CHARACTERIZATION OF ABERRANT DNA METHYLATION CHANGES IN INFERTILITY AND PARAGANGLIOMAS

by

Jahnvi Pflüger

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Oncological Sciences

The University of Utah

August 2014

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Jahnvi Pflüger**

has been approved by the following supervisory committee members:

Bradley R Cairns	, Chair	**4.25.2014** Date Approved
Joshua Schiffman	, Member	**4.25.2014** Date Approved
Dennis Winge	, Member	**4.25.2014** Date Approved
Donald Ayer	, Member	**4.25.2014** Date Approved
David Jones	, Member	**4.25.2014** Date Approved

and by **Bradley R Cairns**, Chair/Dean of the Department/College/School of **Oncological Sciences**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Our work focussed on how germ cell DNA is packaged and if it is poised by distinctive chromatin to influence embryo development. Finally, is misregulation of that poising a common theme observed in infertility and cancer? We profiled the epigenome in mature human sperm and found that packaging in mature sperm revealed the presence of two programs – a future program, involved in guiding embryo development and a past program, involved in spermatogenesis (Chapter 2). Next, the clearest place a chromatin problem can manifest is in infertility. We asked if the DNA methylation status of seven imprinted regions can serve as a diagnostic to inform two groups of infertile patients about the risk of their offspring developing a disorder. Although our results did not provide a causal link for the trans-generational inheritance of DNA methylation defects leading to imprinting disease, it showed a strong correlation between infertility in males and aberrations of DNA methylation at select imprinted loci (Chapter 3). Taken together, our data suggests that germ cell chromatin plays a significant role in early embryonic development and infertility. Finally, we investigated how defects in a metabolic enzyme, succinate dehydrogenase (SDH) can have an impact on chromatin packaging and transcriptome of paragangliomas. We also queried the epigenetic status of paragangliomas lacking mutations in SDH. We compared our two PGL subclasses to a progenitor cell type, neural crest cells (NCCs). Strikingly, we found that both subclasses of PGLs are phenotypically very similar. Furthermore, they share the majority of regions that gain and lose DNA methylation compared to neural crest cells. Whole exome sequencing of both PGL subclasses shows

mutations in many epigenetic modifier genes and hence we speculate that in PGLs lacking SDH mutations, epigenetic enzymes may harbor mutations that could phenocopy the misregulation in SDH deficient tumors (Chapter 4). Together, we hope that by querying the epigenetic status in a normal system and comparing these findings to perturbed systems, we have gained more insight into the role of epigenetic misregulation in infertility and cancer.

“It always seems impossible until it’s done.” - Nelson Mandela

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

As I hum the song by The Beatles, “I get by with a little help from my friends” I realize it is the perfect description of the tremendous support I have received throughout graduate school. It is safe to say that graduate school has been the one experience in my life that has tested everything that I am made of, over and over again. Thankfully, through my journey, I have had many people by side, from my mentor, lab mates, family and friends. I thank my mentor, Brad Cairns, for constantly challenging me to be a better scientist, critical thinker and most of all to be an independent researcher. He is one of the most dynamic and intelligent scientists you can work with and I am grateful for the opportunity of being trained by him. I also want to acknowledge all my other committee members for their support, specifically Joshua Schiffman and Dennis Winge. Josh has inspired me to become more involved in clinically relevant research and helped put a face to the tumor samples I work with. His passion for helping patients and unrelenting positive attitude make him one of a kind. Dennis played a significant role in recruiting me to the University of Utah and has unconditionally supported and believed in me, especially when the going went rough. My lab mates, Cedric, Tanya, Tim, Maggie, Alisha, and Archana have patiently and graciously been my support system both scientifically and emotionally. I thank my friends, Ravi, Angela, Kian, Simon, Patrick, Kristen, Pete, Krista, Jessica, Jeff and Katherine for the many fun adventures and drunk nights that helped me keep my sanity through graduate school. I know the all-consuming experience of graduate school will always bond us together. Finally, I want to thank my pregraduate school friends, Joanna, Sanjivani and Priya for

their genuine friendship, love and respect.

I am who I am because of my parents and my sister, and for this I want to thank them with all my heart. Their unwavering faith in me and my abilities, their constant support and encouragement has been the main reason for my accomplishments. I get my scientific edge and work ethic from my mom, who is a physicist and an incredible woman. She has given me the opportunity to fly higher than she was ever allowed to and for this I am grateful beyond words can express. I get my “never say die” attitude from my father, who has been to hell and back in his lifetime, but has always emerged better and stronger. Finally, my older and extremely talented sister, has always pushed me outside my comfort zone and encouraged me to think outside the box and be creative, regardless of the outcome.

Last but not the least, I want to thank my wonderful, intelligent and handsome husband, Christian. Scientifically, he has helped me grow into a more perceptive and critical scientist and for this I will be forever grateful. Emotionally, words cannot do justice to the amount of love and support he has given me. Many-a-time, I have been close to giving up on pursuing graduate school, but he always told me to, “dig deep” and gave me the hope, strength and comfort I needed to keep going. I could not have asked for a better partner-in-crime to experience graduate school and, now, life with. Thank you for being the wind beneath my wings.

CHAPTER 1

INTRODUCTION

1.1 Role of Chromatin Packaging in Cellular Function

Chromatin structure defines the state in which genetic information in the form of DNA is organized within a cell. In eukaryotic somatic cells, DNA is wrapped around proteins known as histones, which make up the basic unit of chromatin. Chromatin helps compact approximately two metres of DNA into a 2 μm nucleus in a way that influences the abilities of genes to be activated or silenced. This packaging is achieved by several tightly regulated mechanisms including chromatin remodeling, histone modifications and histone variant incorporation[1]. In addition, the presence or absence of covalent modifications on DNA (DNA methylation, DNA hydroxymethylation) also helps set up packaging of the genome in the cell. Thus, chromatin packaging allows for cells to maintain distinct identities while containing the same genetic information. Understanding changes in chromatin packaging is the focus of epigenetics, where heritable information lies in more than just the sequence of the DNA such that it helps cells make decisions about quiescence, proliferation, and differentiation by altering genome accessibility to transcriptional machinery. This ultimately affects the transcription repertoire of the cell and ultimately gene function[2]. Failure of proper maintenance of these heritable genetic marks can result in inappropriate activation or inhibition of various signaling pathways and can lead to disease states such as developmental disorders and cancer[3]. Work documented in this dissertation explores the chromatin packaging in mature human sperm, how aberrations in this packaging may impact infertility in humans and in paragangliomas.

1.1.1 Chromatin Packaging in Mature Human Sperm

In mature human spermatozoa, DNA is packaged by histones as well as protamines. During spermatogenesis, the spermatogonial genome is initially packaged by canonical histones – histone 2A (H2A), histone 2B (H3B), histone 3

(H3) and histone 4 (H4), all of which can be methylated, phosphorylated, acetylated and ubiquitinalated. As meiosis proceeds, the spermatocyte genome incorporates testes-specific histone 2B variant (tH2B). Finally, during spermiogenesis, which is the last stage of spermatogenesis, histones become acetylated and are replaced by protamines via transition proteins. The exact mechanism of the histone to protamine replacement is not fully understood[4,5]. Protamines are sperm specific, basic proteins that compact the DNA in toroids and thus condense the DNA 10-fold more in comparison to packaging achieved by histones in somatic cells[6]. This high level of DNA compaction helps transfer the sperm genome to the egg and maintain DNA integrity[7]. Work published by us (Chapter 2) has shown that 96% of the sperm genome is packaged in protamine and only 4% is packaged via histone[8]. An important question in the field that remained unanswered was whether the remaining histone played a biologically significant role in packaging the sperm genome. Since a majority of the genome is packaged by protamine, it was unlikely that the small remnant of histone played an important role in packaging and so this idea was quickly dismissed. However, we pursued this question in our work and specifically asked: are histones programmatically localized at developmental genes or are they simply left behind to randomly package the sperm genome as a consequence of inefficient replacement of histone to protamine during spermiogenesis? Furthermore, if histones are programmatically localized at developmental genes, are these genes important for guiding embryo development? The idea that totipotency (the potential to differentiate into any cell type) and pluripotency (the potential to differentiate into all three germ layers) can be set up as early as in the germ cells (sperm and oocytes) was extremely interesting and hence we focused on understanding chromatin of human sperm.

We queried these possibilities by performing chromatin-immunoprecipitations followed by high throughput sequencing (ChIP-seq) against canonical histones as

well as specific histone modifications including H3K4me3, H3K4me2, H3K27me3 and testes-specific H2B. To summarize our findings, we observed that nucleosomes appeared to be programmatically localized as they were significantly enriched at promoters of developmental transcription and signaling factors including the *HOX* gene clusters, miRNA clusters and imprinted gene clusters. More specifically, H3K4me2, a histone modification associated with active transcription was enriched at promoters of developmental genes. H3K4me3, another modification also linked with active transcription, was associated with a subset of promoters of developmental genes such as *HOX* clusters, certain noncoding RNAs, paternally expressed imprinted loci as well as genes involved in spermatogenesis. H3K27me3, a histone modification associated with transcriptional repression, was found at the promoters of genes that are bivalent in ES cells and developmental genes that are repressed in early embryo development. Testis-specific H2B was enriched at promoters of genes involved in spermatogenesis. Both H3K4me2 and me3, but not H3K27me3, were associated with genes that are expressed during the 4-8 cell stage in human embryos[9]. We also performed DNA methylation (DNAme) studies and found that promoters of genes involved in embryo development were generally hypomethylated. DNA hypomethylation at promoters of genes is usually associated with active transcription. The findings that genes involved in cell cycle and early development were poised with the presence of histone modifications at their promoters in human sperm were also confirmed by other groups[10,11]. Moreover, this observation was also confirmed in mouse and zebrafish sperm[12] strengthening the possibility that male germ cells are indeed poised for totipotency by the presence of histone modifications and the lack of DNA methylation at genes involved in early embryonic development.

1.1.2 Does the Nature of Infertility Reside Partly in Chromatin?

Since histone modifications and DNA methylation appear to poise the paternal genome at genes involved in early embryonic development, the clearest place a chromatin problem can manifest is infertilty. Hence, we asked an intriguing question: Can infertile men have methylation defects, along with defects in the levels and localization of histone modifications, in their sperm? Previous studies have shown that a small percentage of offspring conceived through assisted reproductive technologies (ART) have an increased risk of developing imprinting disorders. These abnormalities included Beckwith-Wiedemann syndrome, transient neonatal diabetes, Silver-Russell syndrome, Angelman and Prader-Wili syndrome[13,14]. Is the risk of developing these disorders due to using gametes with aberrant DNA methylation at their imprinted loci or due to the ART procedure itself where gametes are manipulated (by hormone stimulation followed by freezing and thawing) and the embryo is cultured *in vitro*?

Recent studies in infertile patients reported aberrant changes in DNA methylation at imprinted loci[13,14]. These patients had either low sperm count (oligozoospermic) or no detectable sperm in their ejaculate (azoospermic). In our study, we focused on two classes of infertile patients: ones that had low sperm count (oligozoospermic) and ones that had abnormal histone to protamine exchange during spermiogenesis. Mature sperm has two sperm-specific nuclear proteins, Protamine 1 and 2 that ensure proper condensation of sperm chromatin. Due to improper exchange of histones to protamines, the ratio of P1:P2 is significantly altered[15,16]. We were interested in evaluateing the relationship between patients with altered chromatin condensation (abnormal P1:P2 ratios) to aberrations in DNA methylation. While DNA methylation defects could occur at both imprinted and nonimprinted regions, we first chose to focus on imprinted regions. At this point, next generation sequencing technologies had not been developed and hence we

were unable to perform high-throughput sequencing on a genome-wide level. Hence, we focused on seven imprinted loci that had been previously associated with imprinting disorders – *H19, IGF2, LIT1, MEST, SNRPN, PLAGL1,* and *PEG3*. Our results were published in the journal of Fertility and Sterility and have been presented in Chapter 3 of this dissertation. In summary, we observed a statistically significant increase in methylation at two of the seven imprinted loci in both infertile patient populations. It is important to realize that while our study did not provide a causal link for the trans-generational inheritance of DNA methylation defects leading to imprinting diseases, it showed a strong correlation between infertility in males and aberrations of DNA methylation at select imprinted loci.

1.1.3 Epigenetics and Cancer

Aberrant changes in DNA methylation have been implicated in the initiation and progression of cancers. In most cancers, genome-wide hypomethylation is observed especially at repetitive elements, retrotransposons, CpG poor promoters and gene deserts. This can lead to an increase in genomic instability. Similarly, DNA hypomethylation can lead to the activation of proto-oncogenes, which may provide a growth and survival advantage to the cells. In addition to this genome-wide DNA hypomethylation, site-specific hypermethylation may contribute to the progression of tumorigenesis by silencing tumor suppressor genes such as *p16, MLH1 and BRCA1*[17,18]. These genes may be involved in an array of cellular processes such as DNA repair, cell adhesion, apoptosis and angiogenesis. Shutting down of these processes may promote cancer initiation and progression. Apart from directly silencing tumor suppressor genes, DNA hypermethylation may silence transcription factors such as *RUNX3* (esophageal cancer), *GATA-4* (colorectal cancer) and *GATA-5* (gastric cancer). This may silence further downstream targets, which may enable the cells to accumulate further lesions leading to the rapid progression

of cancer[19]. This leads us to the idea of the “CpG island methylator phenotype” (CIMP) where large stretches of CpG islands in promoters of candidate genes are aberrantly hypermethylated in cancers. While this aberrant methylation at some genes may help with cancer progression, currently, candidate genes used to study the CIMP (CpG island methylator phenotype) phenomena vary from one cancer type to another. A different list of genes susceptible to gaining DNA methylation at their CpG Island promoters can be generated for each cancer. Hence, this increases the challenge and the burden on how to classify a certain tumor type, paragangliomas in our case, as a CIMP-positive or CIMP-negative phenotype. We do not abide by this nomenclature and take an unbiased approach of identifying all genes that gain or lose DNA methylation in their promoter regions that may fall in CpG Islands, CpG Shores (2 kb upstream or downstream of a CpG Island), CpG Shelves (4 kb upstream or downstream of a CpG Island) and distant regions (not affiliated with a CpG Island) in paragangliomas. We chose this approach since epigenetic misregulation is a dynamic process, and it is most likely that a combination of regions that gain or lose DNA methylation act in concert with histone modifications to give the cell a growth and survival advantage, leading to tumorigenesis. Several studies have looked at the global loss of histone modifications that usually mark genes for transcriptional activation (such as H4K16Ac and H4K20me3), and hence can lead to aberrant gene repression. This process is usually mediated by histone deacetylases (HDACs), which are often found overexpressed in various cancers. HMTs (histone methyltransferases) can also either be overexpressed in cancers or histone demethylases (such as those belonging to the Jumonji family) may be downregulated or their activities can be misregulated such that H3K9me3, H3K27me3 and H3K4me3 levels change further leading to abnormal silencing or activation of genes, respectively [3]. Paragangliomas are of particular interest since mutations in the succinate dehydrogenase complex causes accumulation

of a metabolite (succinate) which can directly impact the activity of many histone demethylates and TET enzymes. Understanding the direct connection between metabolism, epigenetic misregulation, in particular DNA methylation and cancer was a large focus of our interest and dissertation work.

1.2 Paragangliomas (PGLs)

Paragangliomas are rare, highly vascularized, extra-adrenal tumors that are associated with both the parasympathetic nervous system and sympathetic nervous system. They can occur at multiple locations along the paravertebral axis as shown in Figure 1.1[20]. When PGLs occur at the glomus (near the middle ear) or near the carotid body tumor, they are associated with the parasympathetic nervous system and do not secrete catecholamines. Tumors situated along the sympathetic trunk in the abdomen and pelvic regions usually produce catecholamines. Paragangliomas originate from neuroendocrine tissue chromaffin cells that are derived from the neural crest cells[21,22].

Paragangliomas can occur sporadically or can be inherited. Hereditary syndromes known to be associated with development of paragangliomas are multiple endocrine neoplasia (MEN) 2A and 2B, von Hippel-Lindau disease (VHL) and neurofibromatosis type 1 (NF1) caused, respectively, by germline-mutations of the *RET* proto-oncogene and the *VHL* and *NF1* tumor suppressor genes. However, germline mutations in the components of succinate dehydrogenase complex (*SDHA, SDHB, SDHC and SDHD)* occur most commonly in hereditary paragangliomas [23]. The succinate dehydrogenase genes *SDHD* (PGL1), *SDHC* (PGL3), and *SDHB* (PGL4), SDHAF2 (PGL2) appear to function as tumor suppressor genes whereupon loss of the somatic wildtype allele, the enzyme is rendered inactive, thus having severe implications on the Krebs cycle and subsequently levels of metabolites in the cell[21,22]. As a consequence, succinate accumulates in the mitochondria, gets

transported to the cytoplasm and inhibits several α-KG depedent enzymes which include jumonji-histone demethylases (JHDM), ten-eleven translocases (TETs) and prolyl hydroxylases (PHDs) [24-26]. Several previously published studies have reported that inhibiting these enzymes may alter the epigenetic profile of cells, particularly bulk levels of histone modifications, DNA methylation and active DNA demethylation *in vitro* and in cell culture systems. While all these aspects could contribute to neoplastic formation, this misregulation is not fully understood in the neural-crest cell derived tumor context.

1.2.1 Succinate Dehydrogenase Complex (SDH) and Krebs Cycle

The succinate dehydrogenase complex (also known as succinate ubiquinone oxydoreductase or mitochondrial complex II) is a highly conserved heterotetramer. It is anchored within the inner mitochondrial membrane via its two hydrophobic subunits, SDHC and SDHD, where they bind ubiquinone and transfer electrons to the ubiquinone pool as part of the electron-transport chain. The SDH complex also contacts the mitochondrial matrix where its other two hydrophilic domains, SDHA (a flavoprotein) and SDHB (an iron-sulfur protein) together form the catalytic core that are involved in the Kreb's cycle where they oxidize succinate to fumarate. Finally, SDHAF2 (SDH assembly factor 2 or SDH5) is involved in flavination of SDHA. [21,27-29]. (Figure 1.2 [30])

Mutations in different subunits of the SDH complex can cause different disorders. Mutations in SDHA can cause Leigh's syndrome, a rare, neurometabolic disorder that is characterized by degeneration of the central nervous system. Germline heterozygous mutations in SDHA, B, C, D or AF2 subunits of the SDH complex can cause an array of tumors such as paragangliomas, pheochromocytomas (adrenal neuroendocrinal tumors), gastrointestinal stromal tumors (GISTs), renal cell carcinomas, renal oncocytomas, and, rarely, papillary thyroid carcinomas,

neuroblastomas, and seminomas [29,31].

1.3 Metabolic Regulation of Epigenetics

A recently emerging important concept is that a cell's metabolic state can also regulate the cell's epigenetics and transcription and understanding this link can help shed light on the progression of many diseases including cancer. It is known that cells can adjust their metabolic state depending on nutrient ability and extracellular responses. For example, cells that are nonproliferating and differentiated depend on the efficient production of ATP through oxidative phosphorylation (OXPHOS). Oxidative phosphorylation involves the uptake of glucose, which is converted to pyruvate through glycolysis, followed by the complete oxidation of pyruvate to CO_2 in the Kreb's TCA cycle in the mitochondria. This is coupled to respiratory chain activity where oxygen is the final acceptor in an electron transport chain that generates an electrochemical gradient, facilitating ~36 moles of ATP produced per one mole of glucose. In proliferating cells and cancer cells, even in the presence of oxygen, glucose uptake is increased and is metabolized by aerobic glycolysis. During aerobic glycolysis, glucose is converted to pyruvate and then pyruvate is reduced by lactate dehydrogenase in the cytoplasm, resulting in the secretion of lactate yielding only ~4 moles of ATP per mole of glucose. This is known as the Warburg effect and it has been of great interest to understand why proliferating cells and cancer cells would choose aerobic glycolysis versus oxidative phosphorylation as a mode for ATP production. While this phenomenon is not completely understood, one possible explanation for this occurrence is that proliferating and cancer cells keep up with their increased demand of fatty acid generation, lipid generation, protein and nucleic acid synthesis by increasing their nutrient and glucose uptake and metabolizing it via aerobic glycolysis as this provides an energetically favorable state for cells to continue proliferating.

These cells do not experience nutrient deprivation and hence it is also possible that under such conditions inefficient generation of ATP is not a problem. As most chromatin modifying enzymes require substrates or cofactors that are metabolic intermediates, it is now being shown that changes in metabolite levels modulates the activities of these enzymes and therefore impacts chromatin dynamics. [2,26,32-34]

There are several instances where there is crosstalk between metabolism and epigenetics. For example, as glucose enters the glycolytic pathway, a small proportion is utilized by the hexosamine biosynthetic pathway to produce O-linked N-acetylglucosamine (GlcNAc) that is the substrate for histone H2B GlcNAcylation by the enzyme O-GlcNAc transferase (OGT). Flux through glycolysis determines the NAD+/ NADH ratio, which is crucial for the activities of Sirtuin histone deacetylases. Several TCA cycle intermediates can be exported out of mitochondria including citrate and alpha-ketoglutarate (α-KG). Cytosolic citrate is converted to acetyl-CoA, which is used as a donor for histone acetyltransferase-mediated histone acetylation. αKG is used as cofactor for Jumonji-histone demethylases (JHMD) and DNA demethylases (TETs). The substrate for HMT and DNMT is SAM, which is synthesized from essential amino acid methionine. Finally, a low ATP/AMP ratio can activate AMPK, a kinase that phosphorylates histones [2].

1.4 Epigenetic Misregulation in Tumors with Mutations in Kreb's Cycle Enzymes

Paragangliomas that have a mutation in the SDH complex are unable to convert succinate to fumarate. This leads to succinate accumulation in the cell which can competitively inhibit many alpha-ketoglutarate (αKG) dependent dioxygenases[25,35]. αKG is a necessary cofactor for dioxygenase enzymes such as proline hydrolases (PHD), Jumonji histone demethylases (JHMD) and DNA demethylases (TET; ten-eleven translocases). These enzymes use a ferrous ion

and molecular oxygen as cofactors along with alpha-ketoglutarate to hydroxylate their substrates and generate succinate as a product (Figure 1.3)[36]. Thus, if there is an accumulation of succinate in the cell, it can inhibit this reaction competitively and in turn have serious consequences on the epigenome of a cell [24,37].

Succinate accumulation can have an impact on two axes: a hypoxia depedent axis and an epigenetic axis. Under normal conditions, PHDs hydroxylate the two prolyl residues in the oxygen dependent domain (ODD) of HIF (hypoxia inducible factor 1α). This allows VHL,an E3 ubiquitin ligase, to bind the ODD and thus destabilize HIF, which subsequently gets targeted for degradation. Due to a mutation in the SDH complex, the rate of succinate oxidation to fumarate is altered which leads to succinate accumulation in the mitochondria. This then gets transported to the cytosol. Succinate can inhibit PHDs from hydroxylating HIF. VHL is unable to target HIF for degradation allowing for stabilized HIF to dimerize with its partner and get translocated into the nucleus. Here it will transcriptionally upregulate its downstream targets that contribute to angiogenesis and metastasis[25,38.] HIF is also involved in the upregulation of glucose transport and glycolysis, implicating it as a major regulator of the Warburg effect [39,40]. The second axis affected by succinate accumulation is the epigenetic axis, where the activity of important αKG depedent epigenetic enzymes such as histone demethylases in the Jumonji class, and DNA demethylases (TETs) may be inhibited. This can lead to epigenetic changes in paragangliomas where there may be a misregulation of histone modifications and a gain of DNA methylation. Furthermore, this misregulation may impact the transcriptome in a manner that helps drive tumor initiation or growth, perhaps through the improper silencing of tumor suppressor proteins: lack of DNA demethylation could impose/maintain silencing (Figure 1.4)[2,24,25,35,37,41-43].

Recently, another study[44] was the first to perform a restricted methylome analysis on a large cohort of paragangliomas harboring mutations in RET, NF1,

VHL and SDHx genes. They found that SDH deficient tumors gain DNA methylation at genes and very few of these changes correlated with downregulation of gene expression. Genes involved in neuroendocrine differentiation and catecholamine metabolism were mainly affected. They also found a single gene involved in EMT (epithelial-to-mesenchymal transition) and a single tumor suppressor to be affected. While some of their findings were interesting, their analysis was limited to establishing changes in DNA methylation between the tumor subtypes, as opposed to comparing to a control or a progenitor cell type. This highlighted changes that exist between tumorigenic cells with different genetic backgrounds and did not necessarily determine gene candidates that may be involved in tumor initiation and progression. Also, subtle but statistically significant changes in DNA methylation entailed a majority of the reported changes that may or may not have a biologically significant impact on the cell's transcriptome and hence their contribution to tumorigenesis may be limited. Finally, they limited their analysis to changes of DNA methylation in CpG Islands, where as several studies have shown that changes in DNA methylation at CpG Shores (approximately 2kb upstream or downstream from a CpG Island) may also play a significant role in modulating genes that may contribute towards tumorigenesis[45,46].

Related to SDH deficient PGLs, gliomas have mutations in the TCA cycle enzyme, isocitrate dehydrogenase (IDH). IDH catalyzes the oxidative decarboxylation of isocitrate, producing alpha-ketoglutarate. However, mutant IDH loses its normal catalytic activity and instead gains the function of producing an onco-metabolite, 2-hydroxyglutarate (2-HG), which acts as a competitive inhibitor of the αKG-dependent dioxygenases[41]. A recent study profiled changes in DNA methylation in IDH mutant and wildtype gliomas, correlated these changes to gene expression and then further demonstrated that IDH mutations were sufficient to establish these DNA methylation changes and transcriptome changes in

immortalized primary human astrocytes[47]. Further, another publication reported that IDH mutations impair histone demethylation blocking the differentiation of lineage-specific progenitor cells into terminally differentiated cells. They demonstrated bulk gain of several histone modifications in their adipocyte cells expressing IDH mutants and showed enrichment of these modifications at promoters of genes involved in adipocyte differentiation followed by transcriptional repression[42].

Extending the link between the Krebs cycle, epigenomic changes and cancer, another study highlighted divergent global changes in DNA methylation in gastrointestinal stromal tumors (GISTs) that harbored mutations in SDH genes or in other non-SDH related susceptibility genes[48]. They validated their link between SDH mutations and methyl-divergence in tumorigenesis by comparing SDH-deficient, hereditary paragangliomas to adrenal medulla as reference tissue and IDH mutant gliomas to normal glial tissue. Globally, by principal component analysis, they found that the Kreb cycle mutant tumors were more closely related to each other than nonmutant tumors. They also found similar targets that were hypermethylated and hypomethylated in the Krebs cycle mutant tumors. While this study identifies epigenomic homology of tumors from divergent developmental lineages having mutations in related Krebs cycle enzymes, it did not establish a link between DNA methylation changes and gene expression. From previously published reports[44] and our study, only a small fraction (~10-20%) of regions that have differential DNA methylation actually correlate with changes in transcription. Hence, in the Killian *et al.* study the interpretation of the biological contribution of the reported DNA methylation changes in the tumorigenesis of Kreb-cycle mutant tumors is limited.

1.5 Dissertation Overview

Overall, our goal was to study the role of the epigenome in a normal, developmental context (mature human sperm) and in two perturbed systems (infertility and cancer) with a specific focus on aberrations in DNA methylation at imprinted genes and genes involved in oncogenesis.

Chapters 2 and 3 highlight work that I was involved in during the earlier part of my career in graduate school. Chapter 2 focuses on work in understanding if distinct chromatin marks in mature human sperm were poised at developmentally important genes and if this poising could contribute towards early embryonic development. Chapter 3 focuses on aberrant DNA methylation patterns at imprinted genes in infertile patients.

To gain insight on growth and development genes that will impact PGL oncogenesis, we compared the methylomes and transcriptomes of SDH deficient PGLs to a progenitor cell type, neural crest cells (NCCs). NCCs are multipotent by nature and can differentiate into several lineages including peripheral neurons, glia, melanocytes, endocrine cells, chromaffin cells and mesenchymal precursor cells [49,50]. In addition to our SDH deficient (SDHx) PGLs, we have a subclass of PGLs, that may be inherited or sporadic, that do not harbor mutations in any of the reported susceptibility genes (SDHx, VHL, RET, NF1, TMEM127, MAX, EPAS1) and hence will be classified as SDH Present PGLs in this study. This subclass of SDH Present PGLs is of great interest and our work reveals that they are transcriptionally strikingly similar to the SDHx PGLs. Several reports in the literature have confirmed that SDH deficient tumors are transcriptionally more similar to tumors with mutations in VHL and EPAS1; whereas tumors with mutations in RET, NF1, MAX and TMEM127 are related to each other more closely[51]. Sporadic tumors can be transcriptionally related to either group; however, the reasons for why this is are poorly understood. The unique aspect of our work is to understand

more about the connection between SDH deficient and SDH Present PGLs by performing whole exome sequencing, to determine possible, common mutations in gene candidates in shared pathways, and profile common epigenetic changes in DNA methylation and transcription compared to a progenitor cell type, NCCs. Our findings have been documented in Chapter 4, which is currently a manuscript in preparation to be submitted for publication.

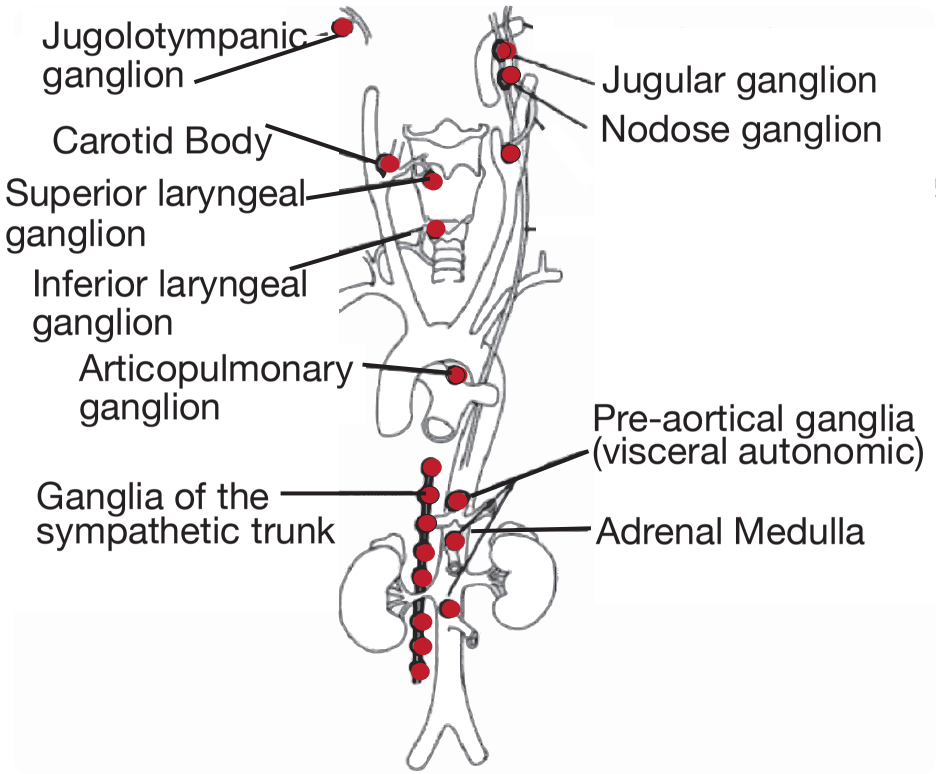


Figure 1.1: Anatomical distribution of paraganglia along the paravertebral axis. Modified and reprinted with permission from Bioscientifica Ltd. Welander J, Soderkvist P, Gimm O (2011) Genetics and clinical characteristics of hereditary pheochromocytomas and paragangliomas. Endocrine Related Cancer 18: R253-R276, copyright 2011.

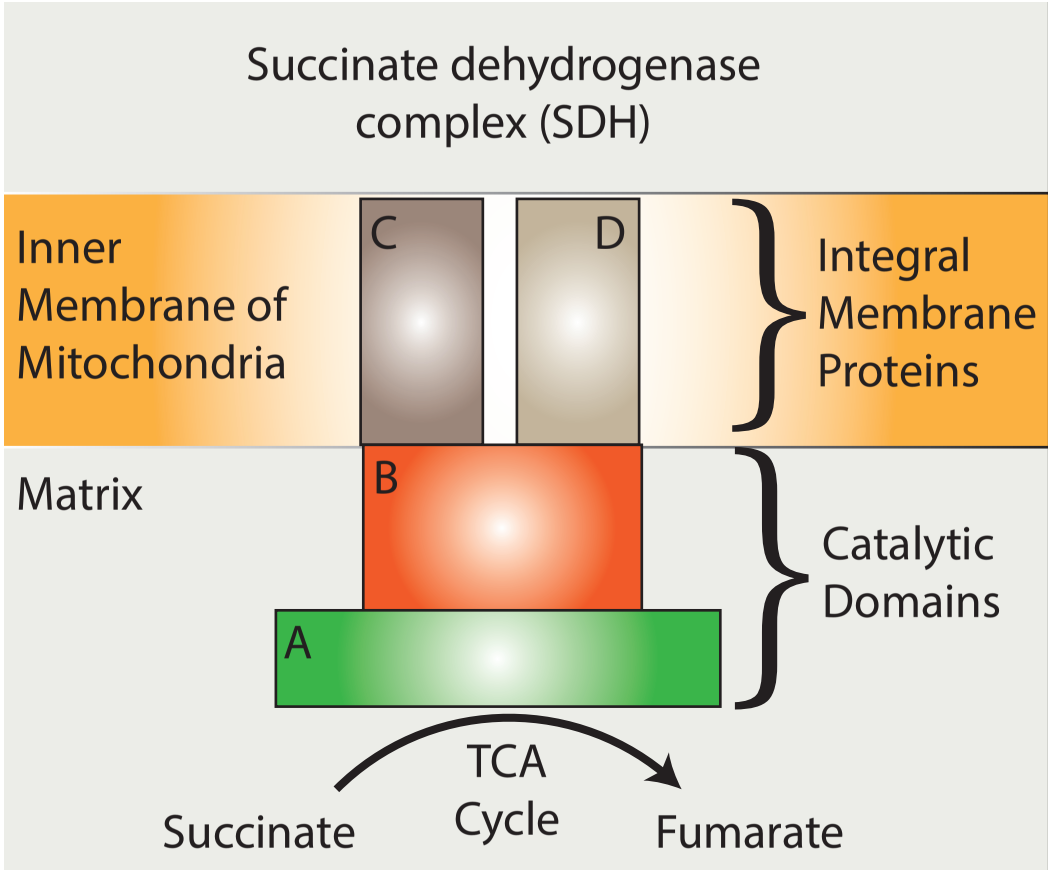


Figure 1.2: Mitochondrial succinate dehydrogenase complex and tricarboxylic acid cycle. Modified and reprinted with permission from Nature Publishing Group. Gottlieb E, Tomlinson IPM (2005) Mitochondrial tumour suppressors: a genetic and biochemical update. Nature Reviews Cancer 5: 857-866, copyright 2005.

