371	https://www.iedb.org/epitope/9289
NC_045512.2	9157	9202	48713	1	45	15	PNTYLEGSVRVVTTF	12.7004	https://www.iedb.org/epitope/48713
NC_045512.2	12976	13021	45682	1	45	15	NRGMVLGSLAATVRL	12.6906	https://www.iedb.org/epitope/45682
NC_045512.2	16920	16965	58429	1	45	15	SHTVMPLSAPTLVPQ	12.6809	https://www.iedb.org/epitope/58429
NC_045512.2	16404	16449	71280	1	45	15	VTDVTQLYLGGMSYY	12.6681	https://www.iedb.org/epitope/71280
NC_045512.2	19416	19461	70900	1	45	15	VSDIDYVPLKSATCI	12.5917	https://www.iedb.org/epitope/70900
NC_045512.2	16671	16716	32185	1	45	15	KLSYGIATVREVLSD	12.58	https://www.iedb.org/epitope/32185
NC_045512.2	12979	13024	53926	1	45	15	RGMVLGSLAATVRLQ	12.5714	https://www.iedb.org/epitope/53926
NC_045512.2	11173	11218	6560	1	45	15	CLFLLPSLATVAYFN	12.5235	https://www.iedb.org/epitope/6560
NC_045512.2	14688	14733	30149	1	45	15	KDFYDFAVSKGFFKE	12.5126	https://www.iedb.org/epitope/30149
NC_045512.2	13794	13839	40004	1	45	15	LTKYTMADLVYALRH	12.4831	https://www.iedb.org/epitope/40004
NC_045512.2	295	340	33697	1	45	15	KTHVQLSLPVLQVRD	12.4831	https://www.iedb.org/epitope/33697
NC_045512.2	12265	12310	35555	1	45	15	LEKMADQAMTQMYKQ	12.4648	https://www.iedb.org/epitope/35555
NC_045512.2	15510	15555	46023	1	45	15	NSVFNICQAVTANVN	12.4648	https://www.iedb.org/epitope/46023
NC_045512.2	29350	29395	74516	1	45	15	YKTFPPTEPKKDKKK	12.4648	https://www.iedb.org/epitope/74516
NC_045512.2	11194	11239	34992	1	45	15	LATVAYFNMVYMPAS	12.4443	https://www.iedb.org/epitope/34992
NC_045512.2	16137	16182	35230	1	45	15	LDMYSVMLTNDNTSR	12.4242	https://www.iedb.org/epitope/35230
NC_045512.2	7825	7870	18245	1	45	15	FVNLDNLRANNTKGS	12.4242	https://www.iedb.org/epitope/18245
NC_045512.2	13009	13054	67134	1	45	15	TVRLQAGNATEVPAN	12.4143	https://www.iedb.org/epitope/67134
NC_045512.2	14991	15036	59754	1	45	15	SMSYEDQDALFAYTK	12.3948	https://www.iedb.org/epitope/59754
NC_045512.2	16620	16665	54616	1	45	15	RLKLFAAETLKATEE	12.3948	https://www.iedb.org/epitope/54616
NC_045512.2	26813	26858	4406	1	45	15	ASFRLFARTRSMWSF	12.3852	https://www.iedb.org/epitope/4406
NC_045512.2	29215	29260	16012	1	45	15	FGMSRIGMEVTPSGT	12.3852	https://www.iedb.org/epitope/16012
NC_045512.2	12145	12190	50555	1	45	15	QEAYEQAVANGDSEV	12.3757	https://www.iedb.org/epitope/50555
NC_045512.2	16665	16710	63688	1	45	15	TFKLSYGIATVREVL	12.3453	https://www.iedb.org/epitope/63688
NC_045512.2	29014	29065	31692	1	51	17	KKSAAEASKKPRQKRTA	12.3294	https://www.iedb.org/epitope/31692
NC_045512.2	17568	17613	46782	1	45	15	PAEIVDTVSALVYDN	12.2574	https://www.iedb.org/epitope/46782
NC_045512.2	17115	17160	20765	1	45	15	GLALYYPSARIVYTA	12.249	https://www.iedb.org/epitope/20765
NC_045512.2	12430	12475	7297	1	45	15	CVPLNIIPLTTAAKL	12.2001	https://www.iedb.org/epitope/7297
NC_045512.2	12964	13009	38283	1	45	15	LNNLNRGMVLGSLAA	12.1902	https://www.iedb.org/epitope/38283
NC_045512.2	18378	18420	23072	1	42	14	GVNLVAVPTGYVDTE	12.1707	https://www.iedb.org/epitope/23072
NC_045512.2	16857	16902	5611	1	45	15	AVVYRGTTTYKLNVG	12.131	https://www.iedb.org/epitope/5611
NC_045512.2	24574	24619	52672	1	45	15	QTYVTQQLIRAAEIR	12.131	https://www.iedb.org/epitope/52672
NC_045512.2	12058	12103	11787	1	45	15	EEMLDNRATLQAIAS	12.1108	https://www.iedb.org/epitope/11787
NC_045512.2	16623	16668	36937	1	45	15	LKLFAAETLKATEET	12.0821	https://www.iedb.org/epitope/36937