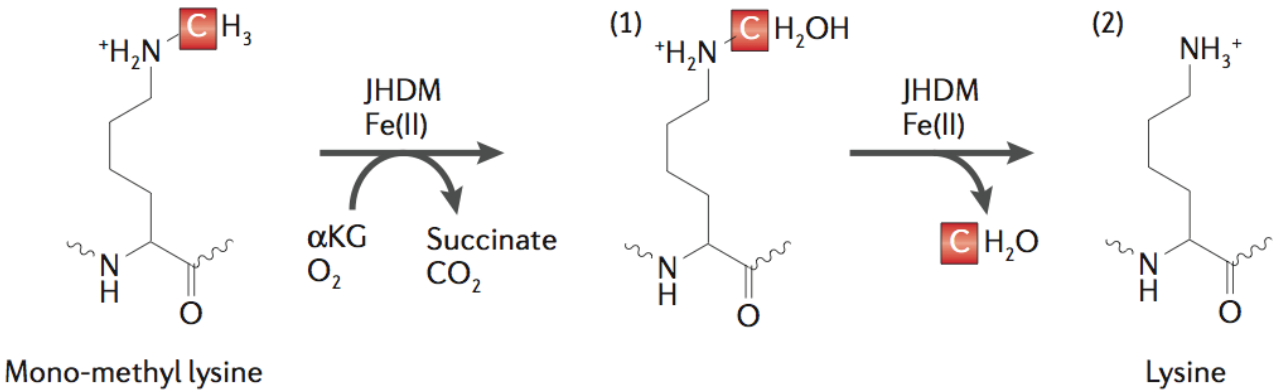
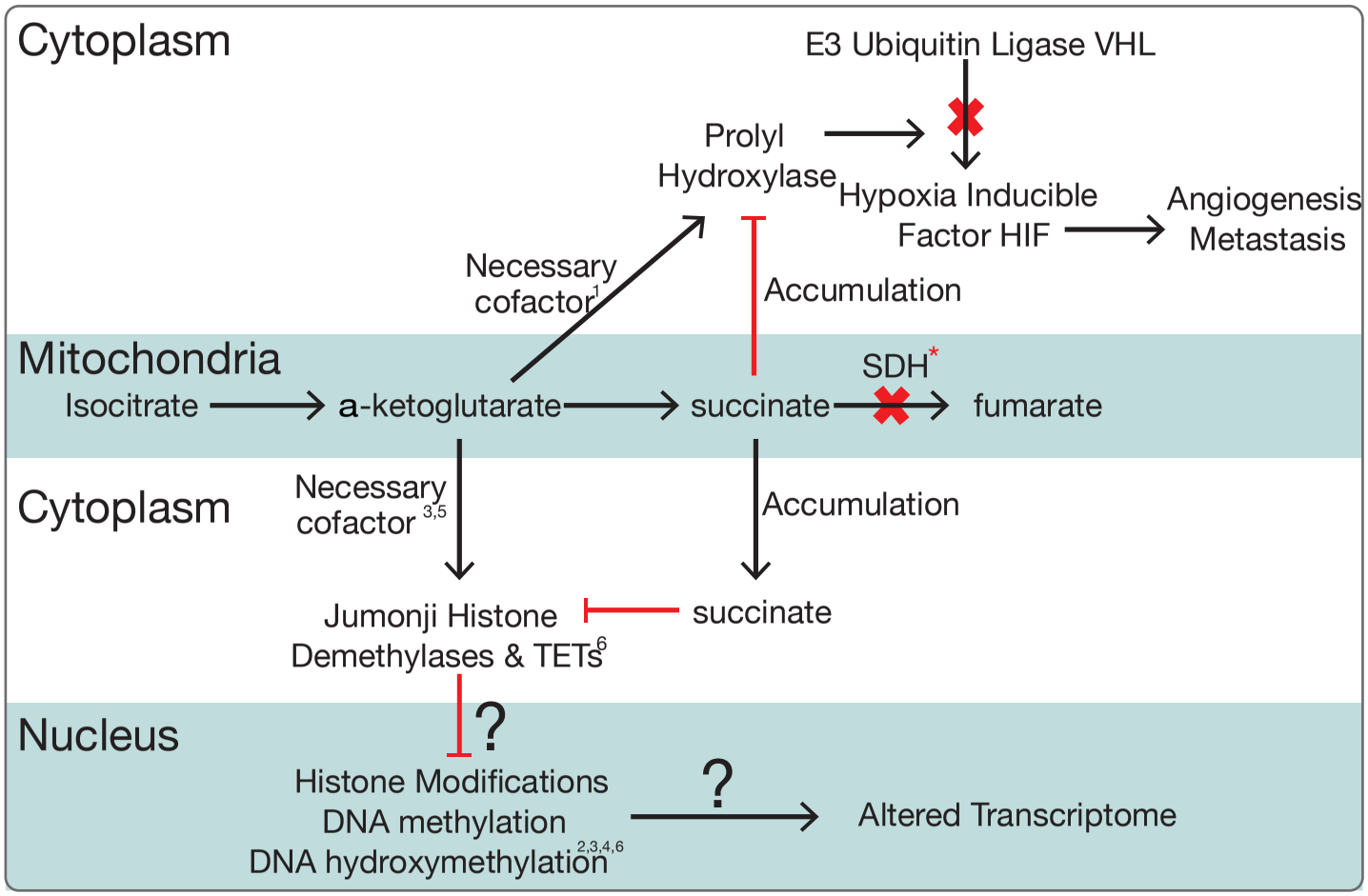


Figure 1.3: Chemical mechanism used by alpha-ketoglutarate depedent Jumonji Histone Demethylases (JHDM). Modified and reprinted with permission from Nature Publishing Group. Klose RJ, Kallin EM, Zhang Y (2006) JmjC-domain-containing proteins and histone demethylation. Nature Reviews Genetics 7: 715-727, copyright 2006.



1 Selak et al, (2005) Cancer cell, 2 Cervera et al (2009) Molecular Cancer, 3 Smith et al (2007) Human Molecular Genetics, 4 Turcan et al, (2012) Nature, 5 Xu et al (2011) Cancer Cell, 6 Xiao et al (2012) Genes and Development

Figure 1.4: Proposed model containing 2 axes: hypoxia and epigenetics in SDH deficient Paragangliomas.

1.6 References

1. Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. Cell 128: 707-719.

2. Lu C, Thompson CB (2012) Metabolic regulation of epigenetics. Cell Metabolism 16: 9-17.

3. Sharma S, Kelly TK, Jones PA (2010) Epigenetics in cancer. Carcinogenesis 31: 27-36.

4. Carrell DT, Hammoud SS (2009) The human sperm epigenome and its potential role in embryonic development. Molecular Human Reproduction 16: 37-47.

5. Ward WS, Coffey DS (1991) DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells. Biology of Reproduction 44: 569-574.

6. Balhorn R, Brewer L, Corzett M (2000) DNA condensation by protamine and arginine-rich peptides: analysis of toroid stability using single DNA molecules. Molecular Reproduction and Development 56: 230-234.

7. Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, et al. (2009) Distinctive chromatin in human sperm packages genes for embryo development. Nature 460: 473-478.

8. Li SS, Liu YH, Tseng CN, Singh S (2006) Analysis of gene expression in single human oocytes and preimplantation embryos. Biochemical Biophysical Research Communications 340: 48-53.

9. Brykczynska U, Hisano M, Erkek S, Ramos L, Oakeley EJ, et al. (2010) Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nat Struct Mol Biol 17: 679-687.

10. Arpanahi A, Brinkworth M, Iles D, Krawetz SA, Paradowska A, et al. (2009) Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. Genome Research 19: 1338-1349.

11. Wu SF, Zhang H, Cairns BR (2011) Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. Genome Research 21: 578-589.

12. Kobayashi H, Sato A, Otsu E, Hiura H, Tomatsu C, et al. (2007) Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. Human Molecular Genetics 16: 2542-2551.

13. Marques CJ, Costa P, Vaz B, Carvalho F, Fernandes S, et al. (2007) Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. Molecular Human Reproduction 14: 67-74.

14. Aoki VW, Emery BR, Liu L, Carrell DT (2006) Protamine levels vary between individual sperm cells of infertile human males and correlate with viability and DNA integrity. Journal of Andrology 27: 890-898.

15. Balhorn R, Reed S, Tanphaichitr N (1988) Aberrant protamine 1/protamine 2 ratios in sperm of infertile human males. Cellular and Molecular Life Sciences 44: 52-55.

16. Jones PA, Baylin SB (2007) The epigenomics of cancer. Cell 128: 683-692.

17. Baylin SB (2005) DNA methylation and gene silencing in cancer. Nature Clinical Practice Oncology 2: S4-S11.

18. Akiyama Y, Watkins N, Suzuki H, Jair KW, van Engeland M, et al. (2003) GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer. Molecular and Cellular Biology 23: 8429-8439.

19. Welander J, Soderkvist P, Gimm O (2011) Genetics and clinical characteristics of hereditary pheochromocytomas and paragangliomas. Endocrine Related Cancer 18: R253-R276.

20. Bardella C, Pollard PJ, Tomlinson I (2011) SDH mutations in cancer. Biochim Biophys Acta 1807: 1432-1443.

21. van Nederveen FH, Gaal J, Favier J, Korpershoek E, Oldenburg RA, et al. (2009) An immunohistochemical procedure to detect patients with paraganglioma and phaeochromocytoma with germline SDHB, SDHC, or SDHD gene mutations: a retrospective and prospective analysis. Lancet Oncology 10: 764-771.

22. Hensen EF, Bayley JP (2011) Recent advances in the genetics of SDH-related paraganglioma and pheochromocytoma. Fam Cancer 10: 355-363.

23. Smith EH, Janknecht R, Maher LJ (2007) Succinate inhibition of alpha-ketoglutarate-dependent enzymes in a yeast model of paraganglioma. Human Molecular Genetics 16: 3136-3148.

24. Xiao M, Yang H, Xu W, Ma S, Lin H, et al. (2012) Inhibition of alpha-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. Genes Dev 26: 1326-1338.

25. Ward PS, Thompson CB (2012) Metabolic reprogramming: a cancer hallmark

even warburg did not anticipate. Cancer Cell 21: 297-308.

26. Eng C, Kiuru M, Fernandez MJ, Aaltonen LA (2003) A role for mitochondrial enzymes in inherited neoplasia and beyond. Nature Reviews Cancer 3: 193-202.

27. Xekouki P, Stratakis CA (2012) Succinate dehydrogenase (SDHx) mutations in pituitary tumors: could this be a new role for mitochondrial complex II and/or Krebs cycle defects? Endocr Relat Cancer 19: C33-40.

28. Kantorovich V, Pacak K (2010) Pheochromocytoma and paraganglioma. Progress in Brain Research 182: 343-373.

29. Gottlieb E, Tomlinson IPM (2005) Mitochondrial tumour suppressors: a genetic and biochemical update. Nature Reviews Cancer 5: 857-866.

30. Gaal J, Stratakis CA, Carney JA, Ball ER, Korpershoek E, et al. (2011) SDHB immunohistochemistry: a useful tool in the diagnosis of Carney-Stratakis and Carney triad gastrointestinal stromal tumors. Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc 24: 147-151.

31. Vander Heiden MG, Cantley LC, Thompson CB (2009) Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science 324: 1029-1033.

32. Hsu PP, Sabatini DM (2008) Cancer cell metabolism: Warburg and beyond. Cell 134: 703-707.

33. Vazquez A, Liu J, Zhou Y, Oltvai ZN (2010) Catabolic efficiency of aerobic glycolysis: the Warburg effect revisited. BMC Syst Biol 4: 58.

34. Selak MA, Armour SM, MacKenzie ED, Boulahbel H, Watson DG, et al. (2005) Succinate links TCA cycle dysfunction to oncogenesis by inhibiting HIF-alpha prolyl hydroxylase. Cancer Cell 7: 77-85.

35. Klose RJ, Kallin EM, Zhang Y (2006) JmjC-domain-containing proteins and histone demethylation. Nature Reviews Genetics 7: 715-727.

36. Hoffmann I, Roatsch M, Schmitt ML, Carlino L, Pippel M, et al. (2012) The role of histone demethylases in cancer therapy. Mol Oncol 6: 683-703.

37. Gilkes DM, Xiang L, Lee SJ, Chaturvedi P, Hubbi ME, et al. (2014) Hypoxia-inducible factors mediate coordinated RhoA-ROCK1 expression and signaling in breast cancer cells. Proceedings of the National Academy of Sciences of the United States of America 111: E384-393.

38. Robey IF, Lien AD, Welsh SJ, Baggett BK, Gillies RJ (2005) Hypoxia-inducible factor-1alpha and the glycolytic phenotype in tumors. Neoplasia 7: 324-330.

39. Semenza GL (2007) HIF-1 mediates the Warburg effect in clear cell renal carcinoma. Journal of Bioenergetics and Biomembranes 39: 231-234.

40. Xu W, Yang H, Liu Y, Yang Y, Wang P, et al. (2011) Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. Cancer Cell 19: 17-30.

41. Lu C, Ward PS, Kapoor GS, Rohle D, Turcan S, et al. (2012) IDH mutation impairs histone demethylation and results in a block to cell differentiation. Nature 483: 1-7.

42. Cervera AM, Bayley J-P, Devilee P, McCreath KJ (2009) Inhibition of succinate dehydrogenase dysregulates histone modification in mammalian cells. Molecular Cancer 8: 89.

43. Letouze E, Martinelli C, Loriot C, Burnichon N, Abermil N, et al. (2013) SDH mutations establish a hypermethylator phenotype in paraganglioma. Cancer Cell 23: 739-752.

44. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nature Genetics 41: 178-186.

45. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics 13: 484-492.

46. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, et al. (2012) IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. Nature 483: 479-483.

47. Killian JK, Kim SY, Miettinen M, Smith C, Merino M, et al. (2013) Succinate dehydrogenase mutation underlies global epigenomic divergence in gastrointestinal stromal tumor. Cancer Discov 3: 648-657.

48. Lee G, Chambers SM, Tomishima MJ, Studer L (2010) Derivation of neural crest cells from human pluripotent stem cells. Nature Protocols 5: 688-701.

49. Knecht AK, Bronner-Fraser M (2002) Induction of the neural crest: a multigene process. Nature Reviews Genetics 3: 453-461.

50. Burnichon N, Vescovo L, Amar L, Libe R, de Reynies A, et al. (2011) Integrative genomic analysis reveals somatic mutations in pheochromocytoma and paraganglioma. Human Molecular Genetics 20: 3974-3985.

CHAPTER 2

DISTINCTIVE CHROMATIN IN HUMAN SPERM PACKAGES GENES FOR EMBRYO DEVELOPMENT

Reprinted with permission from Nature. Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, et al. (2009) Distinctive chromatin in human sperm packages genes for embryo development. Nature 460: 473-478.

Chapter 2 is a published article. My contribution to this work involved identifying the presence and subsequently quantifying histone modification levels in human sperm and performing targeted bisulfite sequencing on sperm DNA.

Vol 460|23 July 2009|doi:10.1038/nature08162 nature

ARTICLES

Distinctive chromatin in human sperm packages genes for embryo development

Saher Sue Hammoud[1,2], David A. Nix[3], Haiying Zhang[1], Jahnvi Purwar[1], Douglas T. Carrell[2] & Bradley R. Cairns[1]

Because nucleosomes are widely replaced by protamine in mature human sperm, the epigenetic contributions of sperm chromatin to embryo development have been considered highly limited. Here we show that the retained nucleosomes are significantly enriched at loci of developmental importance, including imprinted gene clusters, microRNA clusters, *HOX* gene clusters, and the promoters of stand-alone developmental transcription and signalling factors. Notably, histone modifications localize to particular developmental loci. Dimethylated lysine 4 on histone H3 (H3K4me2) is enriched at certain developmental promoters, whereas large blocks of H3K4me3 localize to a subset of developmental promoters, regions in *HOX* clusters, certain noncoding RNAs, and generally to paternally expressed imprinted loci, but not paternally repressed loci. Notably, trimethylated H3K27 (H3K27me3) is significantly enriched at developmental promoters that are repressed in early embryos, including many bivalent (H3K4me3/H3K27me3) promoters in embryonic stem cells. Furthermore, developmental promoters are generally DNA hypomethylated in sperm, but acquire methylation during differentiation. Taken together, epigenetic marking in sperm is extensive, and correlated with developmental regulators.

During spermiogenesis canonical histones are largely exchanged for protamines[1,2], small basic proteins that form tightly packed DNA structures important for normal sperm function[3]. We find about 4% of the haploid genome retained in nucleosomes (Supplementary Fig. 1a). The rare retained nucleosomes in sperm consist of either canonical or histone variant proteins, including a testes-specific histone H2B (TH2B) with an unknown specialized function[4,5]. Their presence may simply be due to inefficient protamine replacement, leading to a low random distribution genome-wide with no impact in the embryo. Alternatively, these retained nucleosomes, along with attendant modifications, might be enriched at particular genes/loci. This latter possibility would raise the possibility for programmatic retention for an epigenetic function in the embryo. To address these questions, we localized the nucleosomes retained in mature sperm from fertile donors using high-resolution genomic approaches.

Developmental loci bear nucleosomes

To address donor variability, we examined nucleosome retention in a single donor (D1) and/or a pool of four donors (donor pool). Sperm chromatin was separated into protamine-bound and histone-bound fractions. In brief, mononucleosomes were isolated (>95% yield) by sequential MNase digestion and sedimentation (Supplementary Fig. 1b–e). This mononucleosome pool was used for chromatin immunoprecipitation (ChIP; to select modified nucleosomes), or the DNA was isolated from the mononucleosome pool to represent all nucleosomes. Purified DNA was subjected to high-throughput sequencing (Illumina GAII), or alternatively, was labelled and hybridized to a high-density promoter-tiling array (9 kilobase (kb) tiled; Supplementary Fig. 2, schematic).

Our initial array approach examined three replicas of D1 (pairwise average $R^2 = 0.85$). Notably, Gene Ontology analysis revealed nucleosomes significantly enriched at promoters that guide embryonic development—primarily developmental transcription factors and signalling molecules (Gene Ontology term false discovery rate (FDR) < 0.01; Box 1 and Supplementary Table 1; for all extended Gene Ontology categories see Supplementary Tables and Supplementary Data Set 1). To conduct genome-wide profiling, we performed high-throughput sequencing of nucleosomes from D1 or the donor pool. Regions significantly enriched for histone relative to the input control (sheared total sperm DNA) were identified using a 300-base-pair (bp) window metric[6]. For display, we depict the normalized difference score and FDR window scores (Fig. 1a, FDR transformation ($-10 \log_{10}$ (q-value FDR)), $20 = 0.01$, $25 = 0.003$, $30 = 0.001$, and $40 = 0.0001$). Histone-enriched loci for one individual (D1) were well correlated with a donor pool ($r = 0.7$). Globally, 76% of the top 9,841 histone-enriched regions (FDR 40 cutoff) intersect genic regions, whereas the expected intersection given random distribution is 36% ($P < 0.001$).

Interestingly, sequencing of D1 or the donor pool revealed significant (FDR < 0.001) histone retention at many loci important for embryo development, including embryonic transcription factors and signalling pathway components (Box 1, Supplementary Tables 2 and 3). We show this enrichment at *HOX* loci (Fig. 1, Supplementary Fig. 3), but also observe this at stand-alone developmental transcription factors (Supplementary Fig. 4) and signalling factors (Supplementary Fig. 5). An FDR of 60 yields 4,556 genes, of which 1,683 are grouped with developmental Gene Ontology categories (2,848 total developmental genes). The magnitude of nucleosome enrichment at developmental loci is modest, with high significance provided by a moderate average increase at a large number of loci. Histones are also significantly enriched at the promoters of microRNAs (miRNAs) ($P < 0.05$; Supplementary Fig. 6) and at the class of imprinted genes ($P < 0.0001$; Fig. 2), addressed in detail later. Selected loci were tested and confirmed by quantitative PCR (qPCR; Supplementary Fig. 7a–e). Outside of these enriched regions, we observe sequencing reads at low levels distributed genome-wide (for example, Figs 1a and 2a), an observation consistent with low levels of nucleosomes genome-wide, although contributions from non-nucleosomal contamination cannot be ruled out.

[1]Howard Hughes Medical Institute, Department of Oncological Sciences, and Huntsman Cancer Institute, [2]IVF and Andrology Laboratories, Departments of Surgery, Obstetrics and Gynecology, and Physiology, [3]Research Informatics and Bioinformatics Core Facility, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah 84112, USA.

Box 1 | Developmental genes are associated with particular chromatin attributes in human sperm

GoMiner was used to identify enriched categories, and all categories displayed have an FDR < 0.01. The top five general categories are listed, after omitting nearly identical/redundant classes. An expanded gene ontology table with the unfiltered top 30–60 categories, the total genes, number of changed genes, enrichment, and FDR are provided in the Supplementary Information.

Nucleosomes, Array D1
(1) Sequence-specific DNA binding; (2) multicellular organismal development; (3) regulation of transcription; (4) developmental process; (5) regulation of metabolic process.

Nucleosomes, Illumina GAII pooled donors
(1) Transcription factor activity; (2) cell fate commitment; (3) WNT receptor signalling; (4) neuron development; (5) embryonic development.

H3K4me2, Array D1
(1) Multicellular organismal development; (2) developmental process; (3) sequence-specific DNA binding; (4) anatomical structure development; (5) system development.

H3K4me3, Array D1
(1) mRNA processing; (2) RNA binding; (3) cell cycle; (4) transcription; (5) RNA splicing.

H3K4me3, Illumina GAII pooled donors
(1) RNA splicing; (2) translation; (3) cell cycle; (4) RNA metabolic process; (5) transcription.

H3K27me3, Illumina GAII pooled donors
(1) WNT receptor signalling; (2) embryonic organ development and morphogenesis; (3) cell fate commitment; (4) neuron differentiation; (5) sequence-specific DNA binding.

DNA hypomethylated promoters D1 and D2
(1) Embryonic development; (2) multicellular organismal development; (3) system development; (4) RNA biosynthetic process; (5) transcription factor activity.

DNA methylated promoters omitting CpG islands, array
(1) Transcription; (2) RNA biosynthetic process; (3) regulation of transcription; (4) embryonic development; (5) embryo morphogenesis.

Protamine occupancy (two replicas, $R^2 = 0.89$, arrays only) yielded 7,151 enriched regions (>2.5-fold), but failed to identify any enriched Gene Ontology term categories, although a few segments of the Y chromosome were notably enriched (including the testis-specific *TSPY* genes, data not shown). Regions of histone enrichment did not exclude protamine, consistent with a nucleosome-protamine mixture existing even at histone-enriched loci. However, as protamine fragments averaged ~750 bp, protamine depletion would have to be extensive (regions >2 kb) to be apparent on our arrays. Taken together, nucleosomes are significantly enriched in sperm at genes important for embryonic development, with transcription factors the most enriched class.

Localization of modified nucleosomes

Because histones replace protamines genome-wide at fertilization[7,8], unmodified histones retained in sperm would seem insufficient to influence gene regulation in embryos. Therefore, we examined three further chromatin properties in sperm: (1) histone variants, (2) histone modifications, and (3) DNA methylation. ChIP combined with promoter microarray analysis (termed ChIP-chip) of TH2B (two replicas, $R^2 = 0.93$) shows 0.3% of gene promoters with relatively high levels of TH2B (>twofold enrichment). Gene Ontology analysis showed significant (FDR < 0.06) enrichment at genes important for sperm biology, capacitation and fertilization (Supplementary Table 4), but not at developmental categories. ChIP sequencing (ChIP-seq) analysis with H2A.Z nucleosomes (at standard conditions, 150–250 mM salt) did not show significant enriched Gene Ontology categories, with high enrichment limited to pericentric heterochromatin (Supplementary Fig. 8), consistent with prior immunostaining[9].

Modified nucleosomes were localized by performing ChIP on mononucleosomes, followed by either array analysis or sequencing (Supplementary Fig. 2, schematic). We normalized the data set for each modification to the data set derived from input mononucleosomes, determined enriched regions (array >twofold; sequencing FDR 40), found the nearest neighbouring gene, and performed Gene Ontology analysis. In somatic cells, H3K4me2 is correlated with euchromatic regions. In sperm, H3K4me2 was enriched at many promoters, and at significant levels at promoters for developmental transcription factors (two replicas $R^2 = 0.94$; Gene Ontology term FDR < 0.06; Box 1 and Supplementary Table 5). In somatic cells, H3K4me3 is localized to: (1) the transcription start sites (TSS) of active genes, (2) genes bearing 'poised' RNA polymerase II (Pol II), and (3) the proximal promoter of inactive developmental regulators in embryonic stem (ES) cells—promoters that also bear the silencing mark H3K27me3 (refs 10, 11), and thus termed bivalent. Mature sperm are transcriptionally inert, and Pol II protein levels are barely detectable (data not shown), so the high H3K4me3 levels we observed in sperm chromatin (Supplementary Fig. 1f) seemed surprising. H3K4me3 was localized by both ChIP-chip (three replicas, $R^2 = 0.96$) and ChIP-seq. The raw data sets were similar ($r = 0.7$) and the thresholded data sets were very similar (array twofold; sequencing, FDR 40; 96% intersection, $P < 0.001$). With both data sets, simple inspection showed small peaks at many 5′ gene ends, with high levels and broader blocks at a subset of genes (that is, *HOX* loci; Fig. 1 and Supplementary Fig. 3). Gene Ontology term analyses with either data set yielded genes that are important for changing nuclear architecture, RNA metabolism, spermatogenesis, and also selected transcription factors important for embryonic development (FDR < 0.01, Box 1, Supplementary Tables 6 and 7 and Supplementary Fig. 9). H3K4me3 at genes related to nuclear architecture and spermatogenesis can presumably be attributed to their prior activation during gametogenesis. RNA metabolism occurs both in gametogenesis and the early embryo, so attribution to a prior program as opposed to a potential poising for a future program cannot be unambiguously attributed. However, several transcription and signalling factors of importance in embryo development exhibited high levels and a broad distribution of H3K4me3, including *EVX1/2*, *ID1*, *STAT3*, *KLF5*, *FGF9*, *SOX7/9*, certain *HOX* genes, and certain noncoding RNAs (Fig. 1 and Supplementary Figs 3 and 6).

Interestingly, ChIP-seq analysis showed significant levels of H3K27me3 at developmental promoters in sperm (Box 1, Fig. 1b, Supplementary Table 8 and Supplementary Figs 3 and 4), and overlapped significantly with H3K27me3-occupied genes in ES cells ($P < 0.01$), which are silent before differentiation. Furthermore, bivalent genes (bearing H3K4me3 and H3K27me3) in ES cells had a significant overlap with bivalent genes in sperm (FDR < 0.001 for each mark). Of the 1,999 genes identified as bivalent in ES cells, 861 were bivalent in sperm ($P < 0.01$; Supplementary Table 9). Also notable but not explored further were many blocks of high H3K4me3 or H3K27me3 in regions lacking annotation (Fig. 1a, oval). Furthermore, H3K9me3 was not detected at the small set developmental promoters tested, but was high at pericentric regions (qPCR only, Supplementary Fig. 7d). Taken together, our results demonstrate extensive histone modification patterns in sperm, and significant similarities to patterns observed in ES cells.

DNA methylation profiles

DNA methylation profiles examined two fertile donors (D2 and D4) using a methylated DNA immunoprecipitation (MeDIP) procedure

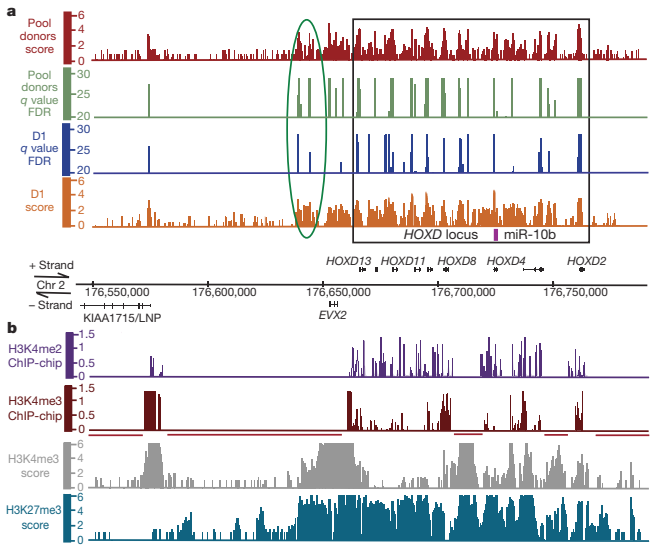


Figure 1 | Profiling of nucleosomes and their modifications at *HOXD*. For high-throughput sequencing, we show the mapped sequencing reads from D1 or a donor pool (red or orange bars, respectively; normalized difference score), and their significance (green or blue bars; FDR of 20 is <1% and FDR of 30 is <0.1%). **a**, The *HOXD* locus (black box) and an uncharacterized flanking locus (green oval). **b**, Profiling of nucleosome modifications at *HOXD* (in part **a**). The *y* axis is signal intensity ($\log_2$, for ChIP-chip), or the normalized difference score for sequencing. The regions not tiled on the array are underlined in red. Chr, chromosome.

and promoter arrays (individual replicates average D2 $R^2 = 0.97$ and D4 $R^2 = 0.89$). Their methylation patterns were highly similar (pairwise $R^2 = 0.86$), and extensive qPCR validated our array threshold (Supplementary Fig. 7e). Gene Ontology analysis of genes with pronounced DNA hypomethylation yielded transcription and signalling factors that guide embryo development (FDR < 0.05; Box 1 and Supplementary Table 10) including *HOX* loci (Fig. 3, blue bars, and Supplementary Figs 4 and 10). Hypomethylation also overlapped very significantly with histone-enriched promoters ($P < 0.02$; Supplementary Table 11). Bisulphite sequencing verified the MeDIP results, revealing extensive hypomethylation at developmental promoters in sperm (Supplementary Fig. 10b, c).

Notably, DNA-hypomethylated promoters in mature sperm overlap greatly with developmental promoters bound by the self-renewal network of transcription factors in human ES cells (for example, OCT4 (also known as POU5F1), SOX2, NANOG, KLF4 and FOXD3 proteins[12]; intersection of OCT4 protein occupancy and DNA hypomethylation, $P < 0.01$). In ES cells, these proteins promote self-renewal and also work with repressive polycomb complexes (PRC2; containing core component SUZ12) to help repress a large set of developmental regulators (including *HOX* genes) to prevent differentiation[10,13–20]. However, the hypomethylation of developmental genes in sperm is extensive (Fig. 3 and Supplementary Fig. 4). In fact, when CpG islands are omitted from the data sets, Gene Ontology term analysis of hypomethylated promoters still yields developmental genes (Box 1 and Supplementary Table 12). Notably, many of these developmental genes become methylated after differentiation; differential analysis of sperm and primary human fibroblasts (MeDIP, two replicas $R^2 = 0.86$) showed that many promoters occupied by PRC2 in human ES cells acquire methylation in fibroblasts (FDR < 0.01, Supplementary Tables 13 and 14; *HOXD* illustrated in Fig. 3, Supplementary Figs 4 and 5). Furthermore, the promoters driving several key members of the self-renewal network are themselves markedly hypermethylated in sperm (*OCT4*, *NANOG* and *FOXD3*, bisulphite sequencing in Supplementary Fig. 10c), whereas their developmental target genes are hypomethylated (bisulphite sequencing in Supplementary Fig. 10b), consistent with recent studies in mice[21–24].

Attributes of *HOX* clusters and miRNAs

Nucleosome enrichment was clear across *HOX* loci and proximal flanking regions, but falls off precipitously outside (*HOXD*, Fig. 1a; *HOXA*, Supplementary Fig. 3a). Histone-enriched *HOXD* regions with a single donor (D1) were largely shared with the donor pool (Fig. 1a; D1 versus donor pool, $r = 0.7$). Notably, retained nucleosomes have regional covalent modifications. For example, distinct and very large (5–20 kb) blocks of H3K4me3 are clearly observed at all *HOX* loci, and also at certain imprinted genes (addressed later). At *HOXD*, high H3K4me3 extends for ~20 kb, encompassing all of *EVX2* and extending to the 3′ region of *HOXD13* (Fig. 1b). Remarkably, a similar profile is observed at the related *HOXA* locus (Supplementary Fig. 3a). At *HOXD* a second block of H3K4me3 is observed in the region between *HOXD4* and *HOXD8* (Fig. 1b), a region that encodes several noncoding RNAs expressed during development. This region represents a marked difference from the chromatin status in ES cells; in ES cells *HOXD8–D11* are all bivalent. The distribution of H3K4me2 (determined from two replicas of D1) is clearly different from H3K4me3 at *HOX* loci (Fig. 1b and Supplementary Fig. 3). For example, at *HOXD*, H3K4me2 is enriched in *HOXD8–D11*, a region deficient in H3K4me3 (Fig. 1b). Notably, high H3K27me3 encompasses all *HOX* loci and their proximal flanking regions. In contrast, high levels of H3K9me (a mark of heterochromatin; Supplementary Fig. 7d) or H2A.Z were not detected at the *HOX* loci tested.

Histones are enriched at many miRNAs, especially miRNA clusters (Supplementary Fig. 6). For example, 16 of the 29 miRNA clusters on autosomes were significantly enriched ($P < 0.05$). Clusters include those bearing *let7e*, *mir-17*, *mir-15a*, *mir-96*, *mir-135b* and *mir-10a/*

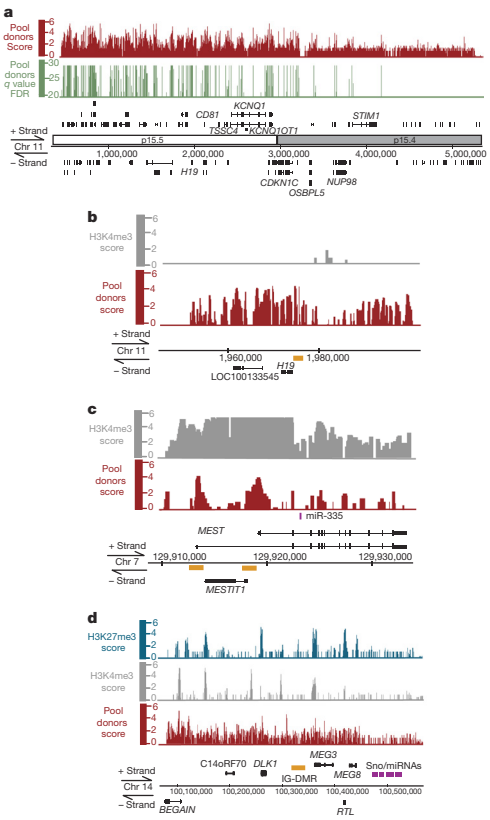


Figure 2 | Nucleosome enrichment at imprinted gene clusters, with high H3K4me3 at paternally expressed noncoding RNAs, and paternally demethylated regions. **a**, Histone enrichment at the 11p15.5 imprinted cluster (ending near *OSBPL5*), but not in the adjacent region. **b**, **c**, An expanded view of the DMRs (yellow rectangles) of *H19* (paternally methylated) (**b**) and *MEST* (paternally demethylated) (**c**). **d**, Moderate H3K4me3 at the promoters of the paternally expressed genes *BEGAIN*, *DLK1* and *RTL*, and the lack of H3K4me3 at the methylated intergenic-differentially methylated region (IG-DMR) of *MEG3* in sperm. Notably, both H3K4me3 and H3K27me3 reside at the promoter of *MEG3*, which later acquires DNA methylation in the embryo. Sno, small nucleolar.

b, as well as the stand-alone miRNAs *mir-153-1*, *mir-488* and *mir-760*. Notably, many histone-occupied miRNAs are associated with embryonic development[25] ($P < 0.01$), and their promoters were largely hypomethylated (Supplementary Fig. 10d). Furthermore, 7 of the 12 miRNAs on autosomes that are occupied by OCT4, NANOG and SOX2 in human ES cells[17] are also significantly occupied by histone (from pooled sequencing data). However, we do not at present understand the logic for their modification status; certain miRNA clusters have high histone and bivalent status, whereas others lack either modification (Supplementary Fig. 6).