NC_045512.2	12982	13027	21358	1	45	15	GMVLGSLAATVRLQA	11.9942	https://www.iedb.org/epitope/21358
NC_045512.2	18519	18564	71813	1	45	15	VVRIKIVQMLSDTLK	11.9942	https://www.iedb.org/epitope/71813
NC_045512.2	12631	12676	5687	1	45	15	AWPLIVTALRANSAV	11.975	https://www.iedb.org/epitope/5687
NC_045512.2	17769	17814	28660	1	45	15	ISPYNSQNAVASKIL	11.975	https://www.iedb.org/epitope/28660
NC_045512.2	29185	29230	3957	1	45	15	AQFAPSASAFFGMSR	11.9639	https://www.iedb.org/epitope/3957
NC_045512.2	25213	25258	36103	1	45	15	LGFIAGLIAIVMVTI	11.9254	https://www.iedb.org/epitope/36103
NC_045512.2	13033	13078	4950	1	45	15	ATEVPANSTVLSFCA	11.8957	https://www.iedb.org/epitope/4950
NC_045512.2	14631	14676	67526	1	45	15	VAALTNNVAFQTVKP	11.7588	https://www.iedb.org/epitope/67526
NC_045512.2	12436	12481	48427	1	45	15	PLNIIPLTTAAKLMV	11.7486	https://www.iedb.org/epitope/48427
NC_045512.2	11818	11863	6425	1	45	15	CIKVATVQSKMSDVK	11.7102	https://www.iedb.org/epitope/6425
NC_045512.2	15156	15201	51217	1	45	15	QKLLKSIAATRGATV	11.7005	https://www.iedb.org/epitope/51217
NC_045512.2	29341	29386	25542	1	45	15	IDAYKTFPPTEPKKD	11.6802	https://www.iedb.org/epitope/25542
NC_045512.2	21408	21450	48380	1	42	14	PLKLRGTAVMSLKEN	11.6216	https://www.iedb.org/epitope/48380
NC_045512.2	14616	14661	66551	1	45	15	TTCFSVAALTNNVAF	11.5823	https://www.iedb.org/epitope/66551
NC_045512.2	15054	15099	29061	1	45	15	ITQMNLKYAISAKNR	11.5823	https://www.iedb.org/epitope/29061
NC_045512.2	11170	11215	35072	1	45	15	LCLFLLPSLATVAYF	11.524	https://www.iedb.org/epitope/35072
NC_045512.2	11167	11212	16538	1	45	15	FLCLFLLPSLATVAY	11.5139	https://www.iedb.org/epitope/16538
NC_045512.2	14151	14193	75796	1	42	14	YSLLMPILTLTRALA	11.4834	https://www.iedb.org/epitope/75796
NC_045512.2	24436	24481	38353	1	45	15	LNTLVKQLSSNFGAI	11.4737	https://www.iedb.org/epitope/38353
NC_045512.2	17799	17844	4502	1	45	15	ASKILGLPTQTVDSS	11.4641	https://www.iedb.org/epitope/4502
NC_045512.2	301	346	25079	1	45	15	HVQLSLPVLQVRDVL	11.4641	https://www.iedb.org/epitope/25079
NC_045512.2	15324	15366	42048	1	42	14	MLRIMASLVLARKHN	11.4267	https://www.iedb.org/epitope/42048
NC_045512.2	28780	28825	19650	1	45	15	GFYAEGSRGGSQASS	11.3474	https://www.iedb.org/epitope/19650
NC_045512.2	15060	15105	51627	1	45	15	QMNLKYAISAKNRAR	11.3141	https://www.iedb.org/epitope/51627
NC_045512.2	24553	24598	63951	1	45	15	TGRLQSLQTYVTQQL	11.1908	https://www.iedb.org/epitope/63951
NC_045512.2	12670	12715	5406	1	45	15	AVKLQNNELSPVALR	11.1209	https://www.iedb.org/epitope/5406
NC_045512.2	25057	25102	36075	1	45	15	LGDISGINASVVNIQ	11.0747	https://www.iedb.org/epitope/36075
NC_045512.2	24562	24607	38990	1	45	15	LQSLQTYVTQQLIRA	11.0122	https://www.iedb.org/epitope/38990
NC_045512.2	24571	24616	39003	1	45	15	LQTYVTQQLIRAAEI	10.9764	https://www.iedb.org/epitope/39003
NC_045512.2	12646	12691	71250	1	45	15	VTALRANSAVKLQNN	10.9086	https://www.iedb.org/epitope/71250
NC_045512.2	24403	24448	38831	1	45	15	LQDVVNQNAQALNTL	10.9086	https://www.iedb.org/epitope/38831
NC_045512.2	12637	12682	48372	1	45	15	PLIVTALRANSAVKL	10.8658	https://www.iedb.org/epitope/48372
NC_045512.2	24460	24505	61229	1	45	15	SSNFGAISSVLNDIL	10.8098	https://www.iedb.org/epitope/61229
NC_045512.2	24589	24634	52057	1	45	15	QQLIRAAEIRASANL	10.7615	https://www.iedb.org/epitope/52057