Attributes of primary and secondary imprinted genes

Nucleosomes are significantly enriched at most imprinted genes in sperm, but at both paternally and maternally expressed loci. However, we observe marked specificity of H3K4me3 localization, with high and broad levels present at genes and noncoding RNAs that are paternally expressed. Locus 11p15.5 (Fig. 2a) is a large imprinted cluster with *IGF2*, *H19* and *KCNQ1* and several miRNAs. Here, increased levels of histone are present throughout the imprinted region (up to *OSBPL5*), but not in the large adjacent region lacking imprinted genes (Fig. 2a). Notably, the paternally silenced *H19* locus upstream of *KCNQ1* has a methylated DMR (Supplementary Fig. 10a) that lacks H3K4me3 (Fig. 2b). In contrast, *MEST* (a paternally expressed gene) has high H3K4me3 that extends from its promoter and first exon (containing the demethylated differentially methylated region (DMR); Fig. 2c and Supplementary Fig. 10a) through the second exon. The antisense noncoding RNA *MESTIT1* (also paternally expressed) is transcribed from the first intron, and is also very high in H3K4me3 (Fig. 2c). Furthermore, the promoter region of the paternally expressed antisense noncoding RNA *KCNQ1OT1* displays H3K4me3 (Fig. 2a and data not shown), and the DMR is DNA demethylated (Supplementary Fig. 10a). Several other examples of paternally expressed loci with blocks of H3K4me3 are provided in Supplementary Fig. 11, including *PEG3*, the noncoding RNAs *AIRN* (antisense to *IGF2R*) and *GNASAS* (antisense to *GNAS*). In contrast, genes flanking *KCNQ1* that are repressed by the noncoding RNA *KCNQOT1* (such as *OSBPL5*, *TSSC4* and *CD81*; Fig. 2a, expanded in Supplementary Fig. 11) contain histone, but lack H3K4me3. Notably, several paternally silenced genes (bearing DNA methylation) bore moderate (2–3-fold) enrichment of H3K9me3, a mark absent at paternally expressed genes (Supplementary Fig. 7d).

The 14q32.33 region (*DLK-DIO3*) is complex and interesting; paternally expressed genes such as *DLK1* and *RTL1* have moderate levels of H3K4me3 in their promoters, and the imprinting control locus (IG-DMR) lacks H3K4me3 (Fig. 2d) and is DNA methylated[26–28]. Notably, the promoter of *MEG3* (also known as *GTL2*; just downstream of the IG-DMR) lacks DNA methylation in sperm, but acquires DNA methylation in the embryo[26–28], termed secondary imprinting. Notably, the *MEG3* promoter region that later acquires DNA methylation initially bears both H3K4me3 and H3K27me3 in sperm; it is bivalent. One interpretation is that for mature sperm and early embryos, H3K4me3 prevents DNA methylation while H3K27me3 promotes silencing, with subsequent H3K4me removal enabling tissue-specific DNA methylation and secondary imprinting. Furthermore, our examination of the X chromosome inactivation centre showed an apparent bivalent status (and DNA hypomethylation) at the TSS of the *XIST* noncoding RNA, but not at *TSIX*,

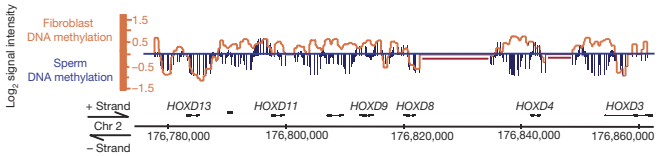


Figure 3 | Developmental promoters in sperm lack DNA methylation, but acquire methylation during development. DNA methylation of the *HOXD* locus in the mature sperm (blue bars) or primary fibroblasts (orange line overlay). The *y* axis is the signal intensity ($\log_2$) and the *x* axis is the annotated physical map (HG17). The regions not tiled on the array are underlined in red.

although future studies are required to determine whether these marks influence the regulation of this locus in the embryo (Supplementary Figs 6 and 10d; note that sequence reads on the X chromosome are half that on autosomes, as it is only present in 50% of sperm).

Modifications and expression timing

Transcriptome analysis has been performed in 4-cell and 8-cell human embryos, with 29 or 65 messenger RNAs identified as enriched, respectively[29]. Notably, genes in sperm bearing H3K4me3 but not H3K27me3 correlated with genes expressed at the 4-cell stage (14 out of 24, $P = 0.059$). Also, genes bearing high H3K4me2 were significantly enriched at genes expressed in the 4–8-cell stage (23 out of 49, $P < 0.02$; only 49 tiled on our array). In contrast, no significant correlation was observed with H3K27me3, which instead associates with transcription factors required for differentiation and organogenesis (discussed earlier). Furthermore, we verified by qPCR the presence of H3K4me2 or H3K4me3 at a subset of these stage-specific gene promoters (Supplementary Fig. 12). Thus, these findings reveal correlations of H3K4me2/3 enrichment, but not H3K27 enrichment, with early expression.

Conclusion

We provide several lines of evidence that the parental genome is packaged and covalently modified in a manner consistent with influencing embryo development. Previous analyses of DNA methylation in sperm identified hypomethylated promoters[23,24,30,31], showed similarities to the pattern in ES cells[24,31], and overlap between PRC2 and CpG islands[15,17,21,22]. We add that hypomethylated developmental promoters in human sperm overlap significantly with developmental promoters (in ES cells) occupied by the self-renewal network. Also, the promoters that acquire methylation in fibroblasts are primarily developmental transcription factors that are bound by PRC2 in human ES cells, consistent with recent work linking PRC2 to DNA methylation in development and neuronal differentiation in mice[21,32,33]. Thus, components of the self-renewal network emerge as candidates for helping to direct DNA hypomethylation in the germ line, and also to guide DNA hypermethylation to particular loci during differentiation, possibly to help 'lock in' differentiation decisions, although this remains to be tested.

The central findings of our work involve the significant enrichment of modified nucleosomes in the sperm genome at genes for embryo development, and a specificity to their modification patterns that might be instructive for the regulation of developmental genes, noncoding RNAs and imprinted loci. For example, histone retention and modification were clear at *HOX* loci and most of the targets of the self-renewal network in ES cells. One key concept in ES cell chromatin is the prevalence of developmental promoters with a bivalent status—bearing both H3K27me3 and H3K4me3 (ref. 10). Many promoters bivalent in ES cells are also bivalent in sperm, although some bear only H3K27me3 in sperm. Notably, H3K27me3 covers essentially all of the four *HOX* loci in sperm, whereas H3K4me3 is present in large blocks at only a subset of locations in *HOX* loci. Our work also provides correlations between H3K4me, but not H3K27me, and early expression in the embryo. In contrast, protamine-enriched loci did not show any significant Gene Ontology categories. However, there were certain segments of the Y chromosome with protamine enrichment, including the testis-specific *TSPY* genes, although the significance is not known.

We also find histones enriched at imprinted gene clusters, and a notable correlation between H3K4me3 and paternally expressed noncoding RNAs and genes; loci that lack DNA methylation in sperm. In contrast, maternally expressed noncoding RNAs/genes, and especially paternally methylated regions, lack H3K4me3 and (for the selected genes tested) contain moderate H3K9me3. Consistent with these observations, recent structural and *in vitro* data show that H3K4 methylation deters DNA methylation by DNMT3A2 and DNMT3L in mice[34]. However, experiments in model organisms are needed to address whether the modification patterns we report influence imprinting patterns *in vivo*. Taken together, we reveal chromatin features in sperm that may contribute to totipotency, developmental decisions and imprinting patterns, and open new questions about whether ageing and lifestyle affects chromatin in a manner that impacts fertility or embryo development.

METHODS SUMMARY

Biological samples. Sperm samples were obtained from four men of known fertility attending the University of Utah Andrology laboratory, consented for research. Samples were collected after 2–5 days abstinence and subjected to a density gradient (to purify viable, motile, mature sperm) and treated with somatic cell lysis buffer (0.1% SDS, 0.5% Triton X-100 in DEPC H_2O) for 20 min on ice to eliminate white blood cell contamination. Samples were centrifuged at 10,000*g* for 3 min, and the sperm pellet was resuspended in PBS and used immediately for chromatin preparation. Clontech human fibroblast cells (Lonza cc-2251) were cultured (37 °C and 5% CO_2) in DMEM containing 10% FBS and supplemented with penicillin and streptomyocin.

Chromatin immunoprecipitation. Standard ChIP methods were used[35], but we omitted crosslinking and used the following salt concentrations in the numbered buffers[35]: (1) 150 mM NaCl, (2) 250 mM NaCl, (3) 200 mM LiCl, and (4) 150 mM NaCl (the PBS wash). Antibodies used were: anti-H3K27me3 (Upstate 07-449), H3K4me3 (Abcam 8580), H3K4me2 (Abcam 32356), TH2B (Upstate 07-680), H2A.Z (Abcam 4174) and H3K9me3 (Abcam 8898). For each, 4 μl of antibody was coupled to 100 μl of Dynabeads (Invitrogen). After ChIP, samples for sequencing were not amplified, whereas for arrays the DNA was amplified (WGA, Sigma) before hybridization.

Methylation profiling using MeDIP. MeDIP procedures for sperm and primary human fibroblasts (Clonetech) were performed as described previously[30].

Sequencing. Sequencing used the Illumina GAII (Illumina Inc.) with standard protocols. Read numbers are final mapped microsatellite filtered reads (26–36 bases). Nucleosomes from D1: 19,658,110, D2–D4: 18,842,467, D1–4: 25,933,196 with equal contribution from each donor (random sub-sampling). Input, human sperm DNA: 17,991,622, H3K4me3: 13,337,105, H3K27me3:10,344,413, and H2A.Z: 5,449,000. All genomics data sets have been deposited in the Gene Expression Omnibus (GEO) under the SuperSeries GSE15594.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 March; accepted 27 May 2009.
Published online 14 June 2009.

1. Ward, W. S. & Coffey, D. S. DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells. *Biol. Reprod.* **44,** 569–574 (1991).
2. Wykes, S. M. & Krawetz, S. A. The structural organization of sperm chromatin. *J. Biol. Chem.* **278,** 29471–29477 (2003).
3. Balhorn, R., Brewer, L. & Corzett, M. DNA condensation by protamine and arginine-rich peptides: analysis of toroid stability using single DNA molecules. *Mol. Reprod. Dev.* **56,** 230–234 (2000).
4. Gatewood, J. M., Cook, G. R., Balhorn, R., Schmid, C. W. & Bradbury, E. M. Isolation of four core histones from human sperm chromatin representing a minor subset of somatic histones. *J. Biol. Chem.* **265,** 20662–20666 (1990).
5. Kimmins, S. & Sassone-Corsi, P. Chromatin remodelling and epigenetic features of germ cells. *Nature* **434,** 583–589 (2005).
6. Nix, D. A., Courdy, S. J. & Boucher, K. M. Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. *BMC Bioinformatics* **9,** 523 (2008).
7. Reik, W., Santos, F. & Dean, W. Mammalian epigenomics: reprogramming the genome for development and therapy. *Theriogenology* **59,** 21–32 (2003).
8. Santos, F., Hendrich, B., Reik, W. & Dean, W. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev. Biol.* **241,** 172–182 (2002).
9. Rangasamy, D., Berven, L., Ridgway, P. & Tremethick, D. J. Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. *EMBO J.* **22,** 1599–1607 (2003).
10. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
11. Bernstein, B. E. *et al.* Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl Acad. Sci. USA* **99,** 8695–8700 (2002).
12. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122,** 947–956 (2005).
13. Muller, J. & Kassis, J. A. Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Curr. Opin. Genet. Dev.* **16,** 476–484 (2006).
14. Schwartz, Y. B. *et al.* Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nature Genet.* **38,** 700–705 (2006).
15. Tanay, A., O'Donnell, A. H., Damelin, M. & Bestor, T. H. Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc. Natl Acad. Sci. USA* **104,** 5521–5526 (2007).

16. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298,** 1039–1043 (2002).
17. Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125,** 301–313 (2006).
18. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
19. Kopp, J. L., Ormsbee, B. D., Desler, M. & Rizzino, A. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells* **26,** 903–911 (2008).
20. Wernig, M. *et al. In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448,** 318–324 (2007).
21. Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30,** 755–766 (2008).
22. Illingworth, R. *et al.* A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.* **6,** e22 (2008).
23. Down, T. A. *et al.* A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature Biotechnol.* **26,** 779–785 (2008).
24. Farthing, C. R. *et al.* Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet.* **4,** e1000116 (2008).
25. Landgraf, P. *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129,** 1401–1414 (2007).
26. Glazov, E. A., McWilliam, S., Barris, W. C. & Dalrymple, B. P. Origin, evolution, and biological role of miRNA cluster in DLK-DIO3 genomic region in placental mammals. *Mol. Biol. Evol.* **25,** 939–948 (2008).
27. Takada, S. *et al. Delta-like* and *Gtl2* are reciprocally expressed, differentially methylated linked imprinted genes on mouse chromosome 12. *Curr. Biol.* **10,** 1135–1138 (2000).
28. da Rocha, S. T., Edwards, C. A., Ito, M., Ogata, T. & Ferguson-Smith, A. C. Genomic imprinting at the mammalian *Dlk1-Dio3* domain. *Trends Genet.* **24,** 306–316 (2008).
29. Li, S. S., Liu, Y. H., Tseng, C. N. & Singh, S. Analysis of gene expression in single human oocytes and preimplantation embryos. *Biochem. Biophys. Res. Commun.* **340,** 48–53 (2006).
30. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genet.* **39,** 457–466 (2007).
31. Fouse, S. D. *et al.* Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2,** 160–169 (2008).
32. Vire, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439,** 871–874 (2006).
33. Schlesinger, Y. *et al.* Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nature Genet.* **39,** 232–236 (2007).
34. Ooi, S. K. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* **448,** 714–717 (2007).
35. Gordon, M. *et al.* Genome-wide dynamics of SAPHIRE, an essential complex for gene activation and chromatin boundaries. *Mol. Cell. Biol.* **27,** 4058–4069 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Dalley for microarray and sequencing expertise, B. Schackmann for oligonucleotides, K. Boucher for statistical analysis, J. Wittmeyer for yeast nucleosomes and helpful comments, and T. Parnell for helpful comments. Financial support was from the Department of Urology (genomics and support of S.S.H.), the Howard Hughes Medical Institute (HHMI) (genomics, biologicals and support of J.P. and H.Z.), CA24014 and CA16056 for core facilities, and the Huntsman Cancer Institute (bioinformatics and support of D.A.N.). B.R.C. is an investigator with the HHMI.

Author Contributions B.R.C., D.T.C. and S.S.H. were involved in the overall design. D.T.C. and S.S.H. were responsible for acquisition of samples, clinical logistics, patient consenting and Institutional Review Board documents. B.R.C., S.S.H., D.A.N. and H.Z. designed detailed molecular and genomics approaches. D.A.N. carried out data processing and array analysis. S.S.H. and D.A.N. performed sequencing analysis. S.S.H. carried out experiments and produced the figures. J.P. carried out immunoblotting and bisulphite sequencing. B.R.C. wrote the manuscript.

Author Information The raw unfiltered reads (fastq format) are deposited at the Gene Expression Omnibus (GEO) under the SuperSeries GSE15594, which encompasses the Subseries entries GSE15690 for ChIP-seq data and GSE15701 for ChIP-chip data. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.T.C. (douglas.carrell@hsc.utah.edu) or B.R.C. (brad.cairns@hci.utah.edu).

doi:10.1038/nature08162 nature

METHODS

Partitioning of histone- and protamine-associated DNA. Chromatin was prepared from 40 million sperm as described previously[36] in the absence of crosslinking reagent, treated with sequential and increasing MNase (10–160 U), and centrifuged to sediment protamine-associated DNA, releasing mononucleosomes. The pooled mononucleosomes were used for ChIP, or the DNA was extracted and gel purified (~140–155 bp) for sequencing and array analysis.

ChIP and preparation for genomics methods. All ChIPs for sequencing were performed using the same pool of mononucleosomes from pooled donors. For arrays, a single pool was used from D1. ChIP methods were as described previously[35] but were performed without a crosslinking agent and slight modifications to the salt levels (250 mM NaCl, 200 mM LiCl), and the TE wash was replaced with a 150 mM PBS wash. ChIP methods used anti-H3K27me3 (Upstate 07-449), H3K4me3 (Abcam 8580), H3K4me2 (Abcam 32356), TH2B (Upstate 07-680), or H2A.Z (Abcam 4174) antibodies. For each, 4 µl of antibody was coupled to 100 µl of Dynabeads (Invitrogen). After the ChIP procedure, the DNA was amplified (WGA, Sigma) before hybridization to arrays, whereas samples used for Solexa were not amplified. For sequencing, DNA lengths corresponding to mononucleosomes with adapters (220–280 bp) were gel purified after the addition of the Illumina adaptors. This size selection was also performed for the nucleosomal DNA from pooled donors not subjected to ChIP.

Methylation profiling using MeDIP. This procedure was described previously[30]. In brief, sonicated sperm DNA was obtained from two different donors and sonicated fibroblast DNA was obtained from Clontech primary human fibroblasts (Lonza CC-2251) (4 µg, 300–1,000-bp fragments). Immunoprecipated DNA was washed, subjected to whole genome amplification (Sigma Aldrich). Amplified DNA (6 µg) was labelled with Cy5, and input DNA (6 µg) was labelled with Cy3 (Bio labs) by standard methods. Samples were hybridized to Agilent expanded promoter arrays, treated according to standard Agilent conditions, and scanned in an Agilent scanner.

Computational analytical methods. The software used in this analysis are open source and available from the TIMAT2 (http://timat2.sourceforge.net) and USeq (http://useq.sourceforge.net) project websites. Human annotation and genomic sequence (May 2004, NCBI Build 35, HG17 and March 2006, NCBI Build 36.1, HG18) were obtained from the UCSC Genome Bioinformatic website.

Low-level ChIP-chip analysis. Processing of the Agilent microarray promoter data was performed in three basic steps: data normalization, sliding window summaries, and enriched region identification. For each data set, the median unadjusted signal intensities from the Cy3 and Cy5 channels were extracted. Probes were then mapped to the HG17 or HG18 builds. Biological replicas were quantile normalized and median scaled to 100 (ref. 37). This normalization was applied to the treatment (ChIP samples) and control (whole genomic input DNA for the MeDIP and protamine data sets or DNA derived from mononucleosomes) replicas separately (see later for replica-averaged R^2). Probe level 'Oligo' summaries were calculated by taking the $\log_2$ ratio (mean treatment replicas/mean control replicas). 'Window' level summaries were generated by identifying windows of a particular size (100 bp for data sets derived from mononucleosomes, 675 bp for MeDIP and protamine data sets) containing a minimum number of oligonucleotide start positions (one for the data sets derived from mononucleosomes, three for the MeDIP and protamine data sets), and calculating an all pair (treatment versus control) relative difference pseudo median. This window summary score was assigned to the centre position of the window 'Pse' or represented as heat map 'PseHM' data. Extended regions of high-scoring windows, called 'intervals', were identified by merging windows that exceed a set threshold and are located within 250 bp of one another. Intervals were then ranked by their best window score. Relative difference pseudo median scores were converted to $\log_2$ ratio values.

The average R^2 values for microarray data were as follows: 0.85 for the three D1 MNase replicas; 0.89 for the three Protamine replicas; 0.96 for the two H3C replicas; 0.94 for the two H3K4me2 replicas; 0.93 for the two TH2B replicas; 0.96 for the three H3K4me3 replicas; and 0.93 for the two H3K27me3 replicas. The average MeDIP R^2 values for the three replicas of each donor were as follows: D2 average $R^2 = 0.97$ and D4 = 0.89, and the correlation between D2 versus D4 was 0.87. The average R^2 for the two primary human fibroblast MeDIP replicas was 0.86.

Low-level Chip-seq analysis. The DNA samples derived from mononucleosomes, and the sonicated control input genomic DNA were prepared for sequencing using Illumina's ChIP-seq kit. The 26-bp and 36-bp reads were generated using Illumina's Genome Analyser II and their standard software pipeline. Reads were mapped to the March 2006 NCBI Build 36.1 human genome using the pipeline's eland_extended aligner.

The USeq package[6] was used to identify regions of histone enrichment relative to input control. This entailed selecting reads that mapped with an alignment score ≥13 ($-10\log_{10}(0.05)$), shifting their centre position 73 bp 3′ to accommodate the 146-bp mononucleosome fragment length, and using a sliding window of 300 bp to score each region in the genome for significant histone enrichment. Significance was determined by calculating a binomial *P* value for each 300-bp window and controlled for multiple testing by applying Storey's *q* value FDR estimation[38,39].

Read numbers. Note the sperm genome has only 4% of the genome in nucleosomes. For nucleosome enrichment D1 had 19,658,110 reads, and the pool of three additional donors had 18,842,467 reads. The raw correlation for D1 versus the donor pool was $r = 0.7$. For all the analysis containing pool donors (D1, and a pooled sample of three additional individuals D2, D3 and D4) we used 25,933,196 mapped filtered reads with equal contribution from each donor (random subsampling). A total of 17,991,622 reads were generated from control input human sperm DNA, 3,337,105 reads from the H3K4me3 sample, 10,344,413 reads for H3K27me3, and 5,449,000 reads for H2Az. The raw unfiltered reads (fastq format) are deposited at GEO under the superseries GSE15594, which encompasses the Subseries entries GSE15690 for ChIP-seq and GSE15701 for ChIP-chip data.

To assess histone enrichment consistency, the QCSeqs application in the USeq package[6] was used to correlate the read counts between the D1 and pooled sample by calculating a Pearson correlation on the basis of the number of mapped reads falling within 500-bp windowed regions stepped every 250 bp across all chromosomes. Only windows with five or more reads in either of the samples were included in the correlation.

To create lists of candidate histone enriched regions, *q*-value thresholds of 20 (0.01) and 30 (0.001) ($-10\log_{10}(q\,\text{value})$) were selected. Overlapping windows that pass a given threshold were merged and scores from the best window assigned to the enriched region. The normalized window score was then used to rank and sort the regions.

A modification was made to score gene promoters and miRNAs for significant histone enrichment. The first step was to define regions for scoring. For gene promoters, the start of the first exon was used to define its hypothetical promoter by selecting a region 9 kb upstream and 2 kb downstream. For miRNAs, the centre position of each was expanded ±300 bp. These defined regions were scored for significant enrichment using the window statistics above.

High-level ChIP-chip and ChIP-seq analysis. Intersect regions. To identify regions of significant intersection between enriched region lists from various data sets, the USeq IntersectRegions application was used. This application counts the number of intersections between two lists of genomic coordinates that occur within a minimum 'max gap' distance. To estimate confidence in the intersections, a thousand 'random' data sets are generated that were matched to the chromosome and size of the original regions, and randomly picked from the interrogated regions on the array or sequenced regions in the genome. These randomized data sets were used to calculate a *P* value for the intersection and fold enrichment (fraction real intersection/fraction average random data set intersection) over random. Initial pilots that imposed a fraction GC match when picking random regions showed little difference with non-GC-matched random data sets and were thus subsequently dropped.

Find neighbouring genes (FNG). Genes associating with histones or histone modifications were determined using the FNG application in the USeq package. The gene lists were uploaded in GoMiner (http://discover.nci.nih.gov/gominer/htgm.jsp) to identify over represented Gene Ontology terms.

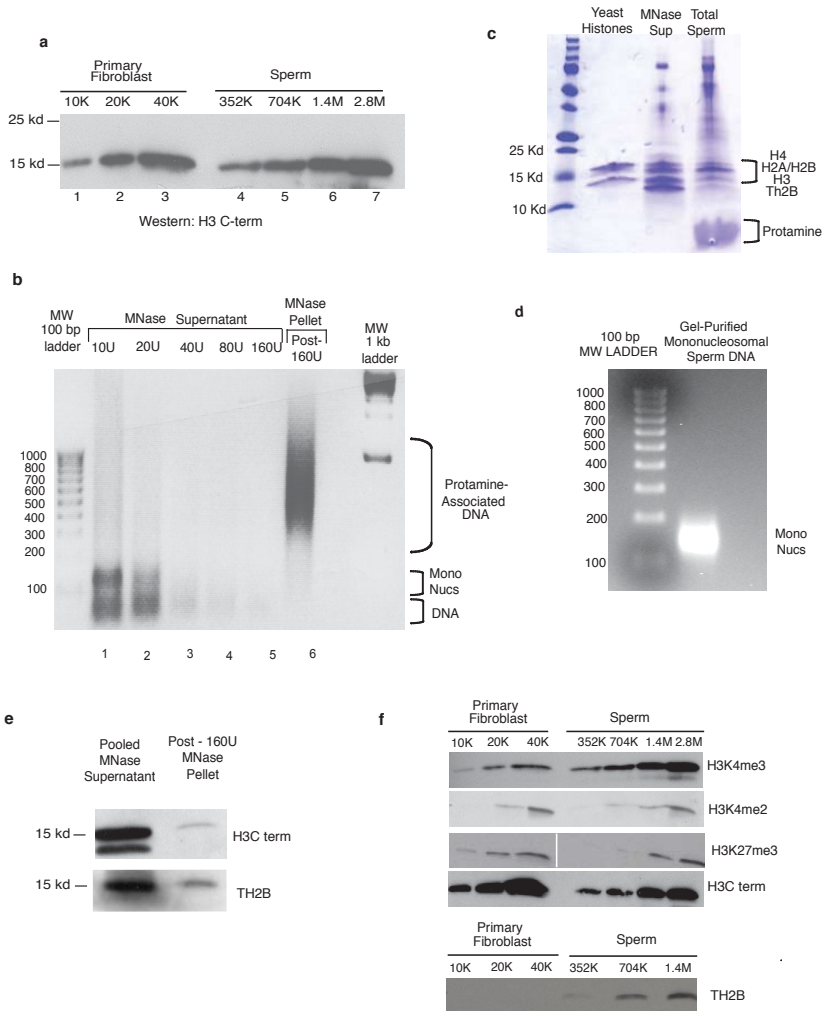
Intersect lists. To determine whether the 4- and 8-cell transcripts identified in early human embryo correlated with any of our histone modifications we used The IntersectLists USeq application which uses random permutation to calculate the significance of intersection between two lists of genes.

Aggregate plots. The USeq AggregatePlots application was used to compare the degree of enrichment and distribution of histone reads surrounding the TSS of developmental and non developmental genes. The gene classes were derived on the basis of Gene Ontology term categories.

36. Zalenskaya, I. A., Bradbury, E. M. & Zalensky, A. O. Chromatin structure of telomere domain in human sperm. *Biochem. Biophys. Res. Commun.* **279,** 213–218 (2000).
37. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19,** 185–193 (2003).
38. Dudoit, S., Gilbert, H. N. & van der Laan, M. J. Resampling-based empirical Bayes multiple testing procedures for controlling generalized tail probability and expected value error rates: focus on the false discovery rate and simulation study. *Biom. J.* **50,** 716–744 (2008).
39. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100,** 9440–9445 (2003).

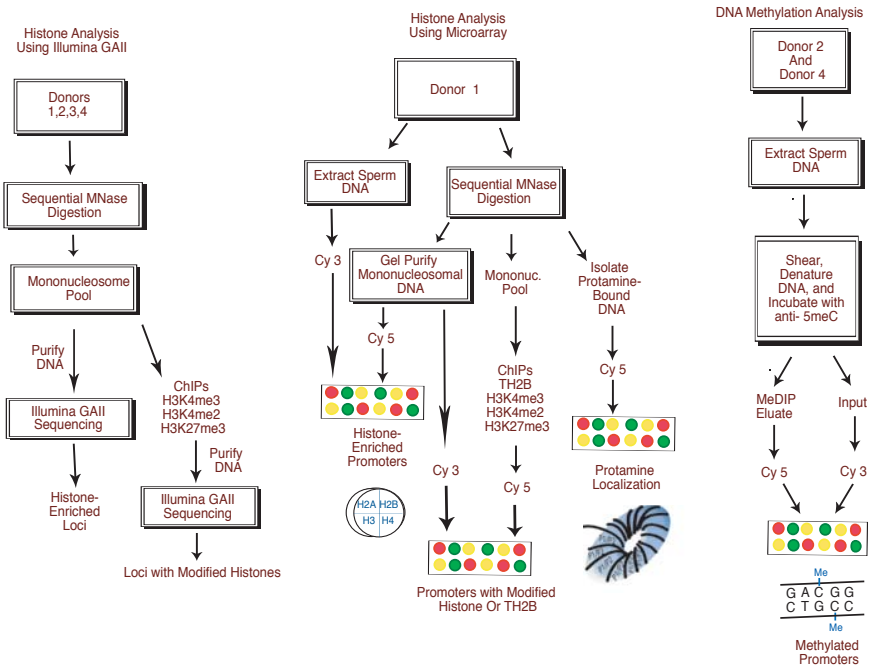
doi: 10.1038/nature08162 nature

SUPPLEMENTARY INFORMATION



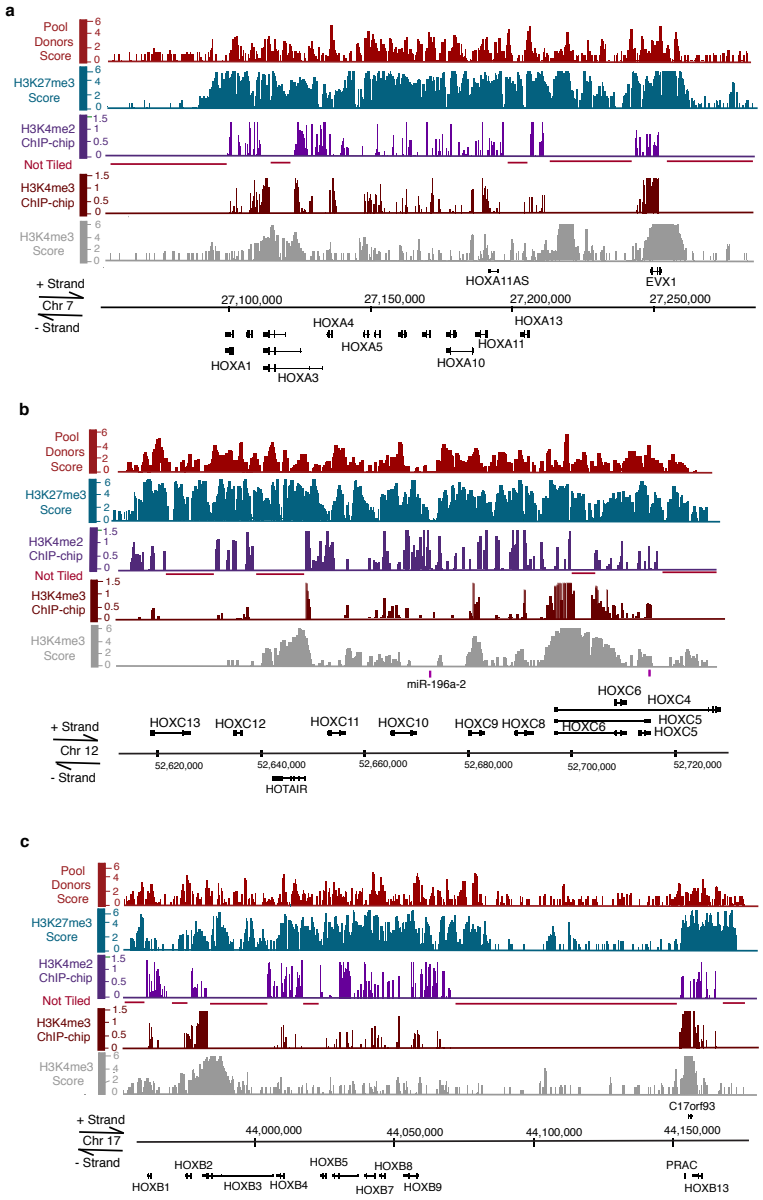
Supplemental Fig. 1: Composition of human sperm chromatin. **a**, Quantifying histone content of primary fibroblast or human sperm cells by immunoblot analysis with the H3C terminus antibody. **b**, Sequential digestion of sperm chromatin with increasing concentrations of micrococal nuclease (MNase) releases mononucleosomes (lanes 1 and 2), whereas protamine-packaged chromatin resists MNase (lane 6). **c**, Characterizing the mononucleosome fraction released into the MNase supernatant pool from panel **b**. **d**, Gel-purified mononucleosomal DNA used for array hybridization or sequencing. **e**, Quantification of the amount of histone released by MNase treatment. Supernatants were pooled. Here, cell equivalents were loaded in each lane; 4% of the total supernatant or protamine pellet. The gel was subjected to immunoblotting and quantified on a Typhoon (Amersham). **f**, Western analysis, involving titrations for bulk levels of H3K4me3, H3K4me2, H3K27me3 in primary fibroblast cells and mature sperm cells. Quanitation by Typhoon (Amersham) reveals that sperm bear ~4% of the histone H3 present in a primary fibroblast.

doi: 10.1038/nature08162



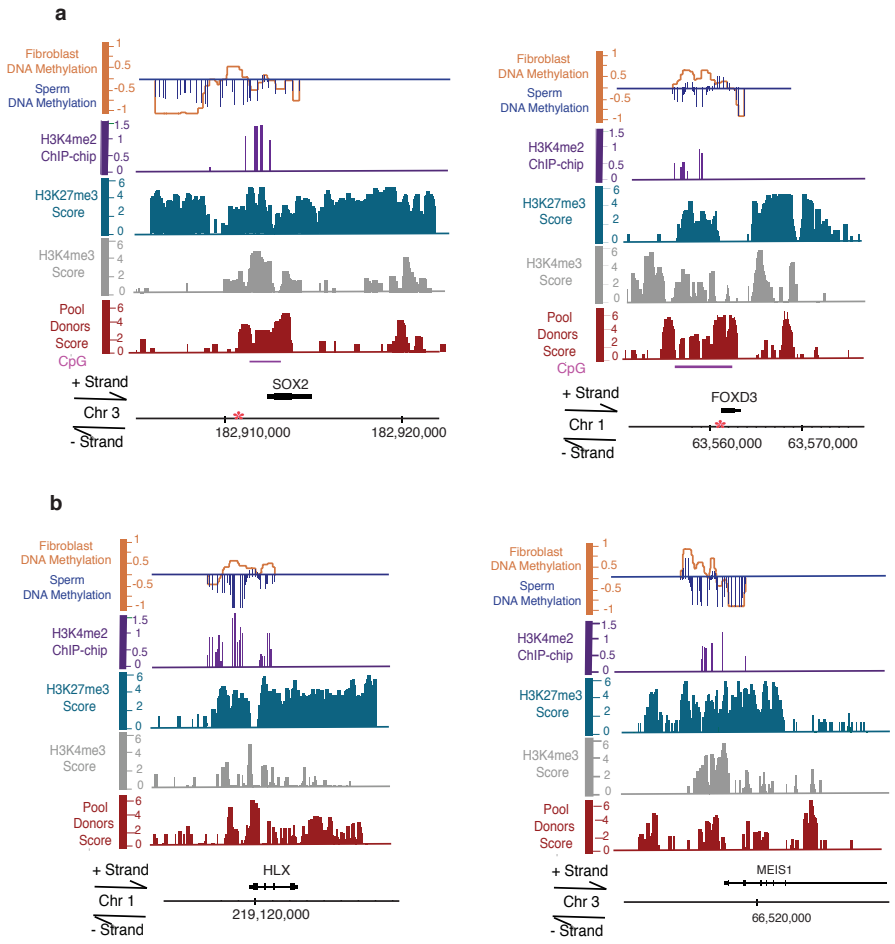
Supplemental Fig. 2: Schematic representation of experimental procedures. Two fertile donors were used for methylation studies, one donor (D1) was used for all histone modifications studied on the arrays. A pool of fertile donors were utilized for mononucleosome localization and characterization and to extend the analysis genome-wide using Illumina GAII.

doi: 10.1038/nature08162



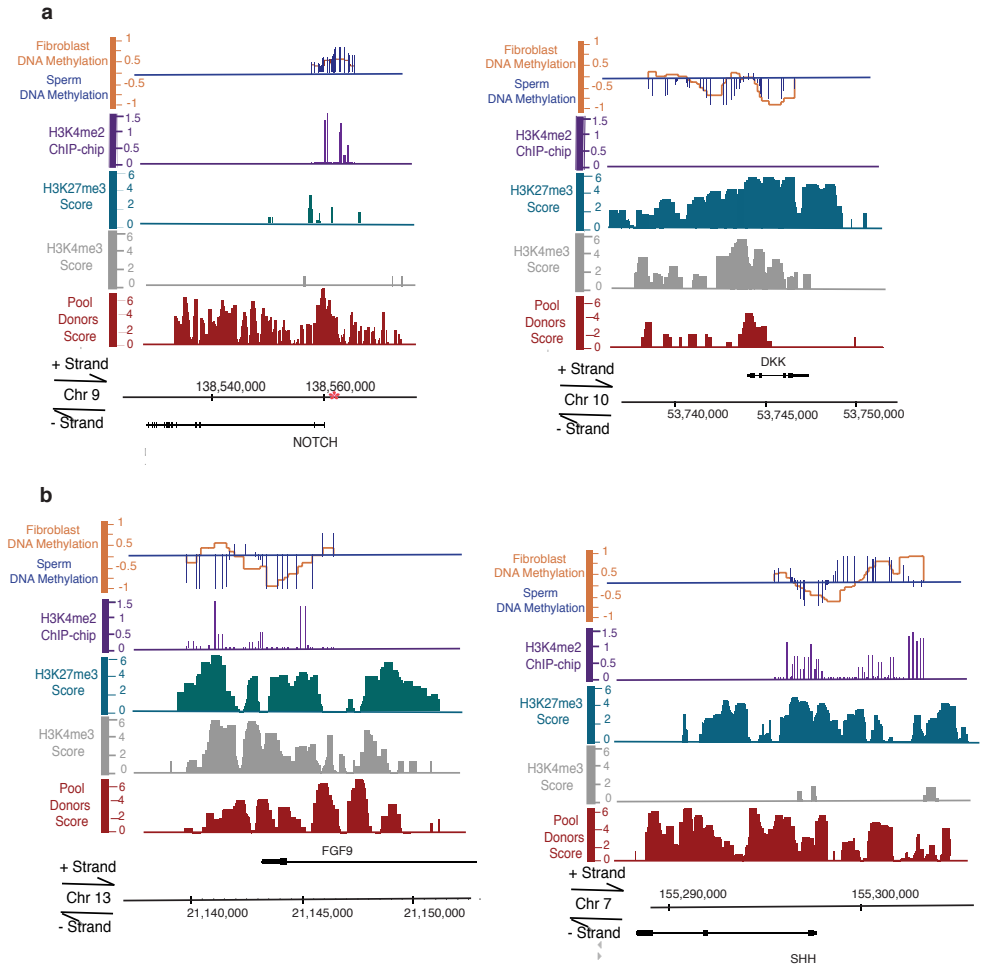
Supplemental Fig. 3: Chromatin attributes of the *HOXA, HOXB,* and *HOXC* loci. Histone enrichment (red bars), or histone modifications (H3K4me3 array results (ruby), H3K4me3 sequencing normalized difference scores (grey), H3K27me3 sequencing normalized difference scores (teal blue) or H3K4me2 (violet)). The y-axis is the signal intensity (log2 for array data, or normalized difference score for Illumina GAII sequencing) and the x-axis is the annotated physical map (HG18). **a**, The *HOXA* locus. **b**, The *HOXC* locus **c,** The *HOXB* locus.

doi: 10.1038/nature08162



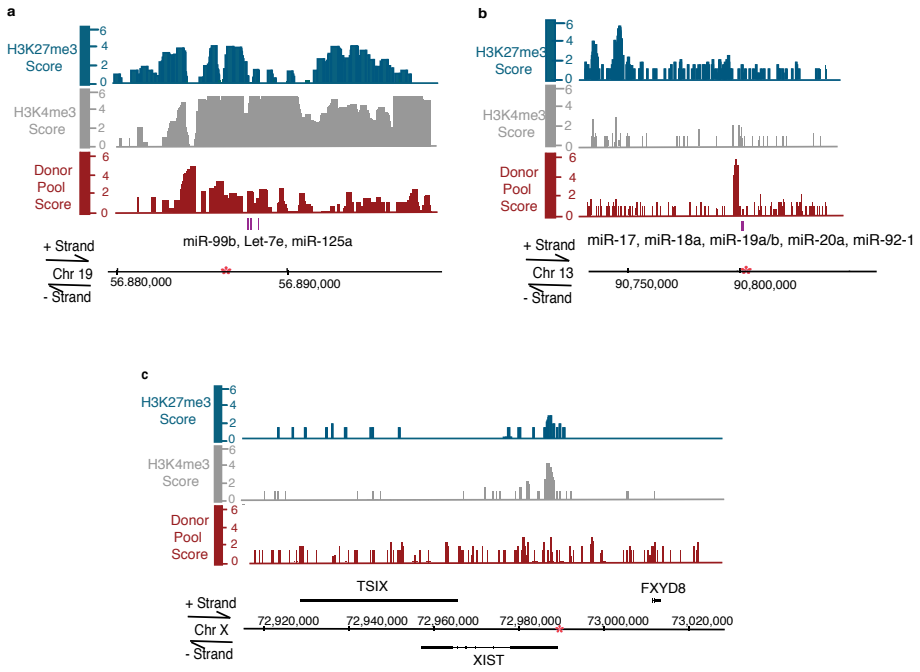
Supplemental Fig. 4: Certain self-renewal genes as well as genes required for embryonic development generally lack DNA methylation and are bivalent. **a**, SOX2 and FOXD3 are member of the pluripotency netwok. *SOX2* is demethylated and characterized by the presence of H3K4me3 and H3K27me3, whereas *FOXD3* is hypermethylated near their transcription start sites. *OCT4* and *NANOG* are also hypermethylated (Supplementary Fig. 10c). **b**, Genes involved in embryonic development are typically DNA hypomethylated, and have high levels of H3K4me2/3 and H3K27me3 around their start sites The red asterisks indicate the region amplified for bisulfite sequencing in Supplementary Fig. 10. The y-axis is the signal intensity (log2 for ChIP-chip arrays, or normalized difference for Illumina GAII sequencing score) and the x-axis is the annotated physical map (HG18).

doi: 10.1038/nature08162



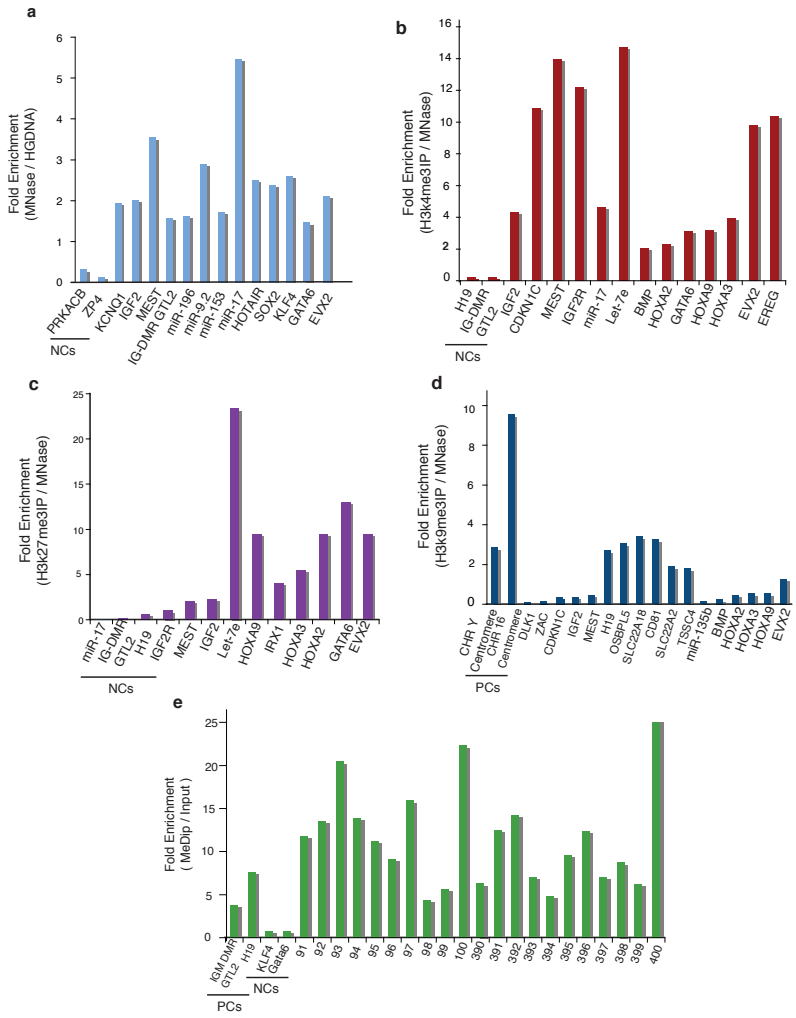
Supplemental Fig 5: Developmental and signaling factors are deficient in DNA methylation, although Notch pathway members are hypermethylated. **a**, Notch signaling pathway members, *DKK1* (hypomethylated) and *NOTCH1* (hypermethylated). **b**, FGF signaling pathway members, *FGF9* (hypomethylated) and regulator *SHH* (hypermehylated). The red asterisks indicates the region amplified for bisulfite sequencing. The y-axis is the signal intensity (log2 for ChIP-chip arrays, or normalized difference for Illumina GAII score) and the x-axis is the annotated physical map (HG18).

doi: 10.1038/nature08162



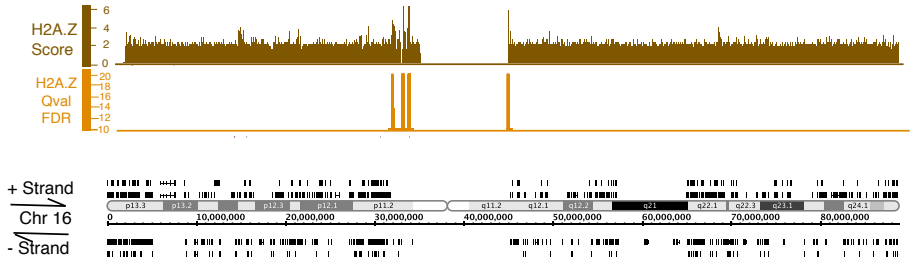
Supplemental Fig6: Histone retention at miRNAs and non-coding RNAs. **a**, A miRNA cluster with high H3K4me3 and H3K27me3. **b,** A miRNA cluster region with high levels of histone in the promoter region of the pri-miRNA, but lacking H3K4me3 and H3K27me3. **c,** The non-coding RNA *XIST* is enriched for H3K4me3 and H3K27me3 at the TSS. The read counts for the *X*-chromosome are half of those on autosomes due to the presence of either *X* or *Y* in sperm. The y-axis is the normalized difference score for sequencing. Asterisks (*) note the locations tested by bisulphite sequencing in Supplementary Fig. 10.

doi: 10.1038/nature08162



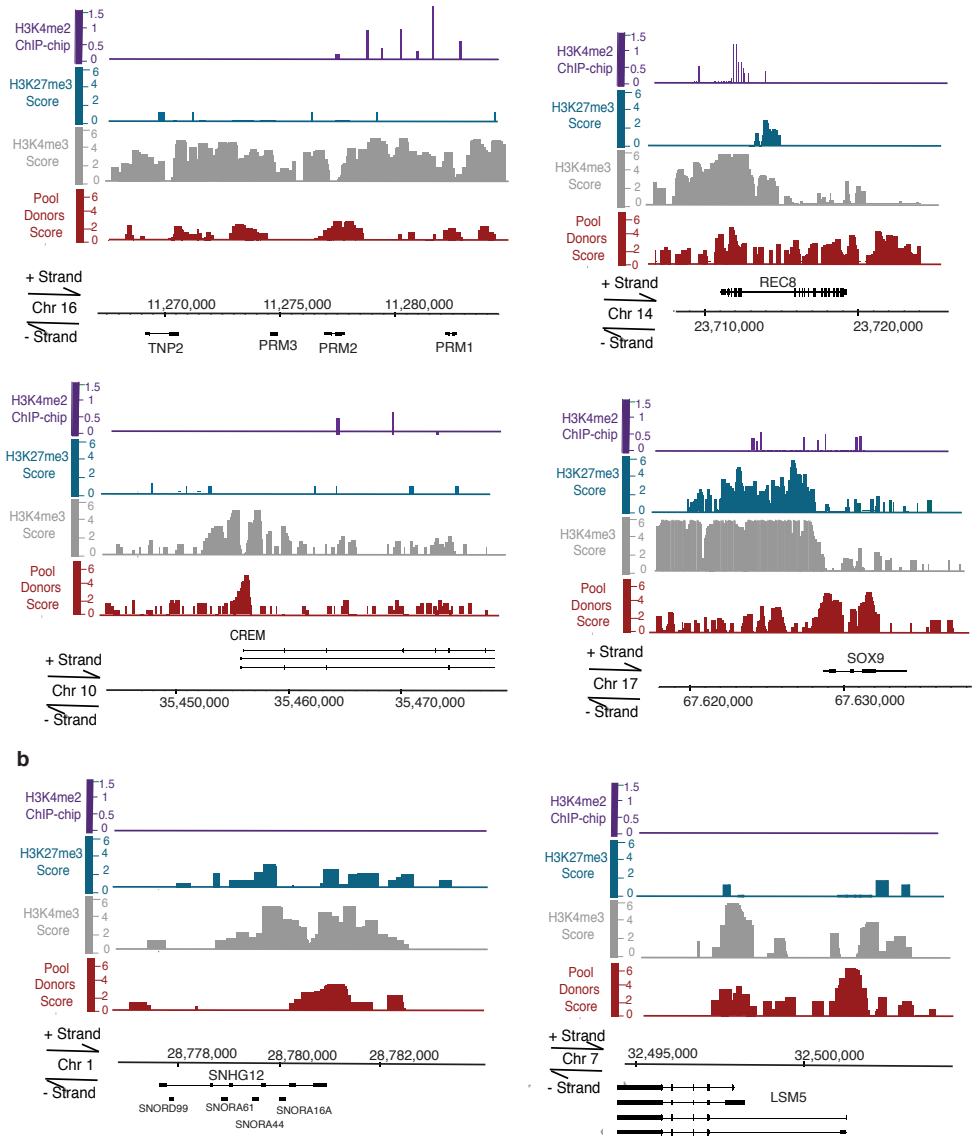
Supplemental Fig 7: qPCR testing of epigenetically modified loci enriched by Illumina GAII sequencing and/or array analysis. **a,** qPCR testing of histone occupancy at both maternally and paternally imprinted genes (*KCNQ1, IGF2, MEST*, and IG-DMR), miRNAs and noncoding RNA (*miR-196*, *miR-9.2*, *miR-153-1*, *miR-17*, and *HOTAIR*), and at a subset of developmental genes. Negative controls (NCs) (*PRKACB* and *ZP4*) are regions that had very low levels of histone by Illumina GAII sequencing and/or array data. Fold enrichment of histone at these promoters was determined by MNase signal divided by the total genomic DNA signal **b**, Fold enrichment of H3K4me3 was determined by normalizing signal from the H3K4me3 IP eluate to the signal from MNase (histone pool). Two maternally-imprinted loci in sperm were used as negative controls. **c**, H3K27me3 and **d**, H3K9me3 enrichment were determined as described above. H3K9me3 positive controls (PCs) were two pericentromeric heterochromatin loci. **e**, qPCR testing of MeDIP data. Enriched loci from MeDIP arrays were binned into the top 100 regions or 400 enriched regions. qPCR of MeDIP eluates were performed for the bottom 10 regions in each of the top 100 and 400 bins. Since all 20 regions enriched for DNA methylation, a cutoff of the top 400 genes (approximately 2-fold) was our stringent cutoff for DNA methylation. qPCR fold enrichment was compared to input (total sheared genomic DNA). Positive controls were two known methylated (imprinted) regions and negative controls were regions that are demethylated in sperm when compared to fibroblast.

doi: 10.1038/nature08162



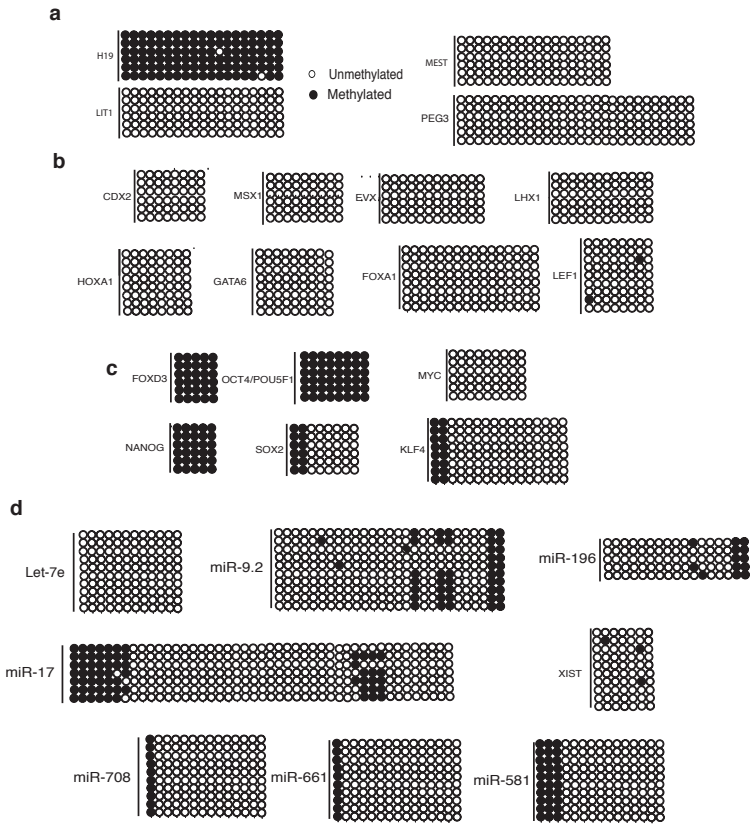
Supplemental Fig 8: H2A.Z localizes to pericentric heterochromatin in the mature human sperm. Brown bars are the normalized difference scores for pooled donor H2A.Z across chromosome 16, and in orange is the FDR. Other chromosomes showed similar peaks flanking the centromere. Pericentric heterochromatin was highly enriched with H2A.Z (FDR <0.05).

doi: 10.1038/nature08162



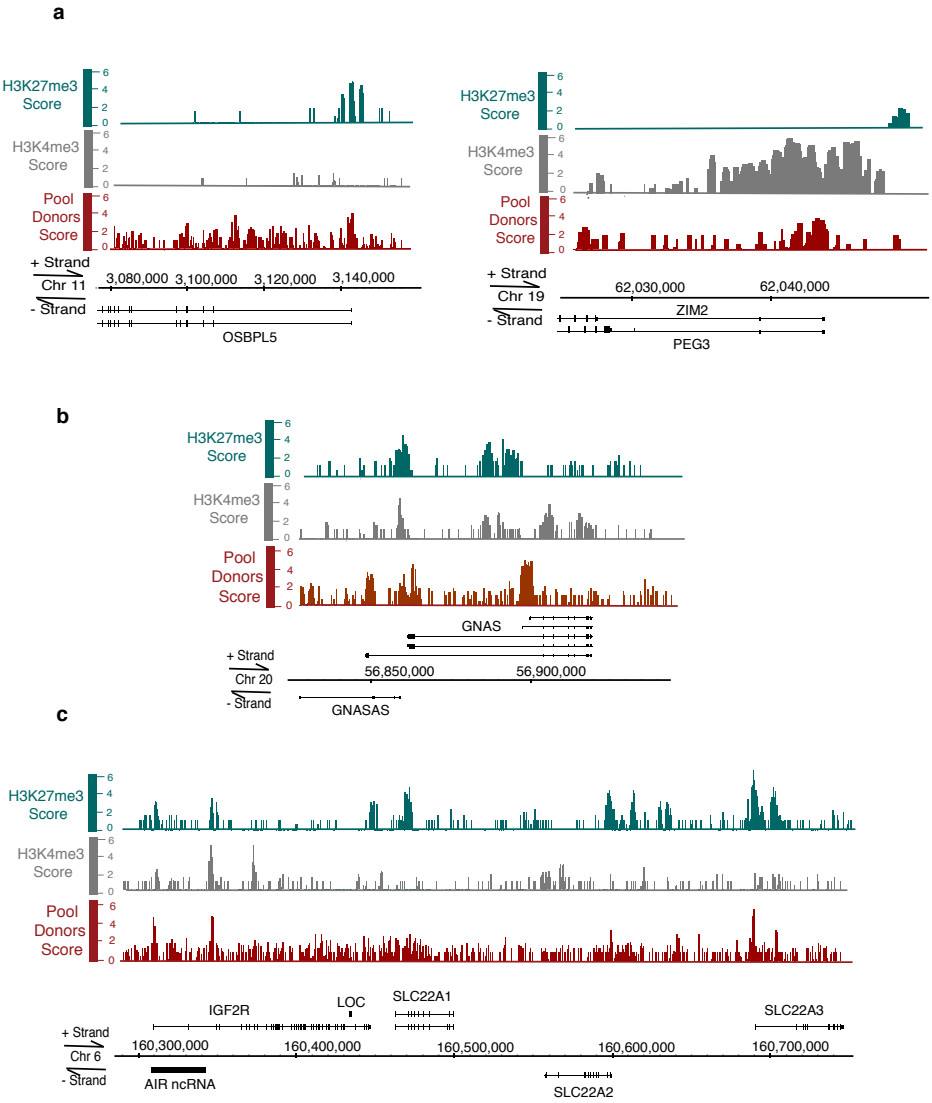
Supplemental Fig 9: Genes required for sperm development generally lack DNA methylation and are bound by H3K4me3. **a,** Four genes expressed at different stages of spermatogenesis remain DNA demethylated and retain H3K4me3 enrichment. **b**, Gene promoters involved in RNA processing, a process utilized intensely during spermiogenesis, are also demethylated and H3K4me3 bound.

doi: 10.1038/nature08162



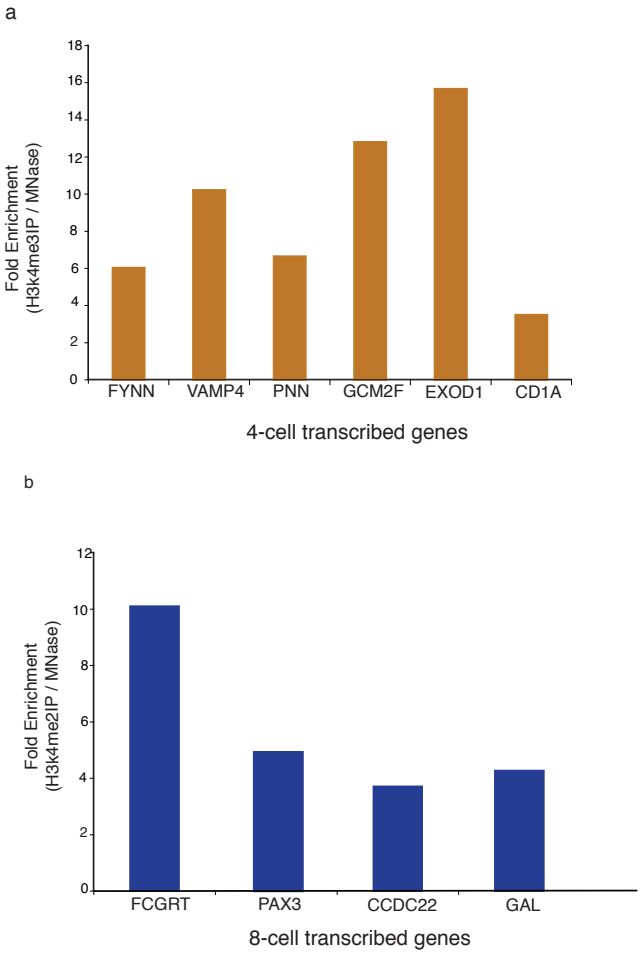
Supplemental Figure 10: DNA hypomethylation at developmental promoters and miRNAs were verified by bisulfite sequencing. **a**, bisulfite sequencing of promoters known to bear (*H19*) or lack (*LIT1, PEG3 and MEST*) paternal methylation in sperm chromatin. CpGs are represented as open dots (if unmethylated) or filled dots (if methylated). **b**, Hypomethylation at developmental transcription factors and **c,** a subset of the pluripotency network promoters. **d**, The TSS of the miRNAs tested were generally hypomethylated.

doi: 10.1038/nature08162



Supplemental Figure 11: H3K4me3 is generally present at paternally-expressed genes and non-coding RNAs. **a**, *OSBPL-5* (a maternally-expressed gene) lacks H3K4me3, whereas *PEG3* (a paternally-expressed gene) has high and broad H3K4me3 . **b,** The TSS of the *AIR* transcript localizes with H3K4me3, whereas promoters silenced by *AIR* (*SLC*s) lack H3K4me3 **c**, Similarly, the *GNASAS* is paternally expressed and has H3K4me3, whereas the remaining promoters are maternally expressed and lack H3K4me3.

doi: 10.1038/nature08162



Supplemental Fig 12: H3K4me2/3 chromatin modifications are correlated with early embryonic genes expression at the 4 and 8 cell stage. **a**, A subset of genes enriched at the 4 cell stage have significant levels of H3K4me3 **b**, whereas genes enriched at the 8-cell stage were associated with high levels of H3K4me2. Fold enrichment for H3K4me3/2 was determined by signal from IP eluate divided the signal derived from the pooled mononucleosomes.

doi: 10.1038/nature08162

SUPPLEMENTARY INFORMATION

Supplemental Table 1: Histone enriched promoters (D1 array)

Go Category	Total Genes	Changed Genes	Enrichment	FDR
Sequence-specific DNA binding	*425*	*32*	*4.601991*	*0*
Transcription factor activity	*755*	*38*	*3.076248*	*0*
Transcription regulator activity	*1090*	*46*	*2.579384*	*0*
Multicellular organismal development	*1620*	*59*	*2.225982*	*0*
DNA binding	*1522*	*54*	*2.168522*	*0*
Regulation of transcription DNA-dependent	*1467*	*51*	*2.124833*	*0*
Transcription DNA-dependent	*1510*	*52*	*2.104801*	*0*
RNA biosynthetic process	*1512*	*52*	*2.102017*	*0*
Regulation of transcription	*1580*	*52*	*2.011551*	*0*
Transcription	*1623*	*53*	*1.995915*	*0*
Developmental process	*1644*	*53*	*1.97042*	*0*
RNA metabolic process	*2265*	*73*	*1.969878*	*0*
Regulation of cellular metabolic process	*1827*	*58*	*1.940324*	*0*
Regulation of metabolic process	*1772*	*55*	*1.897071*	*0*
Regulation of cellular process	*1839*	*57*	*1.894427*	*0*
Regulation of biological process	*2889*	*84*	*1.777119*	*0*
Multicellular organismal process	*3134*	*89*	*1.735704*	*0*
Biological regulation	*2648*	*73*	*1.68496*	*0*
System development	*3396*	*93*	*1.673786*	*0*
Nucleobase nucleoside metabolic process	*1231*	*40*	*1.986034*	*0.0015*
Nucleic acid binding	*2489*	*66*	*1.620704*	*0.001905*
Transcription from RNA polymerase II promoter	*2348*	*63*	*1.639937*	*0.002174*
Anatomical structure development	*460*	*21*	*2.79027*	*0.002273*
Organ development	*1465*	*44*	*1.835692*	*0.0025*
Skeletal development	*869*	*31*	*2.180352*	*0.002692*
Urogenital system development	*174*	*12*	*4.215186*	*0.0028*
Kidney development	*31*	*5*	*9.858096*	*0.004545*
Wnt receptor activity	*29*	*5*	*10.537964*	*0.005*
Growth factor activity	*7*	*3*	*26.194368*	*0.007941*

doi: 10.1038/nature08162

Supplemental Table 2: D1 Histone-enriched loci (Illumina GAII FDR< 0.0001)

Go Category	Total Genes	Changed Genes	Enrichment	FDR
Cell fate commitment	*75*	*60*	*1.59848*	*0*
Sequence-specific DNA binding	*424*	*337*	*1.588112*	*0*
Cellular morphogenesis during differentiation	*125*	*99*	*1.582495*	*0*
Cell projection organization and biogenesis	*169*	*131*	*1.548823*	*0*
Cell part morphogenesis	*169*	*131*	*1.548823*	*0*
Embryonic morphogenesis	*88*	*68*	*1.543986*	*0*
Regionalization	*82*	*63*	*1.535126*	*0*
Neurogenesis	*221*	*168*	*1.518918*	*0*
Wnt receptor signaling pathway	*107*	*80*	*1.493907*	*0*
Regulation of cell differentiation	*119*	*88*	*1.477587*	*0*
Regulation of transcription from RNA polymerase II promoter	*99*	*72*	*1.453164*	*0*
Organ morphogenesis	*304*	*221*	*1.452566*	*0*
Embryonic development	*226*	*164*	*1.449949*	*0*
Regulation of developmental process	*191*	*138*	*1.443653*	*0*
Voltage-gated ion channel activity	*171*	*123*	*1.43723*	*0*
Nervous system development	*604*	*433*	*1.432413*	*0*
Cation channel activity	*228*	*162*	*1.419703*	*0*
Transcription factor activity	*791*	*552*	*1.394376*	*0*
Muscle development	*136*	*94*	*1.38104*	*0*
Central nervous system development	*190*	*129*	*1.356605*	*0*
Skeletal development	*193*	*130*	*1.34587*	*0*
Anatomical structure morphogenesis	*855*	*575*	*1.343751*	*0*
System development	*1396*	*934*	*1.336838*	*0*
Multicellular organismal development	*1868*	*1248*	*1.334919*	*0*
Channel or pore class transporter activity	*363*	*242*	*1.332067*	*0*
Enzyme linked receptor protein signaling pathway	*228*	*152*	*1.332067*	*0*
Positive regulation of transcription DNA-dependent	*180*	*120*	*1.332067*	*0*
Cell morphogenesis	*374*	*249*	*1.330286*	*0*
Cell Differentiation	*1437*	*874*	*1.330286*	*0*
Positive regulation of transcription	*227*	*151*	*1.329133*	*0*
Anatomical structure development	*1679*	*1107*	*1.317389*	*0*
Positive regulation of cell proliferation	*192*	*126*	*1.311253*	*0*
Organ development	*996*	*650*	*1.303981*	*0*
Cell fate commitment	*75*	*60*	*1.59848*	*0*
Sequence-specific DNA binding	*424*	*337*	*1.588112*	*0*
Cellular morphogenesis during differentiation	*125*	*99*	*1.582495*	*0*
Positive regulation of biological process	*850*	*513*	*1.548823*	*0*
51674 localization of cell	*324*	*203*	*1.251896*	*0*
32502 developmental process	*2619*	*1639*	*1.250434*	*0*
06812 cation transport	*432*	*270*	*1.248812*	*0*
15075 ion transporter activity	*622*	*387*	*1.243191*	*0*
42127 regulation of cell proliferation	*383*	*238*	*1.241639*	*0*
65009 regulation of a molecular function	*400*	*246*	*1.228831*	*0*
06366 transcription from RNA polymerase II promoter	*532*	*326*	*1.2244*	*0*
50790 regulation of catalytic activity	*381*	*233*	*1.221935*	*0*
05576 extracellular region	*1056*	*596*	*1.127716*	*0*
06351 transcription DNA-dependent	*1866*	*1050*	*1.124333*	*0*

doi: 10.1038/nature08162

Supplemental Table 2 continued: D1 Histone-enriched loci (Illumina GAII FDR< 0.0001)

Go Category	Total Genes	Changed Genes	Enrichment	FDR
RNA biosynthetic process	*1869*	*1051*	*1.123597*	*0*
Extracellular region	*1056*	*596*	*1.127716*	*0*
Guanyl-nucleotide exchange factor activity	*123*	*84*	*1.364556*	*0.000073*
Blood vessel development	*133*	*90*	*1.352098*	*0.000074*
Ras protein signal transduction	*176*	*115*	*1.305577*	*0.000074*
Negative regulation of developmental process	*65*	*49*	*1.50626*	*0.000075*
Transport	*2094*	*1134*	*1.082066*	*0.000075*
Embryonic development ending in birth or egg hatching	*81*	*59*	*1.455406*	*0.000076*
Extracellular matrix structural constituent	*84*	*61*	*1.451001*	*0.000076*
Transporter activity	*1090*	*611*	*1.120036*	*0.000077*
Cyclic nucleotide metabolic process	*34*	*29*	*1.704262*	*0.000078*
Positive regulation of developmental process	*49*	*39*	*1.590324*	*0.000078*
Vasculature development	*135*	*92*	*1.361668*	*0.000079*
Anion transport	*161*	*107*	*1.32793*	*0.000079*
Extracellular matrix organization and biogenesis	*44*	*36*	*1.634809*	*0.00008*
Heart development	*80*	*59*	*1.473599*	*0.000081*
Extracellular matrix organization and biogenesis	*44*	*36*	*1.634809*	*0.00008*
Heart development	*80*	*59*	*1.473599*	*0.000081*
Voltage-gated potassium channel complex	*80*	*59*	*1.473599*	*0.000081*
Chordate embryonic development	*80*	*59*	*1.473599*	*0.000081*
Developmental maturation	*48*	*38*	*1.581829*	*0.000145*
Kidney development	*29*	*25*	*1.7225*	*0.000201*
Transcriptional activator activity	*243*	*152*	*1.24984*	*0.000203*
Anterior posterior pattern formation	*50*	*39*	*1.558518*	*0.000204*
Cyclic nucleotide biosynthetic process	*26*	*23*	*1.76755*	*0.000205*
Establishment of localization	*2154*	*1162*	*1.077898*	*0.000207*
Extracellular region part	*697*	*400*	*1.146686*	*0.000211*
Anatomical structure formation	*132*	*89*	*1.347204*	*0.000213*
Sensory organ development	*56*	*43*	*1.534255*	*0.000214*
Metanephros development	*23*	*21*	*1.824352*	*0.000216*
Blood vessel morphogenesis	*120*	*81*	*1.348717*	*0.00025*
Ionotropic glutamate receptor activity	*18*	*17*	*1.887094*	*0.000252*
Glutamate-gated ion channel activity	*18*	*17*	*1.887094*	*0.000252*
Muscle contraction	*149*	*98*	*1.314187*	*0.000255*
Brain development	*101*	*70*	*1.384822*	*0.000256*

doi: 10.1038/nature08162

Supplemental Table 3: Donor pool of histone-enriched loci (Illumina GAII FDR< 0.0001)