Chr	Start	Stop	EpitopeID	Identity	Nucleotide length (bp)	AA length	Epitope sequence	-log(eValue)	Link
NC_045512.2	28508	28550	71802	1	42	14	VVQMTKLATTEELPD	10.6209	https://www.iedb.org/epitope/71802
NC_045512.2	25066	25111	28511	1	45	15	ISGINASVVNIQKEI	10.4177	https://www.iedb.org/epitope/28511
NC_045512.2	11977	12022	63292	1	45	15	TEAFEKMVSLLSVLL	10.3374	https://www.iedb.org/epitope/63292
NC_045512.2	26268	26313	63980	1	45	15	TGTLIVNSVLLFLAF	9.93601	https://www.iedb.org/epitope/63980
NC_045512.2	11998	12043	71036	1	45	15	VSLLSVLLSMQGAVD	9.86812	https://www.iedb.org/epitope/71036
NC_045512.2	12556	12601	28211	1	45	15	IQQVVDADSKIVQLS	9.81	https://www.iedb.org/epitope/28211
NC_045512.2	26325	26370	40612	1	45	15	LVTLAILTALRLCAY	9.7795	https://www.iedb.org/epitope/40612
NC_045512.2	27809	27854	16954	1	45	15	FLVLIMLIIFWFSLE	9.71122	https://www.iedb.org/epitope/16954
NC_045512.2	28999	29044	21967	1	45	15	GQTVTKKSAAEASKK	9.71122	https://www.iedb.org/epitope/21967
NC_045512.2	26271	26316	22741	1	45	15	GTLIVNSVLLFLAFV	9.49536	https://www.iedb.org/epitope/22741
NC_045512.2	26570	26609	35672	1	39	13	LEQWNLVIGFLFL	9.46008	https://www.iedb.org/epitope/35672
NC_045512.2	26319	26364	16763	1	45	15	FLLVTLAILTALRLC	8.96348	https://www.iedb.org/epitope/16763
NC_045512.2	27806	27851	35912	1	45	15	LFLVLIMLIIFWFSL	8.55282	https://www.iedb.org/epitope/35912
NC_045512.2	26274	26319	64867	1	45	15	TLIVNSVLLFLAFVV	8.2208	https://www.iedb.org/epitope/64867
NC_045512.2	26313	26358	71664	1	45	15	VVFLLVTLAILTALR	8.04719	https://www.iedb.org/epitope/71664
NC_045512.2	26307	26352	1419	1	45	15	AFVVFLLVTLAILTA	7.99543	https://www.iedb.org/epitope/1419
NC_045512.2	26316	26361	68513	1	45	15	VFLLVTLAILTALRL	7.85967	https://www.iedb.org/epitope/68513
NC_045512.2	11047	11077	59262	1	30	10	SLLILVQSTQWSLFF	4.13517	https://www.iedb.org/epitope/59262
NC_045512.2	11047	11074	66300	1	27	9	TSLLILVQSTQWSLF	2.78062	https://www.iedb.org/epitope/66300