Go Category	Total Genes	Changed Genes	Enrichment	FDR
RNA polymerase II transcription factor activity	*25*	*22*	*2.187319*	*0*
Cell fate commitment	*69*	*53*	*1.909221*	*0*
Regionalization	*86*	*60*	*1.734133*	*0*
Wnt receptor signaling pathway	*121*	*84*	*1.725534*	*0*
Pattern specification process	*123*	*85*	*1.717684*	*0*
Embryonic morphogenesis	*93*	*64*	*1.710514*	*0*
Sensory organ development	*80*	*55*	*1.708843*	*0*
Negative regulation of cell differentiation	*67*	*46*	*1.706525*	*0*
Cellular morphogenesis during differentiation	*124*	*85*	*1.703832*	*0*
Neurogenesis	*257*	*171*	*1.653836*	*0*
Embryonic development	*93*	*61*	*1.630333*	*0*
Chordate embryonic development	*93*	*61*	*1.630333*	*0*
Brain development	*133*	*87*	*1.625912*	*0*
Sequence-specific DNA binding	*488*	*311*	*1.584054*	*0*
Positive regulation of transcription from RNA polymerase II promoter	*145*	*92*	*1.577064*	*0*
Embryonic development	*221*	*140*	*1.574582*	*0*
Cell projection organization and biogenesis	*193*	*121*	*1.558323*	*0*
Cell part morphogenesis	*193*	*121*	*1.558323*	*0*
Regulation of cell differentiation	*157*	*98*	*1.551515*	*0*
Cell morphogenesis	*256*	*158*	*1.534075*	*0*
Cellular structure morphogenesis	*256*	*158*	*1.534075*	*0*
Central nervous system development	*227*	*140*	*1.532963*	*0*
Nervous system development	*675*	*408*	*1.502401*	*0*
Positive regulation of RNA metabolic process	*229*	*138*	*1.497867*	*0*
Skeletal development	*203*	*121*	*1.481559*	*0*
Vasculature development	*165*	*96*	*1.446162*	*0*
Organ morphogenesis	*355*	*205*	*1.435341*	*0*
Cell migration	*222*	*128*	*1.433133*	*0*
Anatomical structure morphogenesis	*823*	*457*	*1.380212*	*0*
Transcription activator activity	*284*	*157*	*1.374076*	*0*
System development	*1538*	*817*	*1.320369*	*0*
Multicellular organismal development	*2093*	*1104*	*1.31108*	*0*
Positive regulation of cellular process	*952*	*501*	*1.308068*	*0*
Anatomical structure development	*1768*	*930*	*1.307465*	*0*
Cell development	*1089*	*565*	*1.289585*	*0*
Cell differentiation	*1636*	*835*	*1.268623*	*0*
Cellular developmental process	*1636*	*835*	*1.268623*	*0*
Organ development	*1106*	*564*	*1.267516*	*0*
Developmental process	*2848*	*1443*	*1.259377*	*0*
Intracellular signaling cascade	*1291*	*653*	*1.257235*	*0*
Regulation of developmental process	*729*	*367*	*1.251319*	*0*
Regulation of RNA metabolic process	*2115*	*1049*	*1.232806*	*0*
Regulation of transcription DNA-dependent	*2103*	*1043*	*1.232749*	*0*
Regulation of transcription	*2228*	*1104*	*1.231639*	*0*
Regulation of gene expression	*2358*	*1159*	*1.221713*	*0*
Transcription DNA-dependent	*2159*	*1061*	*1.221497*	*0*
RNA biosynthetic process	*2163*	*1061*	*1.219238*	*0*

doi: 10.1038/nature08162

Supplemental Table 3 continued: Donor pool of histone-enriched loci (Illumina GAII FDR< 0.0001)

Go Category	Total Genes	Changed Genes	Enrichment	FDR
Regulation of metabolic process	*2629*	*1285*	*1.214904*	*0*
Transcription	*2315*	*1129*	*1.212195*	*0*
Anatomical structure formation	*152*	*89*	*1.455378*	*0.000072*
Transmembrane receptor protein tyrosine kinase activity	*62*	*43*	*1.723877*	*0.000074*
Small GTPase regulator activity	*201*	*112*	*1.385005*	*0.000136*
Respiratory tube development	*43*	*32*	*1.849741*	*0.000138*
Insulin receptor signaling pathway	*31*	*25*	*2.004508*	*0.000139*
Appendage morphogenesis	*37*	*28*	*1.880987*	*0.000189*
Limb morphogenesis	*37*	*28*	*1.880987*	*0.000189*
Appendage development	*37*	*28*	*1.880987*	*0.000189*
Limb development	*37*	*28*	*1.880987*	*0.000189*
Regulation of anatomical structure morphogenesis	*69*	*45*	*1.621037*	*0.000485*
Transcription corepressor activity	*106*	*64*	*1.500734*	*0.000491*
BMP signaling pathway	*18*	*16*	*2.209413*	*0.000539*
Regulation of neuron differentiation	*26*	*21*	*2.007592*	*0.000595*
Localization of cell	*365*	*184*	*1.25301*	*0.000694*
Protein-tyrosine kinase activity	*159*	*89*	*1.391305*	*0.000706*
Rho protein signal transduction	*101*	*61*	*1.501198*	*0.00071*
Small conjugating protein ligase activity	*137*	*78*	*1.415153*	*0.000769*
Forebrain development	*46*	*32*	*1.729106*	*0.000773*
Voltage-gated cation channel activity	*141*	*80*	*1.410264*	*0.000778*
Blood vessel morphogenesis	*145*	*82*	*1.405644*	*0.000787*
Tube development	*114*	*67*	*1.460829*	*0.000795*
Cartilage development	*35*	*26*	*1.846438*	*0.0008*
Regulation of cellular component organization and biogenesis	*241*	*127*	*1.309834*	*0.00082*
Mesoderm formation	*14*	*13*	*2.308048*	*0.000851*
Heart development	*93*	*56*	*1.496699*	*0.000947*
Regulation of neurogenesis	*41*	*29*	*1.7581*	*0.000952*
Negative regulation of developmental process	*314*	*159*	*1.258627*	*0.000964*
Regulation of cell proliferation	*456*	*223*	*1.215541*	*0.000984*
Voltage-gated ion channel activity	*189*	*102*	*1.34143*	*0.00099*
Voltage-gated channel activity	*189*	*102*	*1.34143*	*0.00099*
Actin filament-based process	*206*	*109*	*1.315191*	*0.001133*
Regulation of anatomical structure morphogenesis	*69*	*45*	*1.621037*	*0.000485*
Transcription corepressor activity	*106*	*64*	*1.500734*	*0.000491*
BMP signaling pathway	*18*	*16*	*2.209413*	*0.000539*
Regulation of neuron differentiation	*26*	*21*	*2.007592*	*0.000595*
Localization of cell	*365*	*184*	*1.25301*	*0.000694*
Protein-tyrosine kinase activity	*159*	*89*	*1.391305*	*0.000706*
Rho protein signal transduction	*101*	*61*	*1.501198*	*0.00071*
Small conjugating protein ligase activity	*137*	*78*	*1.415153*	*0.000769*
Forebrain development	*46*	*32*	*1.729106*	*0.000773*
Voltage-gated cation channel activity	*141*	*80*	*1.410264*	*0.000778*
Blood vessel morphogenesis	*145*	*82*	*1.405644*	*0.000787*
Tube development	*114*	*67*	*1.460829*	*0.000795*
Cartilage development	*35*	*26*	*1.846438*	*0.0008*
Regulation of cellular component organization and biogenesis	*241*	*127*	*1.309834*	*0.00082*

doi: 10.1038/nature08162

Supplemental Table 4: TH2B Enriched Promoters (D1 array)

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Beta DNA polymerase activity	*3*	*3*	*22.225524*	*0.018333*
Multidrug transport	*3*	*3*	*22.225524*	*0.018333*
Cation transport	*380*	*35*	*2.047088*	*0.023333*
Metal ion transport	*310*	*30*	*2.150857*	*0.0275*
Voltage-gated potassium channel	*72*	*11*	*3.395566*	*0.0325*
Potassium ion transport	*141*	*17*	*2.679673*	*0.03375*
Alpha-type channel activity	*333*	*32*	*2.135786*	*0.035*
Voltage-gated ion channel activity	*161*	*18*	*2.484841*	*0.035455*
Potassium ion binding	*106*	*14*	*2.935447*	*0.036*
Transporter activity	*1067*	*73*	*1.520584*	*0.037143*
Adenylate cyclase activity	*14*	*5*	*7.937687*	*0.04*
Channel or pore class transporter activity	*338*	*33*	*2.169948*	*0.06*

doi: 10.1038/nature08162

Supplemental Table 5: H3K4me2 enriched promoters (D1 array)

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Multicellular organismal development	*1620*	*148*	*1.394279*	*0.005*
Developmental process	*2265*	*197*	*1.327398*	*0.01*
Sequence-specific DNA binding	*425*	*50*	*1.795495*	*0.026667*
Anatomical structure development	*1465*	*132*	*1.375116*	*0.03*
System development	*1231*	*113*	*1.400953*	*0.031429*
Cell-cell signaling	*525*	*57*	*1.656985*	*0.035*
Organ development	*869*	*84*	*1.47524*	*0.035*
Menstrual cycle	*30*	*9*	*4.578511*	*0.04*
Multicellular organism reproduction	*45*	*10*	*3.39149*	*0.065294*
Reproductive process in a multicellular organism	*45*	*10*	*3.39149*	*0.065294*
Multicellular organismal process	*2648*	*212*	*1.221859*	*0.067333*

doi: 10.1038/nature08162

Supplemental Table 6: H3K4me3 enriched promoters (D1 array)

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Nuclear pore	*44*	*15*	*3.620699*	*0*
mRNA metabolic process	*198*	*44*	*2.36016*	*0*
mRNA processing	*165*	*35*	*2.25288*	*0*
Chromosome	*204*	*40*	*2.082494*	*0*
RNA processing	*266*	*51*	*2.036303*	*0*
Nuclear part	*596*	*109*	*1.94238*	*0*
RNA binding	*481*	*78*	*1.722279*	*0*
Cell cycle	*606*	*98*	*1.717542*	*0*
Cell cycle process	*530*	*82*	*1.643205*	*0*
RNA metabolic process	*1827*	*256*	*1.488179*	*0*
Transcription DNA-dependent	*1510*	*199*	*1.399684*	*0*
RNA biosynthetic process	*1512*	*199*	*1.397833*	*0*
Regulation of transcription DNA-dependent	*1467*	*193*	*1.397272*	*0*
Transcription	*1644*	*216*	*1.395423*	*0*
DNA binding	*1522*	*199*	*1.388648*	*0*
Regulation of transcription	*1580*	*206*	*1.384727*	*0*
Regulation of nucleobase nucleoside nucleotide and nucleic acid metabolic process	*1623*	*211*	*1.380759*	*0*
Regulation of cellular metabolic process	*1772*	*229*	*1.372542*	*0*
Regulation of metabolic process	*1839*	*237*	*1.368739*	*0*
Regulation of cellular process	*2889*	*341*	*1.253605*	*0*
Regulation of biological process	*3134*	*362*	*1.226771*	*0*
RNA splicing	*137*	*30*	*2.325705*	*0.000227*
Macromolecule localization	*548*	*83*	*1.608612*	*0.000233*
Intracellular transport	*494*	*75*	*1.612457*	*0.0004*
Cellular protein metabolic process	*2294*	*271*	*1.254671*	*0.000417*
RNA localization	*36*	*13*	*3.835259*	*0.000426*
Ligase activity	*238*	*42*	*1.874244*	*0.000727*
Establishment of cellular localization	*596*	*86*	*1.53252*	*0.000741*
Specific RNA polymerase II transcription factor activity	*29*	*11*	*4.028548*	*0.000755*
Translation initiation factor activity	*46*	*14*	*3.232393*	*0.000833*
Spliceosome	*88*	*21*	*2.53449*	*0.000847*
Nucleic acid transport	*35*	*12*	*3.641389*	*0.000862*
RNA transport	*35*	*12*	*3.641389*	*0.000862*
Establishment of RNA localization	*35*	*12*	*3.641389*	*0.000862*
Ribonucleoprotein complex	*328*	*53*	*1.716153*	*0.001475*
Nuclear membrane part	*54*	*15*	*2.9502*	*0.001791*
Pore complex	*54*	*15*	*2.9502*	*0.001791*
Tricarboxylic acid cycle	*22*	*9*	*4.344839*	*0.001846*
Acetyl-CoA catabolic process	*22*	*9*	*4.344839*	*0.001846*
Cellular localization	*611*	*86*	*1.494897*	*0.001905*
Ubiquitin cycle	*267*	*45*	*1.790009*	*0.001935*
Translation regulator activity	*99*	*21*	*2.25288*	*0.004405*
Spermatogenesis	*141*	*27*	*2.033755*	*0.004444*
Male gamete generation	*141*	*27*	*2.033755*	*0.004444*

doi: 10.1038/nature08162

Supplemental Table 6 continued: H3K4me3 enriched promoters (D1)

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Transcription regulator activity	*1090*	*136*	*1.325154*	*0.004458*
RNA export from nucleus	*20*	*8*	*4.248287*	*0.004512*
Translation factor activity nucleic acid binding	*86*	*19*	*2.346438*	*0.004828*
Protein transport	*460*	*66*	*1.523842*	*0.004884*
Microtubule-based process	*136*	*26*	*2.030431*	*0.005114*
Protein modification process	*1218*	*149*	*1.29925*	*0.006517*
Nuclear chromosome	*55*	*14*	*2.703456*	*0.007444*
Acetyl-CoA metabolic process	*27*	*9*	*3.540239*	*0.00828*
Transcription from RNA polymerase II promoter	*460*	*65*	*1.500754*	*0.00837*
Nucleobase nucleoside nucleotide and nucleic acid transport	*44*	*12*	*2.89656*	*0.008404*
Organelle organization and biogenesis	*711*	*93*	*1.389208*	*0.008438*
Microtubule cytoskeleton organization and biogenesis	*57*	*14*	*2.608598*	*0.008454*
Sexual reproduction	*218*	*36*	*1.75388*	*0.008526*
Meiotic recombination	*18*	*7*	*4.130279*	*0.010918*
Tricarboxylic acid cycle intermediate metabolic process	*23*	*8*	*3.694163*	*0.01101*
Nuclear export	*29*	*9*	*3.296085*	*0.018762*
Cofactor catabolic process	*29*	*9*	*3.296085*	*0.018762*
Gamete generation	*184*	*31*	*1.78936*	*0.018835*
Protein complex	*1361*	*161*	*1.256382*	*0.01902*
Intracellular protein transport	*289*	*44*	*1.616995*	*0.019208*
Endomembrane system	*331*	*49*	*1.572251*	*0.0194*

doi: 10.1038/nature08162

Supplemental Table 7: H3K4me3-enriched loci as determined from donor pool (Illumina GAII FDR< 0.001)

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
RNA splicing	*64*	*46*	*1.622*	*0*
spliceosome	*119*	*85*	*1.612*	*0*
ATP-dependent helicase activity	*88*	*62*	*1.59*	*0*
mRNA processing	*235*	*156*	*1.48*	*0*
Protein folding	*151*	*99*	*1.48*	*0*
Helicase activity	*129*	*84*	*1.4699*	*0*
ribonucleoprotein complex biogenesis and assembly	*186*	*134*	*1.468*	*0*
mRNA metabolic process	*272*	*177*	*1.463*	*0*
RNA processing	*404*	*253*	*1.413636*	*0*
Ribonucleoprotein complex	*400*	*250*	*1.41*	*0*
nucleolus	*158*	*97*	*1.385*	*0*
Microtubule-based process	*190*	*116*	*1.378*	*0*
Ligase activity	*338*	*204*	*1.362*	*0*
Translation	*351*	*210*	*1.350*	*0*
Mitotic cell cycle	*295*	*168*	*1.285*	*0*
Cell cycle phase	*323*	*183*	*1.285*	*0*
Nucleoplasm	*442*	*250*	*1.279302*	*0*
Nucleoplasm part	*381*	*215*	*1.272843*	*0*
Cell cycle process	*395*	*253*	*1.257553*	*0*
Transcription factor binding	*390*	*214*	*1.238*	*0*
RNA metabolic process	*2624*	*1411*	*1.213*	*0*
Transcription from RNA polymerase II promoter	*610*	*326*	*1.206*	*0*
Nucleobase nucleoside nucleotide metabolic process	*3279*	*1734*	*1.19*	*0*
DNA binding	*2080*	*1099*	*1.199734*	*0*
Cell cycle	*691*	*364*	*1.18*	*0*
Gene expression	*3028*	*1878*	*1.18*	*0*
Transcription	*2315*	*1207*	*1.173*	*0*
Transcription DNA-dependent	*2159*	*1121*	*1.172*	*0*
RNA biosynthetic process	*2163*	*1222*	*1.171*	*0*
Regulation of gene expression	*2358*	*1223*	*1.1698*	*0*
Transcription regulator activity	*1309*	*678*	*1.169*	*0*
Regulation of transcription DNA-dependent	*2103*	*1089*	*1.1689*	*0*
Regulation of transcription	*2228*	*1146*	*1.161*	*0*
Post-translational protein modification	*1346*	*777*	*1.133388*	*0*
Ribonucleotide binding	*1537*	*879*	*1.12284*	*0*

doi: 10.1038/nature08162

Supplemental Table 7 continued: H3K4me3-enriched loci as determined from donor pool (Illumina GAII FDR< 0.001)

GO CATEGORY	Total Genes	Changed Genes	Enrichment	FDR
M phase	*261*	*148*	*1.28*	*0.000098*
Mitochondrion	*807*	*413*	*1.15*	*0.00099*
Ribosome biogenesis and assembly	*86*	*57*	*1.49*	*0.0001*
Regulation of cell cycle	*272*	*154*	*1.278*	*0.000102*
Ubiquitin-dependent protein catabolic process	*164*	*98*	*1.3489*	*0.000185*
RNA helicase activity	*28*	*24*	*1.85*	*0.000187*
Protein RNA complex Assembly	*105*	*67*	*1.440*	*0.000189*
Spindle	*70*	*48*	*1.547*	*0.000192*
Spermatogenesis	*202*	*106*	*1.296*	*0.00082*
Male gamete generation	*202*	*106*	*1.296*	*0.00082*
Response to DNA damage stimulus	*278*	*152*	*1.234*	*0.003*
Mitosis	*198*	*112*	*1.276*	*0.003*
Flagellum	*30*	*23*	*1.73*	*0.00331*
Regulation of translation	*99*	*61*	*1.347*	*0.00349*
Centrosome	*124*	*74*	*1.347*	*0.00349*
Gamete generation	*247*	*136*	*1.24*	*0.00353*
Regulation of RNA cellular biosynthetic process	*128*	*76*	*1.3403*	*0.00357*
Negative regulation of cell cycle	*138*	*81*	*1.323143*	*0.00038*
mRNA splice site selection	*13*	*12*	*2.08*	*0.00397*
rRNA processing	*61*	*40*	*1.48839*	*0.00515*
nuclear chromosome part	*61*	*40*	*1.488*	*0.00515*
Translation initiation factor activity	*58*	*44*	*1.489455*	*0.0056*
Negative regulation of cellular process	*1023*	*579*	*1.111234*	*0.00057*
Chromosome organization and biogenesis	*345*	*211*	*1.200787*	*0.000606*
Regulation of protein metabolic process	*301*	*184*	*1.200201*	*0.001384*
RNA splicing via transesterification reactions	*64*	*47*	*1.441852*	*0.001392*
RNA splicing via transesterification reactions with bulged adenosine as nucleophile	*64*	*47*	*1.441852*	*0.001392*
Nuclear mRNA splicing via spliceosome	*64*	*47*	*1.441852*	*0.001392*
rRNA metabolic process	*64*	*47*	*1.441852*	*0.001392*
Establishment of cellular localization	*766*	*439*	*1.125223*	*0.0012*
Transcription factor complex	*161*	*105*	*1.28046*	*0.001208*
Establishment of protein localization	*674*	*389*	*1.133163*	*0.00125*
Regulation of cyclin-dependent protein kinase activity	*48*	*37*	*1.513433*	*0.001258*
Protein tyrosine phosphatase activity	*97*	*67*	*1.356144*	*0.001438*
Regulation of translation	*99*	*68*	*1.348579*	*0.001657*
Interphase of mitotic cell cycle	*84*	*59*	*1.379036*	*0.001667*
G1 S transition of mitotic cell cycle	*33*	*27*	*1.606396*	*0.001677*
Nucleolar part	*39*	*31*	*1.56063*	*0.001718*
Embryonic developmemt	*220*	*120*	*1.225*	*0.01*

doi: 10.1038/nature08162

Supplemental Table 8: Donor pool H3K27me3-enriched loci (Illumina GAII FDR <0.0001)

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Wnt receptor signaling	*20*	*18*	*2.706*	*0*
Embryonic organ development	*20*	*18*	*2.706*	*0*
Transmembrane receptor protein	*18*	*16*	*2.706*	*0*
Inner ear morphogenesis	*27*	*24*	*2.619*	*0*
Mesenchymal cell development and differentiation	*23*	*19*	*2.484*	*0*
Cell fate commitment	*69*	*54*	*2.353*	*0*
Embryonic morphogenesis	*93*	*71*	*2.295*	*0*
Lung development	*42*	*31*	*2.219*	*0*
Cyclic nucleotide metabolic process	*37*	*27*	*2.194*	*0*
Appendage morphogenesis	*37*	*27*	*2.1943*	*0*
Limb morphogenesis	*37*	*27*	*2.1943*	*0*
Appendage development	*37*	*27*	*2.1943*	*0*
Limb development	*37*	*27*	*2.1943*	*0*
Sensory organ development	*80*	*58*	*2.1800*	*0*
Potassium ion binding	*123*	*89*	*2.171*	*0*
Regionalization	*86*	*62*	*2.16*	*0*
Anterior posterior pattern formation	*54*	*38*	*2.116*	*0*
Axonogenesis	*112*	*77*	*2.06*	*0*
Pattern specification process	*123*	*84*	*2.0535*	*0*
Regulation of anatomical structure morphogenesis	*69*	*47*	*2.048*	*0*
Neuron differentiation	*206*	*139*	*2.029*	*0*
Forebrain development	*46*	*31*	*2.026*	*0*
Developmental maturation	*52*	*35*	*2.02*	*0*
Neuron morphogenesis during differentiation	*118*	*79*	*2.013*	*0*
Skeletal development	*203*	*133*	*1.970*	*0*
Neurite development	*133*	*87*	*1.966*	*0*
Neurogenesis	*265*	*165*	*1.930*	*0*
Cell migration	*222*	*142*	*1.9217*	*0*
Brian development	*133*	*85*	*1.921*	*0*
Embryonic development	*221*	*40*	*1.904*	*0*
Sequence specific DNA binding	*488*	*309*	*1.904*	*0*
Tube Development	*114*	*70*	*1.86*	*0*
Vasculature development	*165*	*101*	*1.846*	*0*
Organ morphogenesis	*335*	*215*	*1.821*	*0*
Blood vessel development	*162*	*98*	*1.819*	*0*
Central nervous system development	*227*	*137*	*1.814*	*0*
Heart development	*93*	*56*	*1.8106*	*0*
Anatomical structure formation	*152*	*91*	*1.8002*	*0*
Bone remodeling	*96*	*57*	*1.785*	*0*
Chordate embryonic development	*93*	*55*	*1.778*	*0*

doi: 10.1038/nature08162

Supplemental Table 8: Donor pool H3K27me3-enriched loci (Illumina GAII FDR <0.0001)

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
System process	*1264*	*504*	*1.1989*	*0*
Ligand gated ion channel	*97*	*54*	*1.674*	*0*
Embryonic limb morphogenesis	*33*	*24*	*2.186*	*0.000044*
Embryonic appendage morphogenesis	*33*	*24*	*2.186*	*0.00044*
Neural crest cell development and differentiation	*14*	*13*	*2.792*	*0.00082*
Metanephros development	*23*	*18*	*2.3533*	*0.000114*
Voltage-gated calcium channel complex	*21*	*17*	*2.434*	*0.000115*
Eye morphogenesis	*21*	*17*	*2.434*	*0.000115*
Eye development	*42*	*28*	*2.004*	*0.000116*
Transcription	*2315*	*570*	*1.183421*	*0.000153*
Dorsal ventral pattern formation	*28*	*18*	*3.089797*	*0.000154*
Endoderm development	*9*	*9*	*4.806351*	*0.000155*
Negative regulation of cell differentiation	*67*	*32*	*2.295571*	*0.000155*
Developmental maturation	*52*	*27*	*2.495605*	*0.000156*
Ligand-gated ion channel activity	*97*	*42*	*2.081101*	*0.000158*
Morphogenesis of an epithelium	*63*	*29*	*2.212447*	*0.00018*
Neuron fate commitment	*14*	*11*	*3.776419*	*0.000181*
Regulation of heart contraction	*42*	*22*	*2.517613*	*0.000182*
Tube morphogenesis	*82*	*35*	*2.051491*	*0.000183*
Tissue remodeling	*105*	*42*	*1.92254*	*0.000183*
Positive regulation of transcription DNA-dependent	*227*	*76*	*1.609175*	*0.000184*
Somitogenesis	*16*	*12*	*3.604763*	*0.000185*
Biological process	*12711*	*2729*	*1.031904*	*0.000186*
Growth factor activity	*164*	*59*	*1.729114*	*0.000187*

doi: 10.1038/nature08162

Supplemental Table 9: Loci enriched for H3K4me3 and H3k27me3 derived from donor pool sequencing data (Illumina GAII FDR< 0.0001)

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Cell fate determination	*27*	*14*	*4.239878*	*0*
Endocrine system development	*33*	*17*	*4.212346*	*0*
Cell fate commitment	*69*	*34*	*4.0292*	*0*
Neuron migration	*36*	*16*	*3.634181*	*0*
Embryonic morphogenesis	*93*	*41*	*3.604873*	*0*
Appendage morphogenesis	*37*	*16*	*3.53596*	*0*
Limb morphogenesis	*37*	*16*	*3.53596*	*0*
Appendage development	*37*	*16*	*3.53596*	*0*
Limb development	*37*	*16*	*3.53596*	*0*
Forebrain development	*46*	*19*	*3.377418*	*0*
Sensory organ development	*80*	*30*	*3.06634*	*0*
Anterior posterior pattern formation	*54*	*20*	*3.028484*	*0*
Brain development	*133*	*47*	*2.889584*	*0*
Regionalization	*86*	*30*	*2.852409*	*0*
Heart development	*93*	*32*	*2.813559*	*0*
Embryonic development	*221*	*74*	*2.737969*	*0*
Pattern specification process	*123*	*41*	*2.725636*	*0*
Homophilic cell adhesion	*133*	*43*	*2.643662*	*0*
Sequence-specific DNA binding	*488*	*155*	*2.597173*	*0*
Central nervous system development	*227*	*69*	*2.485491*	*0*
Chordate embryonic development	*93*	*28*	*2.461864*	*0*
Neurogenesis	*257*	*73*	*2.322623*	*0*
Tube development	*114*	*32*	*2.295272*	*0*
Skeletal development	*203*	*56*	*2.255698*	*0*
Organ morphogenesis	*355*	*97*	*2.234253*	*0*
Positive regulation of transcription from RNA polymerase II promoter	*145*	*39*	*2.199306*	*0*
Circulatory system process	*157*	*42*	*2.187453*	*0*
Blood circulation	*157*	*42*	*2.187453*	*0*
Regulation of cell differentiation	*157*	*42*	*2.187453*	*0*
Muscle development	*166*	*42*	*2.068856*	*0*
Transcription factor activity	*881*	*221*	*2.051188*	*0*
Neuron development	*152*	*38*	*2.044227*	*0*
Nervous system development	*675*	*168*	*2.035141*	*0*
Vasculature development	*165*	*41*	*2.031837*	*0*
Anatomical structure morphogenesis	*823*	*186*	*1.848001*	*0*
Organ development	*1106*	*249*	*1.840913*	*0*
Positive regulation of transcription DNA-dependent	*227*	*51*	*1.837102*	*0*
System development	*1538*	*340*	*1.807639*	*0*
Positive regulation of transcription	*278*	*60*	*1.7648*	*0*
Transcription regulator activity	*1309*	*281*	*1.755318*	*0*
Multicellular organismal development	*2093*	*446*	*1.742427*	*0*
Anatomical structure development	*1768*	*376*	*1.73898*	*0*
Positive regulation of metabolic process	*408*	*84*	*1.683481*	*0*

doi: 10.1038/nature08162

Supplemental Table 9 continued: Loci enriched for H3K4me3 and H3k27me3 derived from donor pool sequencing data (Illumina GAII FDR< 0.0001)

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Positive regulation of cellular metabolic process	*396*	*81*	*1.672549*	*0*
Regulation of transcription from RNA polymerase II promoter	*424*	*87*	*1.677809*	*0*
Biological adhesion	*683*	*139*	*1.664114*	*0*
Cell-cell signaling	*611*	*119*	*1.592556*	*0*
Developmental process	*2848*	*537*	*1.541783*	*0*
Transcription from RNA polymerase II promoter	*610*	*113*	*1.514738*	*0*
Cell differentiation	*1636*	*292*	*1.459448*	*0*
Cellular developmental process	*1636*	*292*	*1.459448*	*0*
Multicellular organismal process	*3267*	*567*	*1.419133*	*0*
Positive regulation of cellular process	*952*	*164*	*1.408627*	*0*
Cell development	*1089*	*186*	*1.396607*	*0*
Positive regulation of biological process	*1046*	*177*	*1.383664*	*0*
Negative regulation of cellular process	*1023*	*171*	*1.366814*	*0*
DNA binding	*2080*	*347*	*1.364128*	*0*
Regulation of transcription	*2228*	*368*	*1.350584*	*0*
Regulation of RNA metabolic process	*2115*	*348*	*1.34542*	*0*
Regulation of transcription DNA-dependent	*2103*	*346*	*1.345321*	*0*
Regulation of nucleobase nucleoside nucleotide and nucleic acid metabolic process	*2282*	*374*	*1.340124*	*0*
Transcription DNA-dependent	*2159*	*351*	*1.329363*	*0*
RNA biosynthetic process	*2163*	*351*	*1.326904*	*0*
Regulation of gene expression	*2358*	*382*	*1.324673*	*0*
Transcription	*2315*	*373*	*1.317489*	*0*
Biological regulation	*4522*	*682*	*1.233227*	*0*
Regulation of biological process	*4060*	*605*	*1.21848*	*0*
Cell communication	*3573*	*524*	*1.199188*	*0*
Positive regulation of nucleobase nucleoside nucleotide and nucleic acid metabolic process	*291*	*61*	*1.714059*	*0.000085*
Negative regulation of biological process	*1089*	*179*	*1.344046*	*0.000086*
Embryonic limb morphogenesis	*33*	*14*	*3.468991*	*0.000088*
Embryonic appendage morphogenesis	*33*	*14*	*3.468991*	*0.000088*
Positive regulation of RNA metabolic process	*229*	*51*	*1.821058*	*0.000089*
Substrate specific channel activity	*365*	*73*	*1.635381*	*0.00009*
Signal transduction	*3247*	*466*	*1.173526*	*0.000164*
Blood vessel development	*162*	*39*	*1.968515*	*0.000165*
Neurotransmitter binding	*101*	*28*	*2.266865*	*0.000167*
Positive regulation of heart contraction	*5*	*5*	*8.176907*	*0.000244*
Morphogenesis of an epithelium	*63*	*20*	*2.595843*	*0.000317*
Regulation of developmental process	*729*	*124*	*1.390859*	*0.000347*
Cellular morphogenesis during differentiation	*124*	*31*	*2.044227*	*0.00035*
Anatomical structure formation	*152*	*36*	*1.936636*	*0.000362*

doi: 10.1038/nature08162

Supplemental Table 10: Promoters deficient in DNA methylation (D2 and D4 array)

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Embryonic development	*199*	*22*	*3.061998*	*0*
Multicellular organismal development	*1620*	*102*	*1.743896*	*0*
System development	*1231*	*83*	*1.867478*	*0*
Nucleus	*2828*	*153*	*1.498468*	*0*
RNA biosynthetic process	*1512*	*95*	*1.740232*	*0*
Transcription	*1644*	*202*	*1.70159*	*0*
Transcription regulator activity	*1090*	*75*	*1.905768*	*0*
Anatomical structure development	*1465*	*92*	*1.739344*	*0*
Regulation of transcription	*1580*	*97*	*1.700396*	*0*
RNA metabolic process	*1827*	*108*	*1.637271*	*0*
Nucleobase nucleoside nucleotide and nucleic acid metabolic process	*2489*	*136*	*1.513385*	*0*
Regulation of cellular metabolic process	*1772*	*105*	*1.641198*	*0*
Nucleic acid binding	*2348*	*130*	*1.533489*	*0*
Regulation of transcription DNA-dependent	*1467*	*91*	*1.718093*	*0*
DNA binding	*1522*	*93*	*1.692402*	*0*
Regulation of metabolic process	*1839*	*106*	*1.596465*	*0*
Organ development	*869*	*61*	*1.94422*	*0*
Biopolymer metabolic process	*3392*	*170*	*1.388125*	*0*
Developmental process	*2265*	*123*	*1.504085*	*0*
Transcription factor activity	*755*	*54*	*1.980989*	*0*
Transcription from RNA polymerase II promoter	*460*	*38*	*2.288027*	*0*
Regulation of transcription from RNA polymerase II promoter	*297*	*26*	*2.424668*	*0.000588*
Female pronucleus	*3*	*3*	*27.697168*	*0.001081*
Nervous system development	*553*	*39*	*1.953326*	*0.001053*
Central nervous system development	*179*	*18*	*2.78519*	*0.00125*
Dorsal ventral pattern formation	*22*	*6*	*7.553773*	*0.001463*
Positive regulation of nucleobase nucleoside nucleotide and nucleic acid metabolic process	*199*	*19*	*2.644453*	*0.002381*
Gamete generation	*184*	*18*	*2.709506*	*0.002558*
Anatomical structure formation	*122*	*14*	*3.178364*	*0.002727*
Anatomical structure morphogenesis	*730*	*46*	*1.745301*	*0.003556*
Notch signaling pathway	*34*	*7*	*5.702358*	*0.003478*
Pronucleus	*4*	*3*	*20.772876*	*0.004681*
M phase	*175*	*17*	*2.690582*	*0.005417*
Multicellular organismal process	*2648*	*127*	*1.328376*	*0.0054*
Regionalization	*72*	*10*	*3.846829*	*0.00549*
Cell cycle phase	*214*	*19*	*2.459094*	*0.005385*
Negative regulation of cellular process	*776*	*47*	*1.677535*	*0.006038*
Sequence-specific DNA binding	*425*	*30*	*1.955094*	*0.006667*
Negative regulation of cellular metabolic process	*256*	*21*	*2.272033*	*0.007636*
Cell cycle process	*530*	*35*	*1.829058*	*0.007544*
Negative regulation of biological process	*807*	*48*	*1.647415*	*0.007414*
Chromosome	*204*	*18*	*2.443868*	*0.007288*
Brain development	*93*	*11*	*3.276009*	*0.011*
Positive regulation of transcription	*192*	*17*	*2.452353*	*0.011148*

doi: 10.1038/nature08162

Supplemental Table 10 continued : Promoters deficient in DNA methylation (D1 and D2 array)