D.6 Appendix D References

1. Hadfield, J., et al., Nextstrain: real-time tracking of pathogen evolution. Bioinformatics, 2018. 34(23): p. 4121-4123.
2. Li, D., et al., WashU Epigenome Browser update 2019. Nucleic Acids Res, 2019. 47(W1): p. W158-W165.
3. Zhou, X., et al., Epigenomic annotation of genetic variants using the Roadmap Epigenome Browser. Nat Biotechnol, 2015. 33(4): p. 345-6.
4. Zhou, X., et al., Exploring long-range genome interactions using the WashU Epigenome Browser. Nat Methods, 2013. 10(5): p. 375-6.
5. Zhou, X., et al., The Human Epigenome Browser at Washington University. Nat Methods, 2011. 8(12): p. 989-90.
6. Blanco-Melo, D., et al., Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell, 2020. 10.
7. Shu, Y. and J. McCauley, GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill, 2017. 22(13).
8. Coordinators, N.R., Database resources of the National Center for Biotechnology Information. Nucleic Acids Res, 2018. 46(D1): p. D8-D13.
9. Myers, E.W. and W. Miller, Optimal alignments in linear space. Comput Appl Biosci, 1988. 4(1): p. 11-7.
10. Broughton, J.P., et al., CRISPR-Cas12-based detection of SARS-CoV-2. Nat Biotechnol, 2020.
11. Campbell, K.M., et al., Prediction of SARS-CoV-2 epitopes across 9360 HLA class I a lleles. bioRxiv, 2020.
12. Vita, R., et al., The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res, 2019. 47(D1): p. D339-D343.
13. Nerli, S. and N.G. Sgourakis, Structure-based modeling of SARS-CoV-2 peptide/HLA-A02 antigens. bioRxiv, 2020.
14. Grifoni, A., et al., Candidate Targets for Immune Responses to 2019-Novel Coronavirus (nCoV): Sequence Homology- and Bioinformatic-Based Predictions. SSRN, 2020: p. 3541361.
15. Poran, A., et al., Sequence-based prediction of vaccine targets for inducing T cell responses to SARS-CoV-2 utilizing the bioinformatics predictor RECON. bioRxiv, 2020.
16. Braun, J., et al., SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. Nature, 2020.
17. Kim, D., et al., The Architecture of SARS-CoV-2 Transcriptome. Cell, 2020.
18. Katoh, K. and D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol, 2013. 30(4): p. 772-80.
19. Katoh, K., et al., MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res, 2002. 30(14): p. 3059-66.
20. Price, M.N., P.S. Dehal, and A.P. Arkin, FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One, 2010. 5(3): p. e9490.
21. Price, M.N., P.S. Dehal, and A.P. Arkin, FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol, 2009. 26(7): p. 1641-50.
22. Liu, P., W. Chen, and J.P. Chen, Viral Metagenomics Revealed Sendai Virus and Coronavirus Infection of Malayan Pangolins (Manis javanica). Viruses, 2019. 11(11).