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Positive regulation of cellular metabolic process	*264*	*21*	*2.203184*	*0.010968*
Gastrulation	*31*	*6*	*5.360742*	*0.013492*
Positive regulation of transcription DNA-dependent	*147*	*14*	*2.637826*	*0.019692*
Meiosis	*44*	*7*	*4.406368*	*0.020149*
M phase of meiotic cell cycle	*44*	*7*	*4.406368*	*0.020149*
Sexual reproduction	*218*	*18*	*2.286922*	*0.021618*
Meiotic cell cycle	*45*	*7*	*4.308448*	*0.022754*
Mitosis	*135*	*13*	*2.667135*	*0.023714*
Cellular protein complex disassembly	*14*	*4*	*7.913477*	*0.026197*
Positive regulation of metabolic process	*280*	*21*	*2.077288*	*0.025833*
Male pronucleus	*2*	*2*	*27.697168*	*0.042405*
Regulation of translational elongation	*2*	*2*	*27.697168*	*0.042405*
Heart development	*75*	*9*	*3.32366*	*0.042683*
Heart morphogenesis	*7*	*3*	*11.870215*	*0.045833*
Vasculature development	*122*	*12*	*2.724312*	*0.045412*
Forebrain development	*25*	*5*	*5.539434*	*0.046292*
Spermatogenesis	*141*	*13*	*2.55364*	*0.045495*
Male gamete generation	*141*	*13*	*2.55364*	*0.045495*

doi: 10.1038/nature08162

Supplemental Table 11: Promoters that share histone enrichment and DNA hypomethylation (array)

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Sequence-specific DNA binding	*425*	*35*	*5.037969*	*0*
Developmental process	*2265*	*37*	*2.186039*	*0*
Multicellular organismal development	*1620*	*30*	*2.478168*	*0*
DNA binding	*1522*	*27*	*2.373961*	*0*
Anatomical structure development	*1465*	*26*	*2.374981*	*0.002*
Transcription factor activity	*755*	*17*	*3.013189*	*0.003333*
RNA metabolic process	*1827*	*29*	*2.124144*	*0.002857*
Nucleic acid binding	*2348*	*34*	*1.937784*	*0.0025*
Regulation of transcription	*1467*	*25*	*2.280522*	*0.003333*
Neural tube patterning	*2*	*2*	*133.821053*	*0.006*
System development	*1231*	*22*	*2.391603*	*0.007273*
Transcription DNA-dependent	*1510*	*25*	*2.21558*	*0.0075*
RNA biosynthetic process	*1512*	*25*	*2.21265*	*0.006923*
Transcription regulator activity	*1090*	*20*	*2.455432*	*0.009286*
Transcription from RNA polymerase II promoter	*460*	*12*	*3.490984*	*0.008667*
Regulation of transcription	*1580*	*25*	*2.117422*	*0.0225*
Heart development	*75*	*5*	*8.921404*	*0.022353*
Regulation of nucleobase nucleoside nucleotide and nucleic acid metabolic process	*1623*	*25*	*2.061322*	*0.022778*
Regulation of metabolic process	*1839*	*27*	*1.964746*	*0.0235*
Skeletal development	*174*	*7*	*5.383606*	*0.023333*
Transcription	*1644*	*25*	*2.034992*	*0.022273*
Nucleobase and nucleic acid metabolic process	*2489*	*33*	*1.774245*	*0.032609*
Regulation of cellular metabolic process	*1772*	*26*	*1.963514*	*0.035*
Regulation of bone remodeling	*21*	*6*	*19.117293*	*0.0368*
Cell-cell signaling	*525*	*12*	*3.058767*	*0.035385*
Multicellular organismal process	*2648*	*34*	*1.718246*	*0.034074*
Voltage-gated potassium channel activity	*93*	*5*	*7.19468*	*0.040345*
Regulation of biological process	*3134*	*38*	*1.622591*	*0.039*
Biological regulation	*3396*	*40*	*1.57622*	*0.04*
Nervous system development	*553*	*12*	*2.903893*	*0.046875*
Alpha-type channel activity	*333*	*9*	*3.616785*	*0.045758*
Channel or pore class transporter activity	*338*	*9*	*3.563282*	*0.046176*
Anatomical structure morphogenesis	*730*	*14*	*2.566431*	*0.045143*
Positive regulation of cell differentiation	*28*	*3*	*14.33797*	*0.058056*
Regulation of cellular process	*2889*	*35*	*1.621231*	*0.057568*
Cellular morphogenesis during differentiation	*108*	*5*	*6.195419*	*0.056053*
Positive regulation of cellular process	*671*	*13*	*2.592658*	*0.056154*
Cell development	*859*	*15*	*2.336805*	*0.065854*
Potassium channel activity	*118*	*5*	*5.670384*	*0.067674*
Organ development	*869*	*15*	*2.309915*	*0.066136*

doi: 10.1038/nature08162

Supplemental Table 12: Sperm DNA demethylation extends beyond CpGs

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Nucleus	*2665*	*103*	*1.668257*	*0*
Transcription DNA-dependent	*1398*	*64*	*1.97604*	*0*
RNA biosynthetic process	*1399*	*64*	*1.974627*	*0*
Regulation of transcription DNA-dependent	*1353*	*62*	*1.977957*	*0*
Regulation of RNA metabolic process	*1365*	*62*	*1.960568*	*0*
Transcription regulator activity	*996*	*50*	*2.166874*	*0*
Transcription	*1518*	*65*	*1.848266*	*0*
Regulation of transcription	*1457*	*63*	*1.866396*	*0*
RNA metabolic process	*1698*	*70*	*1.77944*	*0*
Regulation of nucleobase nucleoside nucleotide and nucleic acid metabolic process	*1500*	*64*	*1.841669*	*0*
Regulation of gene expression	*1540*	*65*	*1.821862*	*0*
Regulation of cellular process	*2683*	*97*	*1.560537*	*0*
Regulation of cellular metabolic process	*1708*	*69*	*1.74375*	*0*
Regulation of biological process	*2873*	*101*	*1.51743*	*0*
Intracellular	*6181*	*180*	*1.257004*	*0*
Gene expression	*2015*	*77*	*1.649448*	*0*
DNA binding	*1398*	*59*	*1.821662*	*0*
Biopolymer metabolic process	*3194*	*108*	*1.459526*	*0*
Regulation of metabolic process	*1757*	*69*	*1.695119*	*0*
Intracellular part	*5850*	*171*	*1.26172*	*0*
Nucleobase nucleoside nucleotide and nucleic acid metabolic process	*2249*	*82*	*1.573792*	*0*
Embryonic development	*183*	*15*	*3.538043*	*0.00037*
Transcription factor activity	*688*	*34*	*2.133111*	*0.000385*
Biological regulation	*3250*	*107*	*1.421096*	*0.0004*
Nucleic acid binding	*2071*	*76*	*1.584004*	*0.000417*
Embryonic morphogenesis	*78*	*10*	*5.533862*	*0.000435*
Macromolecule metabolic process	*4200*	*131*	*1.34631*	*0.000455*
Positive regulation of transcription	*214*	*16*	*3.227224*	*0.00069*
Anatomical structure morphogenesis	*643*	*32*	*2.148137*	*0.000714*
Positive regulation of nucleobase nucleoside nucleotide and nucleic acid metabolic process	*225*	*16*	*3.069449*	*0.000857*
Positive regulation of metabolic process	*324*	*20*	*2.664452*	*0.000882*
Cellular component organization and biogenesis	*1639*	*62*	*1.63281*	*0.000909*
Membrane-bounded organelle	*4209*	*129*	*1.32292*	*0.000938*
Intracellular membrane-bounded organelle	*4207*	*129*	*1.323549*	*0.000968*
Positive regulation of cellular metabolic process	*312*	*20*	*2.766931*	*0.001*
Positive regulation of RNA metabolic process	*171*	*13*	*3.281483*	*0.003864*
Primary metabolic process	*4979*	*145*	*1.257039*	*0.003902*
RNA polymerase II transcription factor activity	*170*	*13*	*3.300786*	*0.003953*

doi: 10.1038/nature08162

Supplemental Table 12 continued: Sperm DNA demethylation extends beyond CpGs

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Positive regulation of transcription DNA-dependent	*170*	*13*	*3.300786*	*0.003953*
Embryonic development ending in birth or egg hatching	*83*	*9*	*4.680447*	*0.004*
Chordate embryonic development	*83*	*9*	*4.680447*	*0.004*
Transcription from RNA polymerase II promoter	*450*	*24*	*2.302087*	*0.004211*
Positive regulation of transcription from RNA polymerase II promoter	*109*	*10*	*3.960011*	*0.004222*
Pattern specification process	*102*	*10*	*4.231777*	*0.004324*
Anatomical structure development	*1378*	*52*	*1.628835*	*0.004681*

doi: 10.1038/nature08162

Supplemental Table 13: Gene promoters occupied by Suz12 in ES cell are DNA demethylated and histone bound in sperm

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Sequence-specific DNA binding	*425*	*51*	*10.377959*	*0*
Transcription factor activity	*755*	*65*	*7.445556*	*0*
Transcription regulator activity	*1090*	*69*	*5.471537*	*0*
Multicellular organismal development	*1620*	*41*	*4.407517*	*0*
DNA binding	*1522*	*72*	*4.348046*	*0*
Regulation of transcription DNA-dependent	*1467*	*72*	*4.112349*	*0*
Transcription DNA-dependent	*1510*	*37*	*4.267268*	*0*
RNA biosynthetic process	*1512*	*37*	*4.261624*	*0*
developmental process	*2265*	*84*	*3.38306*	*0*
Regulation of transcription	*1580*	*72*	*4.078212*	*0*
RNA metabolic process	*1827*	*39*	*3.717502*	*0*
Regulation of nucleobase nucleoside	*1623*	*37*	*3.970163*	*0*
Nucleotide and nucleic acid metabolic process				
Regulation of metabolic process	*1839*	*39*	*3.693245*	*0*
Transcription	*1644*	*72*	*3.78945*	*0*
Regulation of cellular metabolic process	*1772*	*38*	*3.734608*	*0*
Nucleic acid binding	*2348*	*42*	*3.115132*	*0*
Regulation of biological process	*3134*	*48*	*2.667273*	*0*
Regulation of cellular process	*2889*	*45*	*2.712627*	*0*
Multicellular organismal process	*2648*	*43*	*2.827976*	*0*
Biological regulation	*3396*	*48*	*2.461494*	*0*
Nucleobase nucleoside nucleotide and nucleic acid metabolic process	*2489*	*40*	*2.798725*	*0*
System development	*1231*	*58*	*4.07*	*0*
Anatomical structure development	*1465*	*36*	*3.776226*	*0*
Organ development	*869*	*44*	*4.408878*	*0*
Nucleus	*2828*	*39*	*2.401654*	*0*
Nervous system development	*553*	*16*	*5.038718*	*0*
Transcription from RNA polymerase II promoter	*460*	*13*	*4.92165*	*0*
Anatomical structure morphogenesis	*730*	*36*	*4.264915*	*0*
Cellular metabolic process	*5390*	*50*	*1.615498*	*0*
Skeletal development	*174*	*8*	*8.006928*	*0*
Primary metabolic process	*5420*	*50*	*1.606556*	*0*
Lung development	*32*	*4*	*21.768836*	*0.00027*
Neural tube patterning	*2*	*2*	*174.150685*	*0.000263*
Respiratory tube development	*33*	*4*	*21.109174*	*0.000256*
Cellular process	*8815*	*65*	*1.284151*	*0.000476*
Central nervous system development	*179*	*7*	*6.810362*	*0.000465*
Positive regulation of transcription from RNA polymerase II promoter	*80*	*5*	*10.884418*	*0.000455*
Cell differentiation	*1210*	*38*	*2.590671*	*0.001957*
Cellular developmental process	*1210*	*18*	*2.590671*	*0.001957*
Brain development	*93*	*11*	*10.2294*	*0.00383*
Positive regulation of transcription DNA-dependent	*147*	*6*	*7.108191*	*0.00551*
Metabolic process	*6020*	*50*	*1.446434*	*0.0054*

doi: 10.1038/nature08162

Supplemental Table 13: Gene promoters occupied by Suz12 in ES cell are DNA demethylated and histone bound in sperm

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Regulation of transcription from RNA polymerase II promoter	*297*	*8*	*4.690928*	*0.006923*
Neuron fate specification	*5*	*2*	*69.660274*	*0.009811*
Pattern specification process	*111*	*15*	*7.844625*	*0.012321*
Kidney development	*29*	*3*	*18.015588*	*0.01614*
Tube development	*70*	*4*	*9.951468*	*0.019483*
Urogenital system development	*31*	*3*	*16.853292*	*0.020339*
Positive regulation of cellular metabolic process	*264*	*7*	*4.617632*	*0.020667*
positive regulation of transcription	*192*	*6*	*5.442209*	*0.020984*
Cell fate commitment	*74*	*7*	*8.13551*	*0.021452*
Embryonic limb morphogenesis	*34*	*3*	*15.366237*	*0.023231*
Embryonic appendage morphogenesis	*34*	*3*	*15.366237*	*0.023231*
Organ morphogenesis	*274*	*7*	*4.449105*	*0.022879*
Embryonic development	*199*	*36*	*15.250774*	*0.022206*
Positive regulation of nucleobase nucleoside	*199*	*6*	*5.250774*	*0.022206*
Nunucleotide and nucleic acid metabolic process				
Cell development	*859*	*24*	*2.41575*	*0.022899*
Positive regulation of metabolic process	*280*	*7*	*4.353767*	*0.022571*
Appendage morphogenesis	*36*	*4*	*14.512557*	*0.022055*
Limb morphogenesis	*36*	*4*	*14.512557*	*0.022055*

doi: 10.1038/nature08162

Supplemental Table 14: Promoters that acquire methylation in fibroblasts compared to sperm

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Embryonic development	*199*	*43*	*2.393575*	*0*
Anatomical structure morphogenesis	*730*	*105*	*1.593302*	*0*
Regulation of transcription DNA-dependent	*1467*	*205*	*1.547945*	*0*
Transcription DNA-dependent	*1510*	*208*	*1.525872*	*0*
RNA biosynthetic process	*1512*	*208*	*1.523854*	*0*
Regulation of transcription	*1580*	*216*	*1.514357*	*0*
Transcription	*1644*	*220*	*1.482356*	*0*
RNA metabolic process	*1827*	*241*	*1.461202*	*0*
Multicellular organismal development	*1620*	*201*	*1.374399*	*0*
Tube development	*70*	*20*	*3.164927*	*0.000526*
Negative regulation of cell differentiation	*49*	*15*	*3.390993*	*0.001034*
Transcription from RNA polymerase II promoter	*460*	*69*	*1.661587*	*0.001071*
Negative regulation of developmental process	*60*	*17*	*3.138553*	*0.001111*
Cellular component organization and biogenesis	*1763*	*209*	*1.313184*	*0.00125*
Developmental process	*2265*	*259*	*1.266669*	*0.001304*
System development	*1231*	*155*	*1.394779*	*0.001364*
Anatomical structure development	*1465*	*180*	*1.361027*	*0.001429*
Regulation of cell differentiation	*105*	*23*	*2.426444*	*0.001935*
Embryonic morphogenesis	*78*	*19*	*2.698303*	*0.002*
Organ morphogenesis	*274*	*45*	*1.819255*	*0.0025*
Nervous system development	*553*	*77*	*1.542401*	*0.003529*
Lung development	*32*	*11*	*3.807803*	*0.003636*
Synapse organization and biogenesis	*23*	*9*	*4.334574*	*0.003889*
Respiratory tube development	*33*	*11*	*3.692415*	*0.004*
Nucleosome assembly	*41*	*12*	*3.24212*	*0.005405*
Chromosome organization and biogenesis	*236*	*38*	*1.783624*	*0.005417*
Tube morphogenesis	*43*	*12*	*3.091324*	*0.005532*
Calcium-dependent cell-cell adhesion	*20*	*8*	*4.430898*	*0.005641*
Formation of primary germ layer	*21*	*8*	*4.219903*	*0.005652*
Gastrulation	*31*	*10*	*3.573305*	*0.005682*
Regulation of developmental process	*162*	*29*	*1.982963*	*0.005714*
Organ development	*869*	*108*	*1.376689*	*0.005778*
Branching morphogenesis of a tube	*30*	*10*	*3.692415*	*0.005789*
Chromosome organization and biogenesis	*226*	*37*	*1.813531*	*0.005814*
Chromatin assembly or disassembly	*87*	*19*	*2.419168*	*0.005854*
Macromolecular complex assembly	*359*	*53*	*1.635359*	*0.006*
Pattern specification process	*111*	*22*	*2.19549*	*0.006327*
Morphogenesis of a branching structure	*32*	*10*	*3.461639*	*0.0064*
Mesoderm morphogenesis	*22*	*8*	*4.028089*	*0.008039*
Protein-DNA complex assembly	*86*	*18*	*2.318493*	*0.008393*
Cellular component assembly	*389*	*55*	*1.566191*	*0.008462*
Positive regulation of cell differentiation	*28*	*9*	*3.560543*	*0.008491*
Regionalization	*72*	*16*	*2.46161*	*0.008545*

doi: 10.1038/nature08162

Supplemental Table 14: Promoters that acquire methylation in fibroblasts compared to sperm

GO CATEGORY	TOTAL GENES	CHANGED GENES	ENRICHMENT	FDR
Chromatin assembly	*52*	*13*	*2.769311*	*0.008704*
Synaptogenesis	*18*	*7*	*4.307817*	*0.009298*
DNA packaging	*180*	*30*	*1.846207*	*0.009483*
Embryonic arm morphogenesis	*3*	*3*	*11.077244*	*0.012623*
Arm morphogenesis	*3*	*3*	*11.077244*	*0.012623*
Positive regulation of osteoblast differentiation	*3*	*3*	*11.077244*	*0.012623*
Heart development	*75*	*16*	*2.363145*	*0.015625*
Response to hypoxia	*19*	*7*	*4.08109*	*0.016*
Chordate embryonic development	*69*	*15*	*2.408097*	*0.016866*
Anatomical structure formation	*122*	*22*	*1.997536*	*0.018169*
Mesoderm formation	*20*	*7*	*3.877035*	*0.018378*
Embryonic development ending in birth or egg hatching	*70*	*15*	*2.373695*	*0.018429*
Mitotic sister chromatid segregation	*15*	*6*	*4.430898*	*0.018933*
Sensory organ development	*57*	*13*	*2.526389*	*0.019211*
Mesoderm development	*44*	*11*	*2.769311*	*0.019221*
Cell cycle phase	*214*	*33*	*1.708173*	*0.019367*
Cell differentiation	*1210*	*138*	*1.263355*	*0.019634*
Cell cycle	*606*	*76*	*1.389225*	*0.020814*
Sister chromatid segregation	*16*	*6*	*4.153967*	*0.025057*
Cell fate determination	*27*	*8*	*3.282146*	*0.025455*
Protein catabolic process	*185*	*29*	*1.736433*	*0.025495*
Regulation of transcription from RNA polymerase II promoter	*297*	*42*	*1.566479*	*0.025556*
Cell fate commitment	*74*	*15*	*2.245387*	*0.02573*
Macromolecule catabolic process	*326*	*45*	*1.529067*	*0.025895*
Embryonic limb morphogenesis	*34*	*9*	*2.932212*	*0.025957*
Embryonic appendage morphogenesis	*34*	*9*	*2.932212*	*0.025957*
Dorsal ventral pattern formation	*22*	*7*	*3.524578*	*0.025978*
Regulation of gliogenesis	*4*	*3*	*8.307933*	*0.033469*
RNA interference	*4*	*3*	*8.307933*	*0.033469*
Regulation of glial cell differentiation	*4*	*3*	*8.307933*	*0.033469*
Tissue morphogenesis	*55*	*12*	*2.416853*	*0.0372*
Mesodermal cell fate commitment	*8*	*4*	*5.538622*	*0.04181*
Mitotic chromosome condensation	*8*	*4*	*5.538622*	*0.04181*
Pancreas development	*8*	*4*	*5.538622*	*0.04181*
Mesodermal cell differentiation	*8*	*4*	*5.538622*	*0.04181*
Appendage morphogenesis	*36*	*9*	*2.769311*	*0.042364*
Cell cycle process	*530*	*66*	*1.37943*	*0.044123*
Chromosome condensation	*13*	*5*	*4.260479*	*0.044554*
Positive regulation of developmental process	*43*	*10*	*2.576103*	*0.044696*
Anterior posterior pattern formation	*44*	*10*	*2.517556*	*0.048103*
Sex differentiation	*66*	*13*	*2.181881*	*0.050168*
Embryonic pattern specification	*25*	*7*	*3.101628*	*0.050339*

CHAPTER 3

ALTERATIONS IN SPERM DNA METHYLATION PATTERNS AT IMPRINTED LOCI IN TWO CLASSES OF INFERTILITY

Reprinted with permission from Fertility and Sterility. Hammoud SS, Purwar J, Pflueger C, Cairns BR, Carrell DT (2010) Alterations in sperm DNA methylation patterns at imprinted loci in two classes of infertility. Fertility and Sterility 94: 1728-1733.

Chapter 3 is a published article. Both Saher Sue Hammoud and Jahnvi Purwar contributed to this work equally.

Alterations in sperm DNA methylation patterns at imprinted loci in two classes of infertility

Saher Sue Hammoud, B.S.,[a,b] *Jahnvi Purwar, B.S.,*[d] *Christian Pflueger, M.S.,*[d] *Bradley R. Cairns, Ph.D.,*[d,e] *and Douglas T. Carrell, Ph.D.*[a,b,c]

[a] Andrology and IVF Laboratories, Division of Urology, Department of Surgery, [b] Department of Physiology, and [c] Department of Obstetrics and Gynecology, School of Medicine, and [d] Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah; and [e] Howard Hughes Medical Institute, University of Utah School of Medicine, Salt Lake City, Utah

Objective: To evaluate the associations between proper protamine incorporation and DNA methylation at imprinted loci.
Design: Experimental research study.
Setting: Research laboratory.
Patient(s): Three populations were tested—abnormal protamine patients, oligozoospermic patients, and fertile donors.
Intervention(s): The CpG methylation patterns were examined at seven imprinted loci sequenced: *LIT1, MEST, SNRPN, PLAGL1, PEG3, H19,* and *IGF2.*
Main Outcome Measure(s): The DNA methylation patterns were analyzed using bisulfite sequencing. The percentage of methylation was compared between fertile and infertile patients displaying abnormal protamination.
Result(s): At six of the seven imprinted genes, the overall DNA methylation patterns at their respective differentially methylated regions were significantly altered in both infertile patient populations. When comparing the severity of methylation alterations among infertile patients, the oligozoospermic patients were significantly affected at mesoderm-specific transcript (*MEST),* whereas abnormal protamine patients were affected at *KCNQ1,* overlapping transcript 1 *(LIT1),* and at small nuclear ribonucleoprotein polypeptide N *(SNRPN).*
Conclusion(s): Patients with male factor infertility had significantly increased methylation alteration at six of seven imprinted loci tested, with differences in significance observed between oligozoospermic and abnormal protamine patients. This could suggest that risk of transmission of epigenetic alterations may be different with diagnoses. However, this study does not provide a causal link for epigenetic inheritance of imprinting diseases, but does show significant association between male factor infertility and alterations in sperm DNA methylation at imprinted loci. (Fertil Steril® 2010;94:1728–33. ©2010 by American Society for Reproductive Medicine.)

Key Words: Imprinting, Beckwith-Wiedemann syndrome and epigenetic alterations, Angelman syndrome, chromatin, assisted reproductive technology, IVF, ICSI, oligozoospermic, protamines

Genomic imprinting is established and inherited during gametogenesis and preimplantation to ensure parent-of-origin monoallelic gene expression (1, 2). The mechanism by which either one of the two alleles are differentially expressed is not completely understood; however, it is known that the majority of imprinted genes are clustered and are predominately regulated by imprinting control regions (ICRs) (3, 4). At present, approximately 80 imprinted genes have been identified, many of which are implicated in tumorigenesis, fetal growth regulation, and embryonic development (5–8). Pathological perturbation in the methylation imprints during gametogenesis or development can give rise to growth-related syndromes and is frequently observed in cancer (9–20).

After fertilization, both parental genomes are globally demethylated through active or passive demethylation mechanisms, whereas the methylation patterns at imprinted genes are maintained and only erased and re-established in the primordial germ cell. The presence of abnormal methylation patterns residing in gametes raises concerns, as these may be inherited and maintained in the embryo. Meta-analysis showed that children born from assisted reproductive technology (ART) have a fourfold increased incidence of Beckwith-Weidemann syndrome compared with children conceived naturally (21–24). In addition, imprinting syndromes such as Angelman, Prader-Willi, and Silver-Russell have been associated with ART, although no strong correlations were established. Currently, it is unclear whether imprinting abnormalities arise from the ART procedure itself or from pre-existing methylation aberrations in the gametes of infertile patients (25–27).

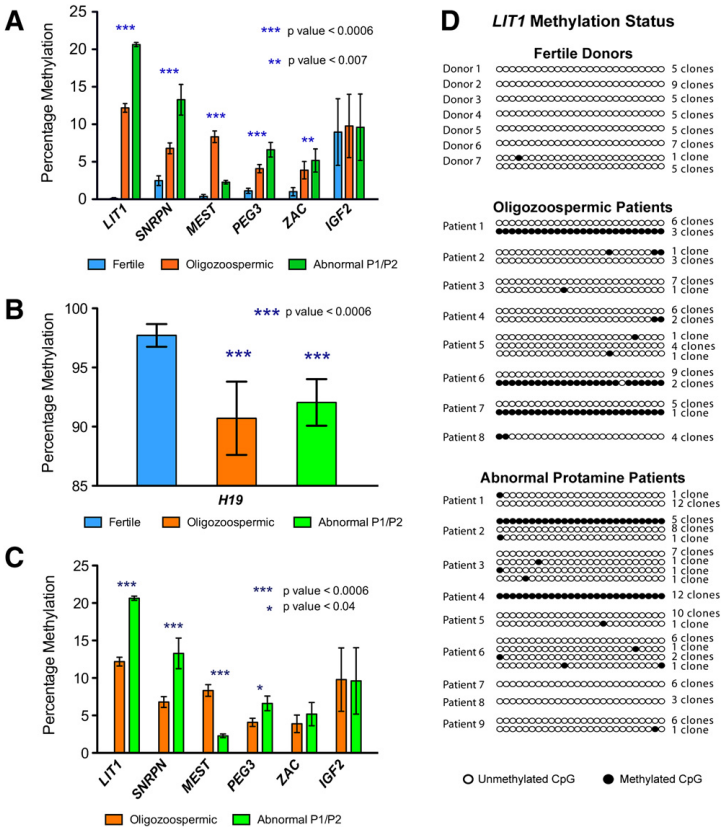
Recent studies have shown that epigenetic abnormalities are common in the sperm of severely oligozoospermic patients, favoring the latter hypothesis (26, 27). Whether epigenetic alterations at imprinted loci of infertile men are limited to oligozoospermic patients or whether epigenetic alterations extend beyond oligozoospermic patients is unknown. In this study we examine methylation changes in patients with an alternative cause for their male factor infertility—patients with abnormal sperm protamine replacement of histones. Protamines 1 and 2 are sperm-specific nuclear proteins that are incorporated into the DNA in a 1:1 ratio and ensure chromatin

Received April 8, 2009; revised August 5, 2009; accepted September 8, 2009; published online November 1, 2009.
S.S.H. has nothing to disclose. J.P. has nothing to disclose. C.P. has nothing to disclose. B.R.C. has nothing to disclose. D.T.C. has nothing to disclose.
Saher Sue Hammoud and Jahnvi Purwar contributed equally.
Reprint requests: Douglas T. Carrell, Ph.D., University of Utah IVF and Andrology Laboratories, 675 Arapeen Drive, Suite 205, Salt Lake City, UT 84108 (FAX: 801-581-6127; E-mail: douglas.carrell@hsc.utah.edu).

0015-0282/$36.00
doi:10.1016/j.fertnstert.2009.09.010

FIGURE 1

The overall methylation patterns at both paternally and maternally imprinted genes were altered in the sperm of infertile patients. (**A,B,C**) The mean percentage of methylation with standard error. $P<.05$ is significant. (**A**) The percentage of methylated CpGs at normally paternally demethylated loci. (**B**) The percentage of demethylation at a paternally methylated DMR of *H19*. (**C**) Comparing methylation changes between the two infertile patient populations. (**D**) Methylation status at the differentially methylated region of *LIT1* for fertile donors, oligozoospermic patients, and abnormal protamine patients.



Hammoud. Imprinting abnormalities in infertile men. Fertil Steril 2010.

condensation. The average P1:P2 ratio in fertile men is ~1, whereas in some infertile patients this ratio is significantly altered (28, 29) and consequently associated with severe sperm defects that can usually be addressed through ART (30, 31). It has been proposed that chromatin packaging may have a role in properly establishing and maintaining methylation patterns, hence, hypothetically, patients with abnormal protamine ratios may be at an increased risk of conceiving an ART offspring with imprinting disease (32, 33). This study evaluates the relationship between protamine ratios and methylation patterns at seven imprinted loci in the sperm of abnormal protamine patients or oligozoospermic patients. We reveal significant changes in the overall DNA methylation patterns at six of these loci, with varying impact on methylation patterns within each class of infertility: oligozoospermic or abnormal protamine levels (p-value < 0.05, Figure 1). These data suggest that aberrant imprinting patterns are observed in patients with abnormal protamine ratios, and that the abnormal patterns may vary among different pathologies, providing a spectrum of risks for transmitting epigenetic abnormalities to the embryo.

MATERIALS AND METHODS

Patient Population

Of the seven tested imprinted loci, six are paternally demethylated and expressed: KCNQ1 overlapping transcript 1 (*LIT1*), insulin-like growth factor 2 (*IGF2*), paternally expressed gene 3 (*PEG3*), pleiomorphic adenoma gene-like 1 (*PLAGL1* also known as *ZAC*), small nuclear ribonucleoprotein

polypeptide N *(SNRPN),* and mesoderm-specific trancript (*MEST*), and one is maternally expressed and is normally DNA methylated in sperm (*H19*). For each locus 10 oligozoospermic (sperm count $\leq 10 \times 10^6$/mL), 10 abnormal protamine replacement patients (average sperm count of $73 \times 10^6 \pm 60$ SD/mL), and 5 known fertile donors were evaluated. For *LIT1* only, eight oligozoospermic patients and nine abnormal protamine patients were evaluated.

Sample Collection and Bisulfite Treatment

Institutional Review Board (IRB) approval was obtained before initiation of this study. Frozen sperm DNA samples were treated with sodium bisulfite to convert unmethylated cytosines to uracil and leaving methylated cytosines unchanged, as previously described by Clark et al. (34). DNA was purified using Qiagen DNeasy clean up kit (Qiagen, Valencia, CA) and eluted twice, each time with 100 μL of elution buffer. The purified DNA was desulfonated by the addition of 20 μL NaOH and incubated at 37°C for 15 minutes. After incubation, 22 μL of 4 M NaOAC, glycogen, and two volumes of ethanol were added to precipitate the DNA overnight at -20°C. Precipitated DNA was washed twice with 70% ethanol and eluted in 30 μL of elution buffer.

PCR Amplification of Bisulfite Converted DNA

Primer sequences and temperatures for *SNRPN, PEG3, ZAC, MEST, LIT1, H19 ICR*, and *IGF2* are available upon request (35, 36). The polymerase chain reaction (PCR) reactions were performed in 50-μL volume reactions containing 5 μL of 10 × PCR buffer–$MgCl_2$ (Invitrogen, Carlsbad, CA), 5 μL of 10 × Enhancer Buffer (Invitrogen), 1.5 μL of $MgCl_2$, 1 μL of 10 mM dNTPs, 0.5 μL of *Taq* (Invitrogen), 2.5 μL of each forward and reverse primer (10 μM stock), and 30 μL of water. The PCR results were analyzed on a 1% agarose gel, and gel purified if multiple products were detected.

TOPO TA Cloning and Sequencing

The PCR products were cloned into a TOPO 2.1 pCR vectors (Invitrogen) and plated onto KAN-X-GAL plates for blue-white screening. Positive colonies were reinoculated into LB-KAN (50 μg/mL), cultured overnight, and plasmids were purified using the Qiagen 96-well clean-up kit. To address sperm sample heterogeneity five or more clones/alleles were sequenced per patient for each of the imprinted loci (sequencing done at Genewiz San Diego Laboratory).

Data Visualization and Analysis

The CG/TG-analyzer, a Perl program, was used to examine the methylation status of a bisulfite-converted sequence and provides an output in the form of 1s and 0s, where 1s represent methylated cytosines and 0s represent unmethylated cytosines (thymine). The CpG positions were defined in a multifasta file, text-based file containing multiple DNA or protein sequences, which includes the CpG position number flanked by four nucleotides on each side. The output was used to calculate the percentage of CpG methylation (program is be available upon request). To compare the overall methylation profile in infertile patients versus fertile donors (Fig. 1), the Wilcoxon-Mann-Whitney test was used. This test is a nonparametric significance test for assessing whether two independent samples of observations came from the same distribution. To determine significance between fertile donors and oligozoospermic patients or fertile and abnormal protamine patients the percentage of methylated CpGs represented in columns 2 and 3 (in Tables 1, 2, and 3) were compared as independent sample populations. A P value $< .05$ was considered significant. The χ^2 analysis was used to compare the percentage of methylated CpGs in the abnormal protamine or oligozoospermic patients with known fertile donors.

RESULTS

Six imprinted genes, that are normally paternally demethylated, were examined: *LIT1, SNRPN, MEST, ZAC, PEG3,* and *IGF2*. Here, all except *IGF2,* showed significant hypermethylation in oligozoospermic and abnormal protamine patients compared with fertile donors (Fig. 1A). Furthermore, the differentially methylated region (DMR) of *H19* (a paternally methylated locus) was

TABLE 1

The percentage of methylated CpGs in the DMR of *LIT1* of oligozoospermic and abnormal protamine patients.

CpG	Abnormal P1/P2 (n = 9)	Oligozoospermic (n = 8)	Fertile donors (n = 7)	Fertile vs. abnormal	Fertile vs. oligozoospermic
CpG 1	25.882	18.181	0	0.0003	0.0035
CpG 2	20	18.181	0	0.0021	0.0035
CpG 3	20	10.909	0	0.0021	0.0271
CpG 4	20	10.909	2.38	0.0066	0.17
CpG 5	21.176	10.909	0	0.0015	0.0271
CpG 6	20	10.909	0	0.0021	0.0271
CpG 7	21.176	10.909	0	0.0015	0.0271
CpG 8	20	10.909	0	0.0021	0.0271
CpG 9	20	10.909	0	0.0021	0.0271
CpG 10	20	10.909	0	0.0021	0.0271
CpG 11	21.176	12.277	0	0.0015	0.0186
CpG 12	20	10.909	0	0.0021	0.0271
CpG 13	21.176	10.909	0	0.0015	0.0271
CpG 14	20	14.454	0	0.0021	0.0101
CpG 15	20	10.909	0	0.0021	0.0271
CpG 16	20	7.272	0	0.0021	0.0742
CpG 17	20	10.909	0	0.0021	0.0271
CpG 18	21.176	12.272	0	0.0015	0.0093
CpG 19	20	10.909	0	0.0021	0.0271
CpG 20	20	10.909	0	0.0021	0.0271
CpG 21	21.176	16.363	0	0.0015	0.0059
CpG 22	21.176	16.363	0	0.0015	0.0059

Note: DMR = differentially methylated region.

Hammoud. Imprinting abnormalities in infertile men. Fertil Steril 2010.

TABLE 2

The percentage of methylated CpG in the DMR of *SNRPN*.