23. Zhou, P., et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature, 2020. 579(7798): p. 270-273.
24. Imai, K., et al., Rapid and Accurate Species Identification of Mitis Group Streptococci Using the MinION Nanopore Sequencer. Front Cell Infect Microbiol, 2020. 10: p. 11.
25. Li, H. and R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics, 2010. 26(5): p. 589-95.
26. Li, H., et al., The Sequence Alignment/Map format and SAMtools. Bioinformatics, 2009. 25(16): p. 2078-9.
27. Moudgil, A., et al., The qBED track: a novel genome browser visualization for point processes. bioRxiv, 2020.
28. Gertz, E.M., et al., Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol, 2006. 4: p. 41.
29. Wu, F., et al., A new coronavirus associated with human respiratory disease in China. Nature, 2020. 579(7798): p. 265-269.
30. Zhang, Z., et al., A greedy algorithm for aligning DNA sequences. J Comput Biol, 2000. 7(1-2): p. 203-14.
31. Ramirez, F., et al., deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res, 2016. 44(W1): p. W160-5.
32. Flynn, J.A., et al., Exploring the coronavirus epidemic using the new WashU Virus Genome Browser. bioRxiv, 2020.

CV

Jennifer Karlow (Flynn)
jaflynn@wustl.edu • 314-799-8243

EDUCATION

Washington University — St. Louis, MO, USA
Doctorate of Philosophy Candidate; Computation and Systems Biology — *2014 – 2021*

Truman State University — Kirksville, MO, USA
Bachelor of Science — *2010 - 2014*
Major: Biology, Minor: Mathematical Biology
Summa Cum Laude (GPA: 3.96/4.0)

RESEARCH EXPERIENCE

Washington University: *Computational and Systems Biology Ph.D. Thesis Work* — St. Louis, MO
Thesis Mentor: Dr. Ting Wang — *June 2015 – October 2021*
Objective of Project: The objective of my thesis work was to better understand how cells undergo epigenetic reprogramming across cancer types and through tumor progression into metastasis. To accomplish this, I identified shared DNA methylation abnormalities in two distinct cancer types, endometrioid adenocarcinoma and glioblastoma, indicating mutual loss of original cell type enhancer functionality. In addition, I globally characterized DNA methylation, transcriptomic, and whole exome changes between primary non-small cell lung cancer (NSCLC) tumors and matched brain metastases to predict regulatory regions where methylation change may be driving a change in gene expression and/or propagating the metastasis phenotype. Using top prioritized alterations, I clinically validating their predictive prognostic power as well as experimentally elucidated their relationship with altered polycomb repressive complex II binding. Additional projects that I contributed heavily to include studying how smoking impacts the epigenetic landscape in NSCLC, the creation of the WashU Virus Genome Browser, and the implementation of Capture MRE-seq, suitable for probing the methylation status within degraded DNA samples.

Washington University: *Computational and Systems Biology Ph.D. Rotation* — St. Louis, MO
Mentor: Dr. Ting Wang — *Jan. 2015 – May 2015*
Objective: I worked with others to determine if cancer-related GWAS SNPs preferentially fell within cell-type-specific enhancer regions. Using Python, I implemented a series of statistical tests using annotated, cell-type-specific BED files from the Roadmap Epigenomics Project and the ENCODE project.

Washington University: *Computational and Systems Biology Ph.D. Rotation* — St. Louis, MO
Mentor: Dr. Li Ding — *Sep. 2014 – Dec. 2014*
Objective: I collaborated with other members of the lab to determine whether viral integration into human host genomes significantly altered expression of genes upstream and downstream of the integration site. I implemented a series of statistical tests on RPKM data at both the exon-level and the gene-level to identify if viral integration was associated with altered expression in tumor and normal samples.

Washington University: *Computational and Systems Biology Ph.D. Rotation* — St. Louis, MO
Mentor: Dr. Kristen Naegle — *June 2014 – Sep. 2014*
Objective: I assisted in building a model for predicting acute myeloid leukemia resistance for the DREAM 9 Challenge, utilizing various machine learning techniques. Using MATLAB, I built support vector machine models, optimized those models, evaluated them, and assigned classification confidences. I also implemented additional machine learning and statistical techniques including PLSR, PCA, kmeans clustering, and random forest classification.

Siteman Cancer Center: *Leah Menshouse Springer Summer Student Program* — St. Louis, MO
Mentors: Dr. Kristen Naegle, Dr. Reid Townsend, Dr. Matthew Ellis, and Dr. Sherri Davies — *Summer 2013*
Objective: The objective of this project was to understand dynamic phosphorylation patterns in Paclitaxol-treated breast tumor tissue using computational techniques. I analyzed high-throughput mass spectrometry data using Scaffold, wrote scripts to automate the pipeline of data processing using MATLAB and Python,

helped modify existing programs for efficiency, partitioned all phospho-peptides in the tumor sample into clusters based on their temporal expression profiles post-treatment, and generated heat maps and co-occurrence matrices.

Truman State University: *Mathematical Biology Program* — Kirksville, MO
Mentors: Dr. John Ma and Dr. Todd Hammond — *2012 – 2013*
Objective: The objective of this project was to determine the effect of elevated CO_2 on phosphorus uptake in common bean plants. To accomplish this, I ordered and assembled equipment for our study. I made growth media, hydroponic solutions, and mixed reagents for the bean plants. I also analyzed samples from each plant jar for phosphorus concentration using a spectrophotometer every four days during growth. I operated the freeze dryer to obtain dry weights of the shoots, roots, and fruits of each sample. I assisted in modeling the rate of phosphorus uptake for each experimental condition, and designed and presented a poster at conferences.

Truman State University: *Volunteer in Biofeedback Laboratory* — Kirksville, MO
Mentor: Dr. Fredric Shaffer — *2011 – 2012*
Objective: The objective of this project was to establish the heart's response to stressors to better understand heartrate variability. I took volunteers' blood pressure, assisted in recording their brain activity using an EEG, and their heart activity using an EKG. I also aided in interpreting the EEG and EKG data.

PUBLICATIONS

Karlow, J.A., Devarakonda, S., Xing, X., Watson, M., Ramaswamy, G., Wang, T., DNA methylation alterations during non-small cell lung cancer metastasis to brain suggest epigenetic reprogramming of brain development pathways. (*In preparation*)

Karlow, J.A. *, Devarakonda, S. *, Sankararaman, S., Kaushal, M., Xing, X., Borgia, J.A., Stinchomb, T., Kalemkeria, G., Gray, J.E., Bradley, J., Alldredge, P., Snider, J., Bernadt, C., Wang, T., Ramaswamy, G. †, Watson, M.A. †, Multiomic comparison of primary and brain metastatic non-small cell lung cancer suggests tumor cell reprogramming towards a glial cell phenotype. (*In preparation*) *equal contribution; †corresponding authors

Xing, X.*, Karlow, J.A.*, Li, D., Wang, T., CG-overhang MRE-Seq for methylome analysis of highly degraded DNA samples. (*Under Review at Methods in Molecular Biology*) *equal contribution

Karlow, J.A.*, Miao, B., Xing, X., Wang, T. *, Zhang, B. *, Common DNA methylation dynamics in endometrioid adenocarcinoma and glioblastoma suggest universal epigenomic alterations in tumorigenesis. *Commun Biol* **4**, (2021) *corresponding authors

Flynn, J.A.*, Purushotham, D.*, Choudhary, M.N.K.*, Zhuo, X.*, Fan, C.*, Matt, G.*, Li, D.†, Wang, T. †, Exploring the coronavirus pandemic with the WashU Virus Genome Browser. *Nat Genet* **52**, 986-991 (2020) *equal contribution; †corresponding authors

Ma, Z., Flynn, J., Libra, G., Shi, Z., Elevated CO2 accelerates depletion of phosphorus by common bean (Phaseolus vulgaris) in association with altered leaf biochemical properties. *Pedosphere* **28**, 422-429 (2017).

Zhang, B. *, Madden, P., Gu, J., Xing, X., Sanker, S., Flynn, J., Kroll, K., Wang, T. *, Uncovering the transcriptomic and epigenomic landscape of nicotinic receptor genes in non-neuronal tissues. *BMC Genomics* **18**, 439 (2017) *corresponding authors

Wyczalkowski, M.A., Wylie, K.M., Cao, S., McLellan, M.D., Flynn, J., Huang, M., Ye., K., Fan, X., Chen, K., Wendl, M.C., Ding, L., BreakPoint Surveyor: a pipeline for structural variant visualization. *Bioinformatics* **33**, 3121-3122 (2017).