CpGs	Abnormal P1/P2 (n = 11)	Oligozoospermic (n = 13)	Fertile donors (n = 5)	Fertile vs. abnormal	Fertile vs. oligozoospermic
CpG 1	4.3	4.0	0	0.152	0.169
CpG 2	5.8	5.0	0	0.09	0.123
CpG 3	5.8	5.0	0	0.09	0.123
CpG 4	5.8	5.0	0	0.09	0.123
CpG 5	5.8	5.0	0	0.09	0.123
CpG 6	10.6	5.0	0	0.026	0.123
CpG 7	14.8	8.0	4.3	0.08	0.413
CpG 8	8.7	5.0	0	0.04	0.123
CpG 9	13.0	5.0	4.3	0.10	0.864
CpG 10	23.1	16	6.5	0.05	0.114
CpG 11	10.1	6.0	0	0.026	0.09
CpG 12	11.6	6.1	8.7	0.618	0.526
CpG 13	15.9	8.0	6.5	0.1	0.753
CpG 14	47.8	10	2.2	0.0001	0.09
CpG 15	11.6	6.1	0	0.017	0.08
CpG 16	5.8	4.0	2.2	0.351	0.566
CpG 17	11.6	13	2.2	0.065	0.039
CpG 18	15.9	12.2	6.5	0.130	0.295
CpG 19	15.9	5.1	6.7	0.140	0.705
CpG 20	17.4	5.1	0	0.003	0.119
CpG 21	17.6	4.1	2.2	0.011	0.560

Note: DMR = differentially methylated region.

Hammoud. Imprinting abnormalities in infertile men. Fertil Steril 2010.

significantly hypomethylated in both infertile classes ($P<.006$ for all except *ZAC*, $P<.002$) (Fig. 1B). Thus, these infertile patients show methylation alterations at six of seven loci tested. However, when comparing overall methylation changes between the two infertile populations, abnormal protamine patients show more extensive hypermethylation at the DMRs of *LIT1* and *SNRPN* in comparison with oligozoospermic patients. In contrast, hypermethylation at *MEST* is significantly higher in oligozoospermic patients (p-value < 0.006, Fig. 1C).

Notably, in both patient populations, the locus that displays the highest number of affected CpGs is *LIT1*. In the DMR of *LIT1*, the percentage of methylated CpGs ranged from 7%–18% or 20%–25% for oligozoospermic or abnormal protamine patients, respectively (Table 1). In contrast, for fertile donors, virtually all CpGs were demethylated. The percentages of methylated CpGs in oligozoospermic and abnormal protamine patients were statistically significant when compared with fertile donors (p-value < 0.05, Table 1). To address the uniformity of methylation changes at *LIT1* in individual sperm from a single patient, we sequenced multiple alleles (5–12) from each patient, and found striking heterogeneity. In three of the eight oligozoospermic patients, *LIT1* was completely methylated in 20%–30% of the alleles, whereas in the other five patients, only sporadic increases were observed (Fig. 1D). Similarly, in the abnormal protamine category one patient always displayed complete methylation, a second displayed methylation on 50% of his alleles, and the remainder (seven) displayed little or no increase.

Consistent with the findings reported previously, the DMR of *SNRPN* was also susceptible to acquiring methylation in infertile men. Abnormal protamine patients had a significant increase in CpG methylation (methylation at individual CpGs typically ranged from 4%–20%) (p-value < 0.05 Table 2). Alterations were also observed in oligozoospermic patients (range of methylation, 4%–8%), but the increase lacked statistical significance (Table 2). At *SNRPN*, alterations in methylation were common (observed at a majority of the alleles) but typically involved only a moderate number of CpGs acquiring methylation. However, in both patient categories, a small number of patients displayed complete methylation at 10% of the alleles tested.

Methylation levels in the DMR of *MEST* (for each CpG) ranged from 7%–19% or 1%–3% in oligozoospermic or abnormal protamine patients, respectively (Table 3). The changes in methylation at many of the CpGs in oligozoospermic patients were near the range of statistical significance ($P=.07$; Table 3). In addition, 3 of 10 oligozoospermic patients had 12%–33% of their alleles completely methylated, whereas the remaining 7 patients displayed very little change. Likewise, in the abnormal protamine class, one patient had 14% of his alleles completely methylated and in the remaining nine patients, there was virtually no change observed. In contrast, very few individual CpGs were significantly ($P<.05$) affected in *PEG3, ZAC*, *IGF2* promoter 3, and *H19* in infertile patients (data not shown).

DISCUSSION

In this study we evaluated the methylation status of seven imprinted loci in two patient populations: oligozoospermic and abnormal protamine ratio patients. The overall methylation patterns in sperm of infertile patients were significantly altered at all imprinted loci (except *IGF2*) when compared with fertile donors. However, when comparing the two infertile patient populations, oligozoospermic patients were hypermethylated at *MEST*, an imprinted gene associated with Silver-Russell syndrome, whereas abnormal protamine patients had significant changes at *LIT1* and *SNRPN* (Figure 1), genes that may be associated with cases of transient neonatal diabetes mellitus and Angelman syndrome. These data suggest that risk of transmission of epigenetic alterations may be different with diagnoses.

TABLE 3

The percentage of methylated CpGs at the DMR of *MEST* in oligozoospermic and abnormal protamine patients.

CpG	Abnormal P1/P2 (n = 10)	Oligozoospermic (n = 10)	Fertile donors (n = 5)	Fertile vs. abnormal	Fertile vs. oligozoospermic
CpG 1	1.785	14.28	0	0.2346	0.0167
CpG 2	1.785	19.04	0	0.2346	0.0063
CpG 3	3.571	7.1428	0	0.1515	0.070
CpG 4	3.571	7.142	3.4	0.483	0.250
CpG 5	1.785	7.1428	0	0.2346	0.070
CpG 6	1.785	9.5238	0	0.2346	0.436
CpG 7	3.571	7.1428	0	0.1515	0.070
CpG 8	1.785	7.1428	0	0.2346	0.070
CpG 9	1.785	7.1428	0	0.2346	0.070
CpG 10	1.785	7.1428	0	0.2346	0.070
CpG 11	1.785	7.1428	0	0.2346	0.070
CpG 12	3.571	7.1428	0	0.1515	0.070
CpG 13	1.785	9.523	3.4	0.642	0.1604
CpG 14	3.571	4.7619	0	0.1515	0.1167
CpG 15	1.785	7.1428	0	0.2346	0.070
CpG 16	3.571	7.1428	0	0.1515	0.070
CpG 17	1.785	7.1428	0	0.2346	0.070
CpG 18	0	7.1428	0	NA	0.070

Note: DMR = differentially methylated region.

Hammoud. Imprinting abnormalities in infertile men. Fertil Steril 2010.

Our data evaluate and demonstrate abnormal imprinting in a different class of abnormal spermatogenesis, abnormal replacement of nuclear proteins by protamine 1 and protamine 2. It was our hypothesis that abnormal chromatin packaging may be associated with methylation defects, which is supported by the data presented from this study. These data, along with previously published data from oligozoospermic patients, reveal that alteration in DNA methylation patterns are common at a handful of imprinted loci tested, suggesting that imprinting abnormalities may reside in the sperm of infertile patients (25–27), but whether these alterations can be inherited is uncertain. Remarkably, when examining normally demethylated DMRs, the alleles of infertile patients are often either unaffected or entirely methylated, suggesting a bistable status, and a susceptibility to complete methylation. Clearly, complete methylation of a normally unmethylated locus may lead to an imprinting disorder in the embryo if proper imprint reestablishment mechanisms are not implemented. Also of note are the small differences in the degree of methylation within some genes and alleles. It is important to determine whether this abnormal methylation has reached a threshold level that might lead to complete methylation in the embryo (at a certain unknown probability) and confer disease, or whether there is a gradual continuum with a threshold for disease.

Whether imprinting diseases in ART offspring arise as a result of abnormal methylation of gametes, or acquire methylation changes during in vitro culture, or both, is still unknown. Current human data suggest that methylation alteration at imprinted loci may reside in gametes and may be inherited by the embryo. Supporting evidence comes from two reports showing that a gain in methylation on the paternal alleles of *LIT1* or *MEST* in sperm is maintained in the baby and associated with transient neonatal diabetes (37) or Silver-Russell syndrome (38). The findings suggest that paternal imprints in sperm may be needed for a healthy and uncomplicated pregnancy. The need to study sperm from fathers of children with imprinting diseases is imperative.

This study does not report a causal link between abnormal methylation of imprinted genes and disease. The relative risk of the defects reported in our study to patients is unknown. However, we demonstrate a link between abnormal spermatogenesis and abnormal methylation of genes associated with rare imprinting diseases previously reported to have elevated incidences in ART offspring (21–24). This suggests that such a link may be strengthened in infertile men with known abnormalities in chromatin packaging. Characterizing these epigenetic alterations in the sperm of infertile men may help predict the likelihood of IVF success rate.

REFERENCES

1. Morison IM, Paton CJ, Cleverley SD. The imprinted gene and parent-of-origin effect database. Nucleic Acids Res 2001;29:275–6.
2. Ferguson-Smith AC, Surani MA. Imprinting and the epigenetic asymmetry between parental genomes. Science 2001;293:1086–9.
3. Verona RI, Mann MR, Bartolomei MS. Genomic imprinting: intricacies of epigenetic regulation in clusters. Annu Rev Cell Dev Biol 2003;19:237–59.
4. Spahn L, Barlow DP. An ICE pattern crystallizes. Nat Genet 2003;35:11–2.
5. Morgan HD, Santos F, Green K, Dean W, Reik W. Epigenetic reprogramming in mammals. Hum Mol Genet 2005. 14 Spec No 1: R47–R58.
6. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 2003;33. Suppl: 245–54.
7. Royo H, Bortolin ML, Seitz H, Cavaille J. Small noncoding RNAs and genomic imprinting. Cytogenet Genome Res 2006;113:99–108.
8. Seitz H, Royo H, Bortolin ML, Lin SP, Ferguson-Smith AC, Cavaille J. A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. Genome Res 2004;14:1741–8.
9. Bjornsson HT, Brown LJ, Fallin MD, Rongione MA, Bibikova M, Wickham E, et al. Epigenetic specificity of loss of imprinting of the IGF2 gene in Wilms tumors. J Natl Cancer Inst 2007;99:1270–3.
10. Chao W, D'Amore PA. IGF2: epigenetic regulation and role in development and disease. Cytokine Growth Factor Rev 2008;19:111–20.

11. Cui H. Loss of imprinting of IGF2 as an epigenetic marker for the risk of human cancer. Dis Markers 2007;23:105–12.
12. Hemberger M. Epigenetic landscape required for placental development. Cell Mol Life Sci 2007;64: 2422–36.
13. Henckel A, Feil R. Differential epigenetic marking on imprinted genes and consequences in human diseases. Med Sci (Paris) 2008;24:747–52.
14. Lemeta S, Jarmalaite S, Pylkkanen L, Bohling T, Husgafvel-Pursiainen K. Preferential loss of the nonimprinted allele for the ZAC1 tumor suppressor gene in human capillary hemangioblastoma. J Neuropathol Exp Neurol 2007;66:860–7.
15. Murrell A, Ito Y, Verde G, Huddleston J, Woodfine K, Silengo MC, et al. Distinct methylation changes at the IGF2-H19 locus in congenital growth disorders and cancer. PLoS ONE 2008;3:e1849.
16. Nakano S, Murakami K, Meguro M, Soejima H, Higashimoto K, Urano T, et al. Expression profile of LIT1/KCNQ1OT1 and epigenetic status at the KvDMR1 in colorectal cancers. Cancer Sci 2006;97:1147–54.
17. Nowaczyk MJ, Carter MT, Xu J, Huggins M, Raca G, Das S, et al. Paternal deletion 6q24.3: a new congenital anomaly syndrome associated with intrauterine growth failure, early developmental delay and characteristic facial appearance. Am J Med Genet A 2008;146:354–60.
18. Shin JY, Fitzpatrick GV, Higgins MJ. Two distinct mechanisms of silencing by the KvDMR1 imprinting control region. EMBO J 2008;27:168–78.
19. Valleley EM, Cordery SF, Bonthron DT. Tissue-specific imprinting of the ZAC/PLAGL1 tumour suppressor gene results from variable utilization of monoallelic and biallelic promoters. Hum Mol Genet 2007;16:972–81.
20. Yoshimizu T, Miroglio A, Ripoche MA, Gabory A, Vernucci M, Riccio A, et al. The H19 locus acts in vivo as a tumor suppressor. Proc Natl Acad Sci U S A 2008;105:12417–22.
21. Cox GF, Burger J, Lip V, Mau UA, Sperling K, Wu BL, et al. Intracytoplasmic sperm injection may increase the risk of imprinting defects. Am J Hum Genet 2002;71:162–4.
22. Niemitz EL, DeBaun MR, Fallon J, Murakami K, Kugoh H, Oshimura M, et al. Microdeletion of LIT1 in familial Beckwith-Wiedemann syndrome. Am J Hum Genet 2004;75:844–9.
23. DeBaun MR, Niemitz EL, Feinberg AP. Association of in vitro fertilization with Beckwith-Wiedemann syndrome and epigenetic alterations of LIT1 and H19. Am J Hum Genet 2003;72:156–60.
24. Gosden R, Trasler J, Lucifero D, Faddy M. Rare congenital disorders, imprinted genes, and assisted reproductive technology. Lancet 2003;361: 1975–7.
25. Kobayashi H, Sato A, Otsu E, Hiura H, Tomatsu C, Utsunomiya T, et al. Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. Hum Mol Genet 2007;16:2542–51.
26. Marques CJ, Carvalho F, Sousa M, Barros A. Genomic imprinting in disruptive spermatogenesis. Lancet 2004;363:1700–2.
27. Marques CJ, Costa P, Vaz B, Carvalho F, Fernandes S, Barros A, et al. Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. Mol Hum Reprod 2008;14:67–74.
28. Aoki VW, Emery BR, Liu L, Carrell DT. Protamine levels vary between individual sperm cells of infertile human males and correlate with viability and DNA integrity. J Androl 2006;27:890–8.
29. Balhorn R, Reed S, Tanphaichitr N. Aberrant protamine 1/protamine 2 ratios in sperm of infertile human males. Experientia 1988;44:52–5.
30. Aoki VW, Moskovtsev SI, Willis J, Liu L, Mullen JB, Carrell DT. DNA integrity is compromised in protamine-deficient human sperm. J Androl 2005;26: 741–8.
31. Aoki VW, Liu L, Jones KP, Hatasaka HH, Gibson M, Peterson CM, et al. Sperm protamine 1/protamine 2 ratios are related to in vitro fertilization pregnancy rates and predictive of fertilization ability. Fertil Steril 2006;86:1408–15.
32. Paldi A. Genomic imprinting: could the chromatin structure be the driving force? Curr Top Dev Biol 2003;53:115–38.
33. Rousseaux S, Caron C, Govin J, Lestrat C, Faure AK, Khochbin S. Establishment of male-specific epigenetic information. Gene 2005;345:139–53.
34. Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res 1994;22:2990–7.
35. El-Maarri O, Seoud M, Coullin P, Herbiniaux U, Oldenburg J, Rouleau G, et al. Maternal alleles acquiring paternal methylation patterns in biparental complete hydatidiform moles. Hum Mol Genet 2003;12:1405–13.
36. Kamikihara T, Arima T, Kato K, Matsuda T, Kato H, Douchi T, et al. Epigenetic silencing of the imprinted gene ZAC by DNA methylation is an early event in the progression of human ovarian cancer. Int J Cancer 2005;115:690–700.
37. Arima T, Kamikihara T, Hayashida T, Kato K, Inoue T, Shirayoshi Y, et al. ZAC, LIT1 (KCNQ1OT1) and p57KIP2 (CDKN1C) are in an imprinted gene network that may play a role in Beckwith-Wiedemann syndrome. Nucleic Acids Res 2005;33:2650–60.
38. Kagami M, Nagai T, Fukami M, Yamazawa K, Ogata T. Silver-Russell syndrome in a girl born after in vitro fertilization: partial hypermethylation at the differentially methylated region of PEG1/MEST. J Assist Reprod Genet 2007;24: 131–6.

CHAPTER 4

MUTATIONS IN HISTONE DEMETHYLASES AND HISTONE METHYLTRANFERASES IN SDH DEFICIENT AND SDH PRESENT PARAGANGLIOMAS MAY EXPLAIN SIMILAR EPIGENETIC MISREGULATION

Chapter 4 is a manuscript in preparation for publication. The authors of this manuscript are Jahnvi Pflüger, Christian Pflüger, Clint Mason, Ashley Chan, Rajini Srinivasan, Joanna Wysocka, Joshua Schiffman and Bradley Cairns.

4.1 Abstract

Depending on whether tumors harbor mutations in the succinate dehydrogenase (SDH) pathway, paragangliomas (PGLs), which are neuroendocrine extra-adrenal tumors, are classified as either SDH Present, or SDH Deficient. To determine the mechanism underlying phenotypic difference between SDH Present and SDH Deficient tumors previous studies have compared the two sub-types directly. Although this method is quite useful for identifying features that distinguish each subtype, commonalities among tumor types, which are likely to cause transformation, are undetectable making it impossible to address important questions relating to paraganglioma tumorigenesis. To identify commonalities among paraganglioma tumor subtypes, and to detect differences that may underlie tumor formation, we profiled genome-wide changes in DNAme in both subclasses as compared to a progenitor cell type, neural crest cells (NCCs). Remarkably, we find that the two subclasses of PGLs have similar DNAme patterns, at regions including CpG Island and Shore promoters. They are hypermethylated and repressed at epigenetically related enzymes, tumor suppressor genes and genes crucial for epithelial-to-mesenchymal transitions (EMT). Interestingly, both SDH Deficient and Present PGLs have very similar transcription profiles when compared to NCCs; genes involved in hypoxia, angiogenesis, and inflammatory response are upregulated, while genes involved in neural crest differentiation, DNA damage response, homologous recombination, and nucleotide excision repair are downregulated. We hypothesized that SDH Present tumors harbored mutations in epigenetic enzymes, and phenocopy the misregulation observed in SDH Deficient tumors – where accumulation of an end product inhibiting alpha-keto-glutarate dependent epigenetic enzymatic activity results in tumorigenesis. To test this hypothesis, we performed whole exome sequencing on SDH Present PGLs and on a subset of SDH Deficient PGLs, with the goal of identifying mutations in novel

candidates and to further understand the genotype of our SDH Present class. Our results revealed mutations in α-KG dependent histone demethylase enzymes such as KDM6B, KDM5C and JMJD4; in histone methyltransferases such as SETD1A, SETD1B, MLL4, NSD1 and PRDM2; in DNMT1 – the maintenance DNA methyltransferase; in JARID2 – which is involved in recruiting PRC2; and in MBD5 – an enzyme that associates with heterochromatin. We speculate that both SDH Deficient and SDH Present PGLs have aberrant epigenetic patterns, resulting from either succinate accumulation or from mutations in enzymes that regulate chromatin structure, leading to aberrant transcriptional activation or repression of genes involved in promoting oncogenesis. Whether these tumors behave in a clinically similar manner is an unknown, making follow up studies relating to patient response, treatment, relapse and outcome, an absolute requirement.

4.2 Introduction

Paragangliomas (PGLs) are neuroendocrine tumors derived from neural crest origin. They are rare, highly vascularized, extra-adrenal tumors that occur at multiple locations along the paravertebral axis; the most common being associated with the carotid body or the glomus (near the middle ear)[1-3]. These tumors can occur sporadically or can be familial, with known germline mutations in susceptibility genes such as RET; NF1 (kinase receptor and signaling regulators); MAX (transcription factors); VHL, EPAS1 (involved in the hypoxia response pathway); TMEM127 (transmembrane protein involved in endosomal signaling) and members of the SDH (succinate dehydrogenase) complex: SDHA, SDHB, SDHC, SDHD; and cofactor SDHAF2 (TCA cycle / energy metabolism)[4]. Hereditary PGLs with germline mutations in the succinate dehydrogenase complex are particularly interesting due to the direct link between the misregulation of metabolism and the resulting alteration of substrate required for the activity of epigenetic enzymes.

Unfortunately, little is known about additional underlying genetic mutations in both SDH-Deficient PGLs and somatic PGLs that are transcriptionally related to SDH Deficient PGLs.

In hereditary paragangliomas, mutations in the SDH complex are commonly observed in one of the catalytic subunits, SDHB and SDHD. However, mutations in the other SDH complex members are also observed in PGLs; albeit at a much lower rate[1,5]. Upon loss of the somatic wildtype allele, the enzyme is rendered inactive due to mutations in the catalytic domain or due to the lack of complex assembly in the mitochondria [1,2]. In both scenarios, succinate is not converted to fumurate at the same rate as in a wildtype cell, which leads to succinate accumulation in the mitochondria[1-3,6]. Following accumulation the metabolite is transported to the cytoplasm, where it competitively inhibits several alpha-keto-glutarate (αKG) dependent enzymes such as Jumonji-histone demethylases (JHDMs) and Ten-eleven Translocases (TETs) family proteins. Furthermore, hydroxylases and prolyl-hydroxylases (PHDs) are also susceptible to succinate inhibition, which has been observed in both *in vitro* and *in vivo* using HEK293T and HeLa cell models[4,7-9]. However, the mechanism for succinate-based inhibition leading to epigenetic misregulation within the context of neural-crest cell derived tumors remains unknown. Here, we investigate how succinate misregulation aberrantly affects the transcriptome of tumor progenitor cells, perhaps allowing for a growth advantage, transformation, and tumorigenesis.

Chromatin structure defines the state in which genetic information is organized within a cell and helps to maintain distinct epigenetic identities. This packaging is achieved by several tightly regulated mechanisms including chromatin remodeling, histone modifications, histone variant incorporation, and covalent modifications on DNA such as 5-methyl-cytosine (5mC) and 5-hydroxymethyl-cytosine (5hmC)[1,5,10]. Failure to properly maintain the correct chromatin structure

can result in inappropriate activation or inhibition of various signaling pathways and can lead to disease states such as cancer[1,2,11].

In general, DNAme at promoters and enhancers is associated with a repressive chromatin state and is known to promote transcriptional repression[12]. In many cancers, genome-wide DNA hypomethylation occurs, and is observed regularly at repetitive elements, retrotransposons, CpG poor promoters and gene deserts. This decrease in DNAme can lead to an increase in genomic instability. Similarly, DNA hypomethylation can lead to the activation of proto-oncogenes, which may provide a growth and survival advantage to the cells. Site-specific hypermethylation may also contribute to the progression of tumorigenesis, by silencing tumor suppressor genes [13,14], which may be involved in an array of cellular processes, including DNA repair, cell adhesion, apoptosis and angiogenesis. Improper function of these processes may lead to cancer initiation and progression. Apart from directly silencing tumor suppressor genes, DNA hypermethylation has been shown to silence transcription factors such as RUNX3 (esophageal cancer), GATA-4 (colorectal cancer) and GATA-5 (gastric cancer). This in turn can lead to silencing of transcription factors gene. Consequently, this may enable cells to accumulate further lesions leading to rapid progression of cancer [15].

Work from several studies[7-9,16,17] lead us to hypothesize that two axes are impacted in PGLs, – a HIF-dependent axis (hypoxia inducible factor) and a chromatin axis. Under normoxic conditions, PHDs hydroxylate two proline residues in the oxygen dependent domain (ODD) of HIF (hypoxia inducible factor), which allows an E3 ubiquitin ligase, VHL, to bind the ODD, destabilizing HIF and targeting it for degradation[18]. However, when succinate accumulates, it competitively inhibits the alpha-ketoglutarate dependent PHDs from acting upon their substrates, deterring VHL from targeting HIF for degradation. As a result, stabilized HIF is allowed to dimerize with its partner, enabling its translocation into the nucleus where it

will transcriptionally upregulate its downstream targets that contribute increased glucose uptake, glycolysis angiogenesis and metastasis[9,19,20]. Along the chromatin axis, succinate accumulation also competitively inhibits JHDMs and TET enzymes, which directly alters the epigenetic profile of cells as well as tumors[9,17,21,22]. These previously published reports have studied the role of these two axes in the context of both *in vitro* and in cell culture systems. Uniquely, we have studied the impact of these axes in patient-derived paraganglioma tumors. Our goal was to identify novel gene candidates involved in the oncogenesis of SDH Deficient and SDH Present PGLs. We achieved this by comparing the methylome and transcriptome of the two tumors subclasses to a progenitor cell type, NCCs. Because uninvolved tissue is not normally collected from patients[23], finding an appropriate control for comparisons was a challenge. Since the majority of the PGLs included in this study are associated with the parasympathetic nervous system, they are derived from the nonchromaffin cells arising from the embryonic neural crest. Hence, we hypothesized that NCCs offered the best available control cell type[24] since they are multipotent and can differentiate into several lineages including peripheral neurons, glia, melanocytes, endocrine cells, chromaffin cells and mesenchymal precursor cells[25] . Our approach to use a progenitor cell type as a comparison is in contrast to a recently published study that compared SDH Deficient PGLs to those with mutations in other known susceptability genes (RET, NF1, and VHL). Hence, their study highlighted differences between two genotypically different but histologically similar tumor subclasses. Their approach did not determine gene candidates involved in driving PGL oncogenesis. In addition, their differential methylation analysis was limited since they only assayed changes in CpG Islands[22]. Recent studies have shown that changes in DNA methylation at CpG Shores (approximately 2kb upstream or downstream from a CpG Island) may also play a significant role in modulating gene expression which in turn may drive

tumorigenesis[12,63]. Hence, to gain insight into changes in DNA methylation on a comprehensive level, we surveyed DMRs in the PGLs that were associated with CpG Islands, CpG Shores, CpG Shelves and Distant Regions.

In addition to SDH Deficient (SDHx) PGLs, we have a subclass of PGLs that do not harbor mutations in any of the known susceptibility genes (VHL, RET, NF1, TMEM127, MAX, EPAS1) and hence will be classified as SDH Present PGLs in this study. This subclass of SDH Present PGLs (sporadic or hereditary) is of great interest to our work, since they are phenotypically very similar to SDH Deficient PGLs, yet they harbor no known SDH related mutations. Notably, several reports in the literature have confirmed that SDH Deficient tumors are transcriptionally more similar to tumors with mutations in VHL and EPAS1; whereas tumors with mutations in RET, NF1, MAX and TMEM127 are more similar to each other [26]. For unknown reasons, sporadic tumors can be transcriptionally related to either group. To identify commonalities among paraganglioma tumor subtypes (SDH Deficient and SDH Present), and to detect differences when compared to a progenitor cell type, that may underlie tumor formation, we performed whole exome sequencing and performed DNAme analysis and correlated these changes with gene expression. Further, we also analyzed published DNAme data and compared their findings to ours[22].

4.3 Materials and Methods

4.3.1 Tumor Samples

In this study, we used tumor samples from patients recruited through the Cancer Genetics Study from 2002 to 2013. Ethical approval for the study was obtained for the institution review board. All patients provided written informed consent for the collection of samples and subsequent use in the study. Table 4.1 details the clinical and genomic characteristics of tumor samples. We used 8 SDH

Deficient tumors and 5 SDH Present tumors for our study. As controls, we used neural crest cells derived from embryonic stem cells as detailed in a previously published study[27].

4.3.2 Immunohistochemistry

Immunohistochemical staining was performed on 4-micron thick sections of formalin-fixed, paraffin-embedded tissues. Sections were air-dried at room temperature and then placed in a 60°C oven for 30 minutes to melt the paraffin. All of the staining steps were performed at 37°C on the automated immunostainer (BenchMarkT Ultra) from Ventana Medical Systems, Tucson, AZ. The sections were de-paraffinized with EZ Prep solution (Ventana Medical Systems). The sections were pretreated with CC1 (Cell Conditioner 1, pH 8.0, Ventana Medical Systems) for 36 minutes at 95°C (for HIF2α & SDHB) or the sections were pre-treated with CC1 (Cell Conditioner 1, pH 8.0) for 64 minutes at 95°C (HIF1α). Rabbit polyclonal antibodies (HIF2α, Lifespan Biosciences, Seattle, WA, catalog # LS-B517/39990; SDHB, Sigma Aldrich, St. Louis, MO, catalog # HPA002868) and mouse monoclonal antibody (HIF1α, Abcam, Cambridge, MA, clone H1alpha67, catalog # ab1) were applied for 1 hour at 35°C (HIF2α, 1:500 dilution) or for 2 hours at 37°C (HIF1α, 1:400 dilution) or 44 minutes at 37°C (SDHB, 1:600 dilution). An amplification kit (Ventana Medical Systems) was used to increase the signal of each antibody. The sections were detected using the UltraView DAB detection kit (Ventana Medical Systems), which is a HRP-Multimer system, utilizing DAB (3-3’ diaminobenzidine) as the chromogen. The sections were counterstained with hematoxylin (Ventana Medical Systems) for 12 minutes. A bluing solution was added for 20 minutes to adjust the counterstain color from a purple blue to a true blue. The sections were removed from the immunostainer and were gently washed in a dH2O/Dawn mixture (1 ml of Dawn/ 500 ml of dH2O) to remove any coverslip

oil applied by the automated instrument, followed by gentle rinsing in dH2O. The sections were placed in iodine for 30 seconds to remove any precipitates from fixation then dipped in sodium thiosulfate to clear the iodine. The sections were dehydrated in graded alcohols (70%, 95% x2 and 100% x2), cleared in xylene, and then cover slips were added. Controls were run with each antibody and stained appropriately. IHC data for SDHB are shown in Table 4.1

4.3.3 DNA and RNA extraction

Tumor samples were stored in liquid nitrogen. Each sample was examined histologically with haematoxylin-and-eosin-stained sections by a neuropathologist and representative sections were microdissected from the corresponding OCT embedded tumors (~10 mg). Genomic DNA was extracted with the Qiagen DNeasy Blood and Tissue Kit using the manufacturer's instructions; specifically, tumor tissue was lysed using the Qiagen Tissue Lyser with 5mm single dispenser beads at 25Hertz for 3 minutes, twice. Lysates for DNA were treated with 20 uL proteinase K (Qiagen DNeasy kit) at 56°C for 5 hours and with 4 uL of RNAseA (100mg/mL, Qiagen) at room temperature for 1 hour. Lysates for RNA were extracted with Qiagen RNeasy Kit using the manufacturer's instructions. Genomic DNA 500 ng of Genomic DNA was bisulfite-converted using the EZ DNA Methylation Kit (Zymo Research), whole-genome amplified, enzymatically fragmented and hybridized to the Illumina Infinium 450K Human Methylation arrays per the manufacturers instructions. Total RNA was subjected to DNase treatment using TURBO DNA-free Kit (Ambion) according to manufacturer's procedure, followed by RiboMinus treatment (Eukaryote Kit) according to the manufacturer's procedure.

4.3.4 DNA Methylation Data Comparison and Regional CpG Analysis

We used the DNA methylation data from Illumina infinium 450K Human methylation array from 22 tumor samples published in a previous study[22] and compared their data to DNA methylation data from our study. The genomic and clinical information associated with the tumor samples from Letouze *et al.* have been detailed in Table 4.1. Methylation data were extracted using GenomeStudio software (Illumina). Methylation values for each site are expressed as a β value, representing a continuous measurement from 0 (completely unmethylated) to 1 (completely methylated). This value is based on following calculation: β value = (signal intensity of methylation-detection probe) / (signal intensity of methylation-detection probe + signal intensity of nonmethylation detection probe). For methylation analysis, Illumina data were imported into Partek software. We performed a regional analysis where we first parsed our data into CpG Islands, CpG Shores, CpG Shelves using Partek. We also surveyed CpGs in Distant Regions (divided into 2kb regions). Regions with three probes or more were considered and average fraction methylation was calculated using the β value of the probes in that region and were further normalized by logit-transformation. Analysis of variance (ANOVA) with false discovery correction (FDR) was used to identify genes that were differentially methylated between the SDH Deficient PGLs and NCCs and SDH Present PGLs and NCCs. Significant changes were defined as genes having an FDR-corrected p-value < 0.05 as well having greater than 15% fraction DNA methylation. Figure 4.1 details the workflow for DNA methylation analysis followed by clustering and correlation to transcription data. Region information with associated annotations and probe numbers created for this study are available upon request.

4.3.5 High throughput RNA Sequencing and Analysis

A directional RNA library was prepared according to Illumina's protocol and sequenced on a 50bp single-end format on Illumina HiSeq 2000.

Alignments were generated from Illumina Fastq files to the hg19 genome with all known and theoretical splice junctions using Novocraft's novoalign aligner with the following parameters: -r None -t 90 –o SAM –m –l17

SamTranscriptomeParser application from the USeq package was used to parse SAM alignment files, that were aligned against chromosomes and extended splice junctions, and convert the coordinates to genomic space and sorts and save the alignments in BAM format. The maximum alignment score was set at 120, alignments were limited to unique matches. If the maximum number of locations threshold failed, one randomly picked repeat read per alignment was saved (-a 120 -n 1 -d).

DefineRegionDifferentialSeqs application was used to define differentially expressed genes between tumors and control.

To obtain read coverage tracks, BAM format files were generated from SAM alignments. Scaled read coverage tracks were generated for stranded data using the Sam2Useq application and were visualized on GBrowse.

4.3.6 Whole Exome Sequencing

Clonal segregation was noted in five SDH Present PGLs and one SDHD Deficient PGL from which DNA was extracted in the manner described in this section. Target enrichment for whole exome sequencing was performed with Agilent Technologies Inc. SureSelect Human All Exon V5+UTR kit and was analyzed on a HiSeq2000 sequencer (Illumina, Inc.) with two samples multiplexed per lane. Sequencing data were aligned with Novoalign (c) and filtered with standard GATK protocols through HCI's Tomato Pipeline. VarScan2 was utilized

for variant calling of paired tumor-normal samples, and variants were annotated using Annovar and in-house coding. This coding reflected the use of dozens of previous (non-PGL) cancer cases and healthy controls run through the identical pipeline and sequencers to filter out common variants and/or sequencing artifacts and misalignments. In addition, variants found to be present at >0.2% prevalence in any of 1000 genomes, ESP6500, or UCSC database when at least 400 alleles have been reported, were eliminated as being unlikely somatic mutation candidates. Further, only nonsynonmous mutations with greater than 25% variance frequency were included in consequent analysis.

4.4 Results

4.4.1 Definition of Stringent Regional CpG Analysis for Determining DMRs using the Illumina 450K Methylation Array

Accumulation of succinate in the cell can inhibit many αKG-dependent enzymes in the cell, including TET enzymes, involved in active DNA demethylation[9]. We hypothesized that SDH Deficient PGLs would gain DNA methylation at focal regions. To test this, we profiled genome wide changes in DNA methylation in SDH Deficient PGLs compared to NCCs. Additionally, we also profiled changes in DNA methylation in SDH Present PGLs compared to NCCs. Since the majority of the SDH Present PGLs do not accumulate succinate, we would not anticipate similar regions gaining DNA methylation. To test this, we used Illumina's 450K DNA methylation array, which provides single nucleotide resolution over ~485,000 CpGs in the human genome and covers >90% of genes. However, unlike recently published work[22,30], instead of performing a single CpG analysis to determine variation, we used a regional analysis involving the status of multiple CpGs in a region. In this new method, we used annotations defined from the array to first parse all regions into four groups: CpG Islands, Shores (regions that are 2 kb

upstream and downstream of CpG Islands), Shelves (2 kb regions that are either 4 kb upstream or downstream of CpG Islands) and Distant Regions (2 kb regions that are not associated with a CpG Island) where each region contained at least 3 CpGs or more (Figure 4.2). Analyzing data from all four regions is another novel aspect of our work, where unlike recently published work on SDH Deficient paragangliomas, we did not limit ourselves to looking at differentially methylated regions in CpG Islands. Regions in each CpG related category represented all regulatory regions in the genome (±200 bp from Transcription Start Site-TSS200, ±1500 bp from Transcription Start Site-TSS1500, 5'UTR and 1st Exon), Body, 3'UTR and intergenic (Supplementary Figure 4.1). We pursued a regional analysis for the following reasons. Globally, we observed that DNA methylation levels did not significantly change and were consistent between the tumors and controls (Figure 4.4). Hence, when the methylation level at one CpG changed among several that remained unchanged, while this change might be statistically significant, there is not much precedance of a single CpG being targetted by either DNA methyltransferases or DNA demethylases for misregulation. These enzymes target multiple CpGs in a region. As a result, any future analysis for DNA methylation change included at least three CpGs which also served as a stringent threshold allowing us to consider data from the majority of the CpGs tiled on the array (57%, Figure 4.3). Also, we calculated mean DNA methylation of all CpGs in a 2 kb region that inturn allowed us to call differentially methylated regions (DMRs) with high confidence (FDR-corrected p-value < 0.05). Consequently, to identify potential regulatory CpG sites that are biologically meaningful we expect more than one CpG in a defined region to be affected. Finally, DNA methylation analysis using high throughput whole genome bisulfite sequencing techniques, adopt a minimum of five CpGs to define a DMR[28,29]. Keeping all these factors in mind, we decided to take a regional approach as this allowed us to survey the majority of genomic regions of interest

and defined DMRs as regions that had 15% or more change in fraction methylation with a p-value ≤ 0.05.