PRESENTATIONS

"Epigenetic reprogramming of brain development pathways during non-small cell lung cancer metastasis to brain"

- Washington University Molecular Genetics and Genomics, Computational and Systems Biology, Human and Statistical Genetics Retreat (Poster) — Potosi, MO *October 2019*

"DNA methylation alterations associated with non-small cell lung cancer metastasis to Brain suggest epigenetic reprogramming of brain development pathways

- 15th Course on Epigenetics (Poster) — Paris, France *March 2019*

"Exploring the activation of developmental programs during non-small cell lung cancer metastasis to brain"

- Center for Genome Sciences at Washington University Research Update (Talk) — St. Louis, MO *January 2019*

"Exploring DNA methylation valley changes during non-small cell lung cancer metastasis to brain"

- Washington University Molecular Genetics and Genomics, Computational and Systems Biology, Human and Statistical Genetics Retreat (Talk) — Potosi, MO *October 2018*

"Exploring gene body methylation increase during disease progression of non-small cell lung cancer metastasis to brain"

- Epigenomics at VARI Workshop (Poster) — Grand Rapids, MI *August 2018*
- Washington University Department of Genetics Research Talk (Talk) — St. Louis, MO *April 2018*

"Understanding the role of aberrant methylation in metastasis progression"

- Siteman Cancer Center Cancer Biology Predoctoral Pathway Retreat (Talk) — St. Louis, MO *January 2018*
- Washington University Computational and Systems Biology Program Retreat (Poster) — St. Louis, MO *September 2017*
- Gordon Research Conference: Cancer Genetics & Epigenetics (Poster) — Lucca, Italy *April 2017*
- Washington University Department of Genetics Research Talk (Talk) — St. Louis, MO *February 2017*
- Siteman Cancer Center Cancer Biology Predoctoral Pathway Retreat (Poster) — St. Louis, MO *January 2017*
- Center for Genome Sciences at Washington University Research Update (Talk) — St. Louis, MO *December 2016*
- Washington University Department of Genetics Research Talk (Talk) — St. Louis, MO *December 2016*

"Abnormal DNA methylation at regulatory regions suggests cancer cell identity crisis"

- The Systems Biology: Global Regulation of Gene Expression conference at Cold Spring Harbor (Poster) — Cold Spring Harbor, NY *March 2016*

"Epigenome reference aids identification of cancer cell identity crisis"

- National Institute of Environmental Health Services laboratory site visit for the TaRGET II Consortium at Washington University (Talk) — St. Louis, MO *February 2016*

"Drug response predictions using phosphoprotein signaling data and machine learning techniques"

- Siteman Cancer Center Seminar: Developmental & Regenerative Biology & Cell-Cell Communications in Cancer Research at Washington University (Talk) — St. Louis, MO *March 2015*

"Modeling the effect of elevated CO_2 on phosphorus uptake in common bean (*Phaseolus vulgaris*)"

- The Student Research Conference at Truman State University (Talk) — Kirksville, MO *April 2013*
- The 27th National Conference on Undergraduate Research (NCUR) at the University of Wisconsin – La Crosse (Poster) — La Crosse, WI *April 2013*
- The NIMBioS Undergraduate Research Conference at the Interface of Biology and Mathematics at the University of Tennessee in Knoxville (Poster) — Knoxville, TN *November 2012*

AWARDS, FELLOWSHIPS, and FUNDING

- Recipient of the Steve Johnson Award for best student poster at the MGG/CSB/HSG retreat — *2019*
- Recipient of the best poster award at the 15th course on Epigenetics at the Curie Institute — *2019*
- Selected from the Epigenomics at VARI Workshop to attend the 15th Course on Epigenetics at the Curie Institute — *2019*
- Selected to attend the Epigenomics at VARI Workshop — *2018*
- Siteman Cancer Center Cancer Biology Pre-Doctoral Trainee (fully funded) — *2016 – 2018*
- Washington University Precision Medicine Pathway Trainee (partial funding) — *2015 – 2017*
- Monsanto Fellowship Finalist (1 of 4) — *2016*
- National Science Foundation Graduate Research Fellowship Program – Honorable Mention — *2016*
- Summa Cum Laude graduate with Bachelor of Science in Biology from Truman State University — *2014*
- President's Combined Ability Scholarship at Truman State University — *2010 – 2014*
- President's Honorary Scholarship at Truman State University — *2010 – 2014*
- Named to the President's List at Truman State University for finishing the semester with a GPA of 4.0 (7 semesters), and named to the Provost's List at Truman State University for — *2010 – 2014*

finishing the semester with a GPA between 3.50 and 3.99 (1 semester)

- Ferguson Rotary Club Scholarship *2010*
- Federal Employee Education and Assistance Academic Scholarship *2010*

TEACHING EXPERIENCE

Gordon Research Conference:
Cancer Genetics & Epigenetics: *Assisted in running the Washington University EpiGenome Browser Workshop* Lucca, Italy
Responsibilities: I assisted in organizing the materials needed for this workshop, which included updating a workshop booklet and building demo browser sessions. I also helped give the workshop, where I demonstrated some of the browser functionality and assisted attendees with their hands-on demos. *April 2017*

Washington University: *Teaching assistant for Dr. Michael Brent's Computational Molecular Biology Course* St. Louis, MO
Responsibilities: I was responsible for grading all assignments throughout the course and administering all quizzes. Additionally, I gave two lectures during the semester. We hosted office hours twice weekly. *Fall 2015*

Truman State University: *Teaching assistant for Dr. Stephen Hudman's Biometry Course* Kirksville, MO
Responsibilities: I helped redesign some of the course material, which included compiling learning modules and reformatting exams. *Spring 2014*

Truman State University: *Teaching assistant for Dr. Stephen Hudman's Introduction to Ecology with Lab* Kirksville, MO
Responsibilities: I assisted in grading exams and laboratory reports. *Spring 2013*

Truman State University: *Teaching assistant for Instructor Maureen McHale's Intro. Biology II with Lab* Kirksville, MO
Responsibilities: I was responsible for setting up lab experiments, explaining lab procedures to students, answering students' questions about lab procedures, and assisting in grading students' assignments, presentations, quizzes, and exams. *Fall 2011*

LEADERSHIP

Truman State University: *Member of the Beta Beta Beta National Biological Honor Society* Kirksville, MO
Executive Board as Research Chair *2013 – 2014*
Responsibilities: I lead all committee meetings, organized Beta Beta Beta Research Day, and ran all weakly journal club meetings for the organization.

CAMPUS INVOLVEMENT

Washington University: *FoldIt volunteer* St. Louis, MO
Responsibilities: I assisted in the FoldIt program where I helped teach middle-school students from the inner city of St. Louis about protein folding. *2014*

Washington University: *Member of the Women's Club Volleyball Team* St. Louis, MO
Responsibilities: I served as co-captain of one of the two teams. We practiced twice per week and participated in approximately 4 tournaments per semester throughout the Midwest. *2014 – 2017*

Truman State University: *Executive member of the Beta Beta Beta National Biological Honor Society* Kirksville, MO
Responsibilities: I helped determine the criteria for Beta Beta Beta National Biological Honor Society scholarship. I put together a portfolio of the current research being done at Truman State and A. T. Still University for students to reference, and I designed a poster informing students about specific tests such as the MCAT and GRE. I also assisted in community upkeep. *2010 – 2014*

Truman State University: *Student mentor, Science on Saturdays Program* Kirksville, MO
Responsibilities: A few times a year, middle-school-aged children from rural Missouri communities would come to Truman State, and we would teach them science lessons, ranging from anatomy to basic physics and chemistry. *2012*

Truman State University: *Member, Women's Club Volleyball Team* Kirksville, MO
Responsibilities: Practiced 3 days/week and participated in approximately 4 tournaments/semester. *2011 – 2014*

Truman State University: *Member, Women's Club Ultimate Frisbee Team* Kirksville, MO
Responsibilities: Practiced 3 days/week and participated in approximately 4 tournaments/semester. *2010*