4.4.2 Global DNA Hypomethylation for SDH Deficient and SDH Present Paragangliomas Compared to Progenitor Cells

Globally, the DNA methylation patterns of the SDH Deficient and SDH Present paragangliomas in our study were very similar to each other as well as to the progenitor cell controls, neural crest cells (NCCs) and ES cells. We plotted Pearson coefficients in a hierarchical clustering heatmap/dendrogram between the SDH Deficient and Present tumors from our study (black text), NCCs and ES cells (grey text) and SDH Deficient, WT tumors and RET, NF1, and MAX PGLs (blue text) from the previously published study. All tumors were highly correlated to each other with a Pearson coefficient ranging from 0.94 to 0.99 (Figure 4.4). Surprisingly, the malignant (metastatic or recurrent) PGLs did not particularly cluster together; instead they correlated well and clustered together with benign PGLs. Also of note, the SDH Deficient tumors (both malignant and benign) from our study were less anticorrelated with the RET, NF1 (somatic, benign) and WT benign PGLs from the Letouze *et al*. study. These observations represent either technical variability or tumor variability between the SDH Deficient tumors used in both studies. This variability can also be observed in the principle component analysis (Figure 4.5), since the SDH Deficient tumors from our study and the previously published study did cluster together, but not as tightly. Global median DNA methylation (Figure 4.5) revealed that ES cells and neural crests fall in the expected range and are 73% and 75% methylated, respectively. SDH Deficient and Present tumors, however, ranged from 55% to 71% methylation and 47% to 55% methylation, respectively. These findings detail a global reduction of DNA methylation in SDH Deficient and Present tumors compared to progenitor cells. Remarkably, SDH Deficient tumors

do not lose as much DNA methylation as SDH Present tumors compared to progenitor cells. These trends were also observed in SDH Deficient and Present tumors from Letouze *et al.* where their SDH Deficient tumors had 48% to 54% median methylation and the SDH Present tumors (VHL, NF1, RET, MAX mutant tumors and wt tumors) had 30% to 54% median methylation. Overall, there is very little global DNA methylation variability in both SDH Deficient and SDH Present, nor between our PGL data and the PGL data from the previously published study[22]. This makes a strong case for regional methylation changes that potentially drive PGL tumorigenesis.

4.4.3 DNA Methylation Dynamics at CpG Islands, CpG Shores and Distant Regions Reveal Focal Regions of both Hypomethylation and Hypermethylation

To determine focal changes in differentially methylated regions, we compared methylation data from SDH Deficient and Present tumors, from our study and the previously published study[22], to NCCs. We identified 2747, 1884 and 841 DMRs from CpG Islands, CpG Shores and Distant Regions (containing three or more CpGs in a region), respectively, based on a p-value ≤ 0.05 and performed k-means clustering on these DMRs (Figure 4.7, 4.8 and 4.9). Data for CpG Shelves is not shown, since none of these regions passed the stringent thresholds for differentially methylated regions. CpG Islands, CpG Shores and Distant Regions all partitioned in five distinct clusters where a striking the majority of the regions lost DNA methylation in the tumors compared to control (Figure 4.7: clusters 1,2,4 and 5; Figure 4.8: clusters 1,3 and 5; Figure 4.9: clusters 1,3 and 4). Another interesting observation from the heatmaps was that the DMRs identified from both subclasses of tumors, when compared to the NCCs, had an overall similar pattern of DNA methylation change.

While the majority of the changes were hypomethylated in the tumors compared to NCCs, we also identified clusters where regions gained DNA methylation in the tumors compared to the control. These hypermethylated regions were identified in all categories including CpG Islands, Shores and Distant Regions (Figure 4.7: cluster 3; Figure 4.8: clusters 2 and 4; Figure 4.9: clusters 2 and 5). In CpG Islands, cluster 3 consisted of 450 DMRs, which were further partitioned into two clusters, namely cluster 3a and 3b. Regions in cluster 3a appeared to gain DNA methylation exclusively in the SDH Deficient tumors, whereas regions in cluster 3b gained DNA methylation in both SDH Deficient and Present tumors; albeit to a comparatively a lesser extent in the latter. CpG Island DMRs were partitioned into functional genomic locations where a total number of 450 DMRs became hypermethylated and a total number of 2427 DMRs lost DNA methylation (Figure 4.10). Notably only 30% of the regions that gained DNA methylation and 23% of regions that lose DNA methylation fell into promoter regions. Interestingly the majority of the regions that gained (66%) or lost DNA methylation (75%) in the tumors compared to the control were enriched in either the body of genes or intergenic regions. These regions did not overlap significantly with either repeat regions or enhancers. In CpG Shores, clusters 2 and 4, consisted of 575 hypermethylated DMRs and 1612 hypomethylated DMRs. These DMRs were partitioned into functional groups (Figure 4.11) where 40% of hypermethylated regions enriched in the promoters of genes while 60% enriched in body, 3'UTR and intergenic regions. Conversely to the aforementioned CpG Islands, 60% of hypomethylated regions at CpG Shores enriched in promoters of genes whilst the remaining 40% enriched in nonpromoter regions. In Distant Regions, (not associated with CpG Islands) 31% of hypermethylated regions and 42% of hypomethylated regions are associated with promoters while the remaining regions (69% and 58%) fell into nonpromoter regions. This is in contrast to hypomethylated CpG Islands that are mostly affiliated

with nonpromoter regions (Figure 4.10).

4.4.4 Focal Hypermethylated Changes Reveal Novel Gene Candidates that may Contribute to PGL Oncogenesis

To gain further insight into the genes that may be affected by DNA methylation changes we sought to parse hypermethylated and hypomethylated promoters and correlate these changes to transcriptional changes at CpG Islands, Shores and Distant Regions (Figure 4.13, 4.14, 4.15 and 4.16). CpG Islands that were hypermethylated, were partitioned into two clusters, 3a and 3b (Figure 4.7), to parse differences between SDH Deficient and SDH Present tumors. Genes that were hypermethylated only in SDH Deficient tumors (cluster 3a) included potential tumor suppressors genes previously reported in a vast array of other cancers. These epigenetically silenced genes include DRD4, KRT19, and FRZB (Figure 4.13 and Table 4.2; genome snapshots Figures 4.20 and 4.21). Interestingly, while SDH Present tumors did not gain DNA methylation in the CpG Island promoters of these genes, the impact on transcriptional repression was similar to that observed in SDH Deficient tumors. Strikingly, genes that were hypermethylated in both tumor subclasses (cluster 3b) also included genes that are either transcriptionally silenced or mutated in other cancers such as DNMT3A, GRHL2, KAZALD1, NSD1, ATP5G2, and TOX3 (Figure 4.14 and Table 4.3; genome snapshots Figures 4.22 and 4.23). A small subset of genes showed gain in gene expression regardless of the gain in DNA methylation in CpG Island promoters in both tumor subclasses; this could potentially be explained by the presence of 5hmC (Supplementary Figure 4.2 and 4.3). We also correlated genes from clusters 3a and 3b that were hypermethylated in their nonpromoter CpG Islands, to gene expression (Supplementary Figure 4.4 and 4.5) but did not find any striking pattern.

Figure 4.15 highlights genes that were hypermethylated in their CpG

Shore promoters (cluster 2 and 4, Figure 4.8) and correlated with transcriptional downregulation. Comparatively, 52 genes were hypermethylated in their CpG Shore promoters in comparison to the 22 genes that were hypermethylated in their CpG Island promoters (Figure 4.15 compared to Figures 4.13, 4.14). Only five genes gained DNA methylation exclusively in SDH Deficient tumors in their promoter CpG Shores (Supplementary Figure 4.6). The majority of remaining genes were affected in both PGL subclasses (Figure 4.15). These included genes involved in cell morphology changes, cell migration (FAM60A, KRT8, TRIP6), tumor suppressors (FGFR2, PAX6) and genes that are epigenetically silenced in other tumors such as SFRP2 and SOX9 were also found affected in PGLs (Table 4.4). Intriguingly, JARID2, TET1, SALL1 and SALL4 are genes known to be involved in epigenetic regulation and were also found to be hypermethylated and transcriptionally downregulated in both SDH Deficient and Present PGLs. (genome snapshots Figures 4.24 and 4.25). We observed a small group of genes that gained DNA methylation in their CpG Shores and were transcriptionally upregulated (Supplementary Figure 4.7) as well as genes from clusters 2 and 4 that were hypermethylated in their nonpromoter CpG Shores, to gene expression (Supplementary Figure 4.8 and 4.9), again, but did not find any striking patterns.

Only 15 genes that were hypermethylated in promoters, that are distant from CpG Islands, (clusters 2 and 5, Figure 4.9) correlated with loss in gene expression (Figure 4.16). Only three of these genes gained DNA methylation exclusively in SDH Deficient tumors (Supplementary Figure 4.10). Interestingly, FABP3 was also seen downregulated in SDH Present tumors although it did not pass our statistical thresholds (genome snapshot Figure 4.26). The remaining 12 genes were affected in both PGL subclasses which included two previously reported potential tumor suppressors SOX5 and SOX6 (Table 4.5). These genes are hypermethylated in SDH Deficient and Present tumors compared to NCCs

and are also transcriptionally repressed. Genes that were upregulated regardless of hypermethylated promoters in Distant Regions are shown in Supplementary Figure 4.11. Genes that gained methylation in the nonpromoter Distant Regions were correlated to gene expression and are plotted in Supplementary Figures 4.12 and 4.13; however, no striking patterns were observed.

4.5 Few Hypermethylated and Downregulated Genes in our SDH Deficient and Present PGLs Overlap with Those Previously Published Datasets of Cells or Tissues Lacking SDH or IDH

We were interested in determining if any of the gene candidates that gain focal hypermethylation at their promoter regions and lose gene expression in paragangliomas are also observed in previously published datasets of non-PGL tumors having mutations in either SDH or another TCA cycle gene, IDH. We did this to identify common genes that were impacted in tumors with mutations in TCA cycle genes. We compared genes to a list of genes from Killian *et al*., that gained 0.15 or more fraction DNA methylation in IDH1 gliomas, SDH Deficient GISTs or SDH Deficient PGLs compared to their respective controls (transcription data were not provided). We also compared our genes to two list of genes from Letouze *et al.* featuring hypermethylated regions in SDHB knockout mouse chromaffin cells compared to wildtype chromaffin cells and in SDH Deficient PGLs compared to SDH Present PGLs. The third list we compared our genes to were IDH1 gliomas that contained hypermethylated genes that were repressed[30]. The first list consisted of genes that that gained 0.15 or more fraction DNA methylation and were downregulated in SDH Deficient PGLs compared to SDH Present. The second list consisted of genes that were hypermethylated (0.15 or more gain in fraction DNA methylation) in SDHB knock out mouse chromaffin cells compared to wildtype (transcription data not provided). We observed four genes, BMP4,

FABP3, FRZB and TRIP6 that had statistically significant overlap ($p < 0.001$), with hypermethylated genes in IDH1 gliomas and SDH Deficient GISTs and PGLs (Figure 4.27) from Killian *et al*. We observed two other genes, KRT19 and RPP25, that overlapped with less statistical significance ($p < 0.0127$) with hypermethylated genes in SDH Deficient PGLs and SDHB knock out mouse chromaffin cells from Letouze *et al.* (Figure 4.28). Remarkably, however, we observed 17 genes in total that overlapped exclusively between the SDHB knockout mouse chromaffin cells and genes that were hypermethylated and downregulated in PGLs from our study but not with their PGLs ($p < 0.001$). These genes included tumor suppressors FABP3 and FRZB, which were also observed to be hypermethylated in IDH1 gliomas, SDH Deficient GISTs, and PGLs as mentioned above. Finally, we found six genes in common between our PGLs and IDH1 gliomas from Turcan *et al.* (Figure 4.29, $p< 0.0085$), which included TGIF1, RPP25, TMEM159, DNMT3A, TRIP6, and GAP43. By comparing genes misregulated in tumors with mutations in Krebs-cycle enzymes, we may have identified a few of the interesting gene candidates involved in the oncogenesis of these tumors most susceptible to changes in metabolism. Interestingly, our SDH Present PGLs accrue similar misregulation of tumor suppressor genes and epigenetic modifier genes despite the lack of mutations in Krebs-cycle enzymes. This may suggest an alternate route of misregulation.

4.5.1 Regions that Lose DNA Methylation in their CpG Island Promoters may be Associated in Genes that also Contribute to PGL Oncogenesis

The majority of our DMRs at CpG Islands, Shores and Distant regions were hypomethylated in SDH Deficient and Present tumors compared to NCCs (Figure 4.7, 4.8 and 4.9). We found 73 CpG Island promoters, 143 CpG Shore promoters

and 110 promoters distant from CpG Islands that were hypomethylated and correlated with gain in transcription (selected genes graphed in Figure 4.17, 4.18 and 4.19). Among these genes we found several interesting candidates involved in promoting invasion and metastasis (ACP5, CHL1, CPEB4, DOCK2, LY6K), cell proliferation and tumor growth (GNA14, PMEPA1, TACSTD2), inhibition of cell death (BCL2L1, CFLAR, DGKA, SRPK3), promoting angiogenesis (ALK1 and FSTL3), chronic inflammation (CD14) and genomic instability (REC8) (Table 6). We also observed long non-coding RNAs such as ncRNA00182 and psiTPTE22 affected in this category.

4.5.2 Hypomethylated CpG Island and CpG Shores Promoters Intersect with Small DMRs from Colon Cancers

Five hundred and fifty eight CpG Island promoters and 968 CpG Shore promoter regions were hypomethylated in both SDH Deficient and Present PGLs compared to neural crest cells. Remarkably, we found 22 CpG Island hypomethylated promoters and 61 CpG Shore hypomethylated promoters in PGLs overlap with statistical significance ($p<0.009$ and $p < 0.001$, respectively) with small DMR regions, in CpG Shores, from colon cancer (Supplementary Figure 4.14). Presumably this small overlap is due to differences in tumor type, and effects that are secondary to carcinogenesis. Notably, the few regions that do overlap might represent gene regions important for tumorigenesis in a variety of tissue type. We did not see any statistically significant overlap with CpG Island and CpG Shore nonpromoter (Body, 3'UTR and intergenic) regions.

4.5.3 Hypomethylated CpG Shore and Distant Promoters Intersect with Repeat Elements

We also found 45% of CpG Shore promoters overlap with LINE elements, 71% overlap with SINE elements, 17% and 20% overlap with LTRs and DNA repeats respectively (p-value for all comparisons <0.001, Supplementary figure 4.15). We found similar overlap of CpG Shore body and 3'UTRs with SINEs (55%) and DNA repeats (16%) and CpG Shore intergenic regions with LINEs (40%) and SINEs (61%) (p-value for all comparisons < 0.001, Supplementary Figures 4.16 and 4.17). Similarly, we found 34% of Distant Region promoters overlap with LINE elements, 55% overlap with SINE elements, 14% and 14% overlap with LTRs and DNA repeats respectively (p<0.001, Supplementary Figure 4.18). Distant region body and 3'UTRs overlapped significantly (p<0.001) with LINEs (32%), SINEs (57%) and DNA repeats (15%) (Supplementary figure 4.19). Finally, intergenic regions that associated with Distant Regions also overlapped with significantly (p<0.001) with LINEs (40%), SINEs (68%), LTRs (30%) and DNA repeats (21%) (Supplementary Figure 4.20). These findings suggest a bias for DNA hypomethylation at repeat elements.

4.5.4 SDH Deficient and Present PGLs have Very Similar Transcriptome Profiles Compared to Neural Crest Cells

We identified genes that were differentially expressed in the two PGL subclasses (SDH Deficient and SDH Present) compared to NCCs (using p-value < 0.001 and a fold change of 3 or more). We compared gene lists that were upregulated or downregulated in both SDH Deficient PGLs and SDH Present PGLs over NCCs and performed GSEA (gene set enrichment analysis). GSEA revealed an NES score of 4.63 with a p-value < 0.001 for upregulated genes in

both data sets (top panel, Figure 4.30) and a normalized enrichment score (NES) of -5.08 with a p-value < 0.001 for downregulated genes in both data sets (bottom panel, Figure 4.30). We further performed k-means clustering of log2 transformed RPKM values of 4549 genes that were differentially expressed (p-value < 0.001 and fold change ≥ 3, Figure 4.31). Six distinct clusters were observed from the clustering analysis where cluster 1, 3 and 6 represented genes that were upregulated in tumors compared to control and clusters 2, 4 and 5 represented genes that were downregulated in tumors compared to control. Strikingly, GO-term analysis of upregulated genes in clusters 1, 3 and 6 enriched for inflammatory response, defense response, angiogenesis, blood vessel development, hypoxia, myc transcription, negative regulation of apoptosis and cap-dependent translation ($p < 0.05$). In contrast, downregulated genes in clusters 2, 4 and 5 enriched for GO-term categories such as DNA damage response, homologous recombination, base excision repair, nucleotide excision repair, neural crest differentiation, histone modification ($p < 0.05$). Supplementary Figures 4.21 and 4.22 represent the genes in each enriched GO-term category and their fold change in SDH Deficient and SDH Present PGLs compared to NCCs. No significant GO-terms were enriched in the few genes that were differentially expressed between the two tumor groups (data not shown), which further helped confirm that both classes of PGLs used in this study had very similar transcription profiles.

4.5.5 Whole exome sequencing of SDH Present PGLs and a Subset of SDH Deficient PGLs Reveals Mutations in Many Genes Including Key Epigenetic Enzymes

From our methylation and transcriptome data we observed a striking resemblance in affected regions in SDH Deficient and SDH Present PGLs. We

speculated that our two subclasses of PGLs were epigenetically similar to each other because the SDH Present PGLs may have mutations in either the SDH complex or genes that were affected in the same pathway as those that accumulated succinate. In order to understand this possible link, we performed whole exome sequencing on five SDH Present tumors (with three matching germline DNA) and an SDHD Deficient PGL. Sequencing results confirmed that the SDH Present PGLs did not harbor mutations in any of the SDH complex genes or other previously reported susceptibility genes (RET, NF1, TMEM127, MAX, VHL, EPAS1). Strikingly, we found mutations in a Jumonji histone demethylase, KDM6B, in five of our six surveyed PGLs, including the SDH Deficient PGL. The mutations reported in this gene were not previously found in population databases. Also, this gene was enriched with statistical significance ($p<0.005$) in pairwise comparisons of the PGLs. Given this finding; we focused on confirming novel mutations found in PGLs in several genes involved in epigenetic regulation. Genes such as NSD1, KDM6B, JARID2, JMJD4, PRDM2, MBD5, KDM5C, DNMT1, YY1, YY1AP1, SETD1A, SETD1B were a few of the epigenetic candidates that were mutated in at least one of the sequenced tumors (Table 4.7). While this list and analysis is not an exhaustive representation of all possible genes that are affected in SDH Deficient and SDH Present tumors, these genes may provide a possible explanation for the similar epigenetic and transcriptional profiles observed in SDH Deficient and Present PGLs.

4.6 Discussion

In our study, we have demonstrated that SDH Deficient and SDH Present PGLs share very similar DNAme changes compared to a progenitor cell type, neural crest cells (NCCs). The two PGL subclasses also share very similar transcriptome profiles compared to NCCs. This is in stark contrast to previously

published reports that have focused on changes between tumor subtypes – those having mutations in SDH and those having mutations in other susceptibility genes such as RET, NF1, TMEM127 and MAX. Although our DNAme data from SDH Deficient tumors closely resembled data generated by others, we also observed some differences[22]. We presume these differences come from a combination of technical and biological attributes, because PGLs are highly vascular and heterogeneous tumors chief cells must be stained and enriched prior to isolation of DNA and RNA. To minimize contamination from nontumor tissue we implemented a very stringent PGL tissue recovery pipeline, where each sample was examined histologically with haematoxylin-and-eosin-stained sections by a neuropathologist and representative sections were microdissected from the corresponding OCT embedded tumors (~10 mg). This rigorous procedure enables us to have high confidence in attributing our findings of changes in DNAme and RNA levels to PGLs. Furthermore, our analysis identified hypermethylated candidate genes that had not previously been reported in PGLs. Presumably this is due to several factors: 1) rather than comparing our tumor subclasses to one another, we compared them to a progenitor cell type. Notably, because all of our tumors lacked mutations in the non-SDH susceptibility genes and previously published data did not, we may be studying different classes of PGLs. 2) For determining differentially methylated regions we included a minimum of 15% change in fraction DNAme and a cut off of p-value ≤ 0.05 but we also adopted a regional analysis instead of a single probe wise analysis. Although there is an inherent bias toward promoters in the Illumina 450K array, which focuses on CpG Islands, we also included CpG Shores and Distant Regions in our analysis. These criteria were not used by other studies, thus allowing us to be more comprehensive in assessing regions that were either gaining or losing DNAme and to correlate transcriptional status.

Interestingly, our analysis revealed hypermethylation at CpG Island

promoters and transcriptional repression of genes such as DRD4, KRT19, and FRZB exclusively in SDH Deficient paragangliomas[31-38]. All other gene candidates were equally affected in both tumor subclasses. These genes have been reported to be epigenetically silenced in a range of different cancers and may serve as potential tumor suppressors. DNMT3A, GRHL2, KAZALD1, NSD1, ATP5G2 and TOX3 were hypermethylated in their CpG Islands promoters in both tumor subclasses and were either transcriptionally repressed or mutated in other cancers[39-45]. In promoters located in CpG Shores, genes involved in cell morphology, cell migration (FAM60A, KRT8, TRIP6), tumor suppressors (FGFR2, PAX6) and genes that are epigenetically silenced in other tumors such as SFRP2 and SOX9 were also found hypermethylated and downregulated in PGLs[46-55]. Intriguingly, JARID2, TET1, SALL1 and SALL4 are all genes involved in epigenetic regulation and were also found to be hypermethylated at their CpG Shores and transcriptionally downregulated in both PGL subclasses. At promoters which fell into regions distant from CpG Islands three potential tumor suppressors are also downregulated and DNA hypermethylated, FABP3, SOX5 and SOX6, which were also reported previously to be affected in a variety of other cancers[56-59].

As a confirmation of our method, we did see a small yet statistically significant overlap of affected genes when we compared our data to other datasets from tumors with mutations in TCA cycle genes[21,22,30]. We observed four genes, BMP4, FABP3, FRZB and TRIP6 that overlapped with high statistical significance ($p < 0.001$) with hypermethylated genes in IDH1 gliomas and SDH Deficient GISTs and PGLs[21]. We observed two other genes, KRT19 and RPP25, that overlapped with less statistical significance ($p < 0.0127$) with hypermethylated genes in SDH Deficient PGLs and SDHB knock out mouse chromaffin cells[22]. Finally, we observed six genes, TGIF1, RPP25, TMEM159, DNMT3A, TRIP6, GAP43, in common with IDH1 gliomas ($p < 0.0085$)[30]. Notably, these genes were found hypermethylated (in most cases) and

repressed in both SDH Deficient and Present PGLs used in our study. Hence, while mutations in Krebs-cycle genes may be strongly associated with misregulation of these genes, we speculate that both tumors may adopt different mechanisms to achieve misregulation of similar gene targets. While in SDH Deficient tumors misregulation is driven by the accumulation of a co-factor required for the activity of epigenetic enzymes, SDH Present tumors may harbor mutations in epigenetic enzymes that could potentially phenocopy the misregulation observed in SDH Deficient tumors.

After utilizing a stringent analysis pipeline, DNAme changes only correlated with transcriptional level to a minimal degree (13.4% at CpG Island promoters, 16.2% at CpG Shore promoters, and 32% at Distant Region promoters). Other factors, including chromatin modifications, histone variants, and transcription factors, are likely to account for the majority of transcriptional changes. We also observed that several nonpromoter regions gained DNAme in tumors. These regions did not correlate with enhancers or repeats, and a large number of them correlated with increased expression of associated genes. Because transcriptionally active gene bodies are often hypermethylated, this result is not terribly surprising[60,61].

Surprisingly, another interesting observation from our data is that the majority of the DMRs actually lose DNAme. We anticipated that succinate accumulation would inhibit TET enzymes from hydroxylating 5mC to 5hmC and hence we expected a gain in DNAme[9]. When we analyzed our data and other's data, on a global level we observed hypomethylation of tumors, and the number of DMRs that are hypomethylated are almost 3 to 5 fold greater than those that are hypermethylated. Hypomethylation has been observed in many cancers and is usually associated with repetitive regions that are normally DNA methylated. Repetitive region hypomethylation is correlated with decreased genome stability, which is mediated by recombination between non-allelic repeats causing an

increase in chromosome rearrangements or translocations. Hypomethylation of repeats may also cause an increase in retrotransposon activity, which can lead to wide spread gene disruption[62,63]. While this level of hypomethylation in SDH disrupted tumors would not be predicted from our working model, it highlights the fact that epigenetic misregulation in PGLs is a dynamic process involving a combination of gain and loss of DNAme acting in concert with mislocalization of histone modifications to potentially promote tumorigenesis. It is important to note, however, that PGLs Deficient in SDH globally lose less DNAme than SDH Present PGLs when compared to progenitor cells, arguing strongly for SDH's role in inhibiting DNA demethylation.

Among genes that were hypomethylated and gained gene expression, we found several interesting candidates involved in promoting invasion and metastasis[64-66] (ACP5, CHL1, CPEB4, DOCK2, LY6K), cell proliferation and tumor growth[67-69] (GNA14, PMEPA1,TACSTD2), inhibition of cell death[70-75] (BCL2L1, CFLAR, DGKA, SRPK3), promoting angiogenesis[76,77] (ALK1 and FSTL3), chronic inflammation[78] (CD14) and genomic instability (REC8)[79]. None of these genes have been previously reported to be affected in PGLs, presumably do to the fact that previously published studies focused mainly on regions that gained DNAme, and failed to adequately consider loss of DNAme. These data demonstrate that it is important to look at all changes in the genome regardless of their location (CpG Island, Shore or Distant Regions) and or expected methylation change, in order to get a complete picture of all the aberrant changes that may be occurring and hence potentially contributing to tumorigenesis.

Finally, we looked at the transcriptional changes in SDH Deficient and SDH Present PGLs and compared them to NCCs. Our data agreed with previously published reports[26,80,81] that showed upregulation of genes involved in hypoxia, angiogenesis, blood vessel development, inflammatory response,

defense response, myc transcription, negative regulation of apoptosis and cap-dependent translation. Furthermore, our data also agrees with downregulation of genes involved in DNA damage response, homologous recombination, base excision repair, nucleotide excision repair, neural crest differentiation and histone modifications. Strikingly, both SDH Deficient and SDH Present PGLs had very similar transcriptome profiles compared to NCCs. This is in agreement with previous reports that have demonstrated that wildtype PGLs can be transcriptionally similar to either SDH Deficient and VHL mutated tumors or to RET, NF1 and TMEM127 tumors[26].

To identify a potential link between SDH Deficient and Present PGLs and to understand why their epigenetic profiles and transcriptomes look so similar, we performed whole exome sequencing. Preliminary data points towards mutations in epigenetic genes in SDH Present tumors. These genes may be part of the same epigenetic axis affected in SDH Deficient, succinate-accumulating PGLs. For example, KDM6B is an H3K27me2 and me3 demethylase and is found in a complex with MLL4, an H3K4 methyltransferase. MLL4 and KDM6B are perhaps recruited to promoters of genes by transcription factors to methylate H3K4 and demethylate H3K27 and hence allow for transcriptional activation of the gene[82]. Mutations in MLL4 and KDM6B were not found in their catalytic domains, however, mutations may effect partner interactions or target recognition and hence have downstream effects on gene transcription. This would cause the affected gene to remain aberrantly silenced. Mutations in other H3K4 methyltransferases, SETD1A and SETD1B have also been observed that might deter them from being recruited to their appropriate target regions leading to inactivation of the gene. JARID2, a member of the Jumonji family of proteins lacking demethylase activity is known to bind GC-rich DNA and recruit the Ezh2/PRC2 complex to its target sites where Ezh2 methylates H3K27 and transcriptionally silences the genes[83].

We hypothesize that mutations in JARID2 disrupt its ability to interact or recruit PRC2 to its target sites, leading to aberrant gene activation. This is the first report to implicate a mutation in MBD5 that is associated with cancer. MBD5, which contains a methyl-binding domain and does not bind methylated DNA, is known to associate with heterochromatin and hence may contribute to its formation[84]. It is linked to developmental disorders as knockout mice show growth retardation and preweaning lethality[85]. While its exact role is not fully understood, mutations in this gene might affect its interaction with heterochromatin. The knock-out mouse for KDM5C, which encodes an H3K4me2 and me3 demethylase, have neurulation and cardiac looping defects[86]. Mutations in KDM5C are associated with mental retardation, austism and renal carcinoma[87]. Taken together, while this does not represent an exhaustive list of genes that are either mutated or transcriptionally misregulated in SDH Deficient and SDH Present tumors, these candidate genes may provide a novel and fascinating explaination for the strikingly similar epigenetic and transcriptional profiles observed in our SDH Deficient and SDH Present PGLs.

Although mutations in different susceptibility genes (SDHx, VHL, EPAS1, RET, NF1, TMEM127, MAX) have been reported in paragangliomas, misregulation in each tumor subclass can mechanistically converge on the shared pathways towards proliferation. For example, gain of function mutations in RET, and loss of function mutations in NF1 and TMEM127 can activate the PI3K pathway which will consequently activate mTOR. Myc, no longer bound by MAX (due to loss of function mutations), cooperates with mTOR and thus activates it. mTOR activation regulates cell growth through increased synthesis of nucleic acids, lipids, fatty acid, proteins and most importantly can activate HIF[4]. Accordingly, mutations in the SDH complex cause accumulation of succinate, which competitively inhibits PHDs leading to stabilization of HIF. Activated HIF regulates transcription of its downstream targets involved in increased glucose uptake, glycolysis, angiogenesis and metastasis.

Succinate accumulation can also inhibit the activity of α-KG dependent epigenetic enzymes, which in turn can lead to epigenetic misregulation of downstream targets that may contribute further to PGL oncogenesis as previously discussed. Our work shows that SDH Deficient and Present PGLs are phenotypically very similar. Similar genes are upregulated and downregulated in both tumor subclasses. Furthermore, they share the majority of regions that gain and lose DNA methylation when compared to neural crest cells. Strikingly, more regions are hypomethylated in PGLs than hypermethylated. Whole exome sequencing of both PGL subclasses shows mutations in many epigenetic modifier genes and hence we speculate that in PGLs lacking SDH mutations, epigenetic enzymes may harbor mutations that could phenocopy the misregulation in SDH Deficient tumors. To test this hypothesis, further research using cell culture systems and animal models will need to be performed. In addition, these systems can be used to further probe the changes in the chromatin landscape with respect to identifying regions of histone modification accumulation and mislocalization in PGLs and elucidating its impact on neural-crest derived tumors. Furthermore, identifying downstream targets of key transcription factors involved in PGL oncogenesis such as HIF and Myc could help us gain further insight into understanding PGL's selective growth advantage leading to progression and metastasis of the cancer.

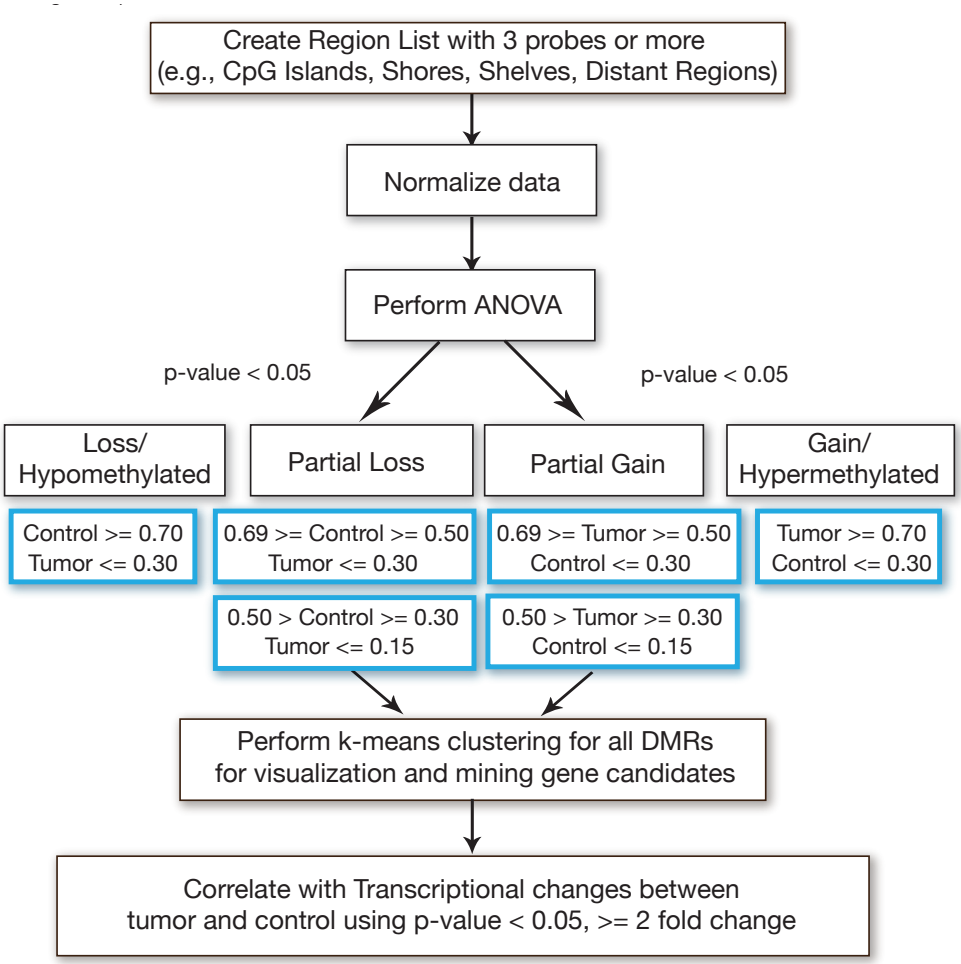


Figure 4.1: Overall workflow for analysis of genome wide DNA methylation data from 450K array and correlating it to transcriptional changes. Workflow describes p-value cut offs as well as absolute DNA methylation value cut-offs used to identify differentially methylated regions in tumors vs neural crest cells.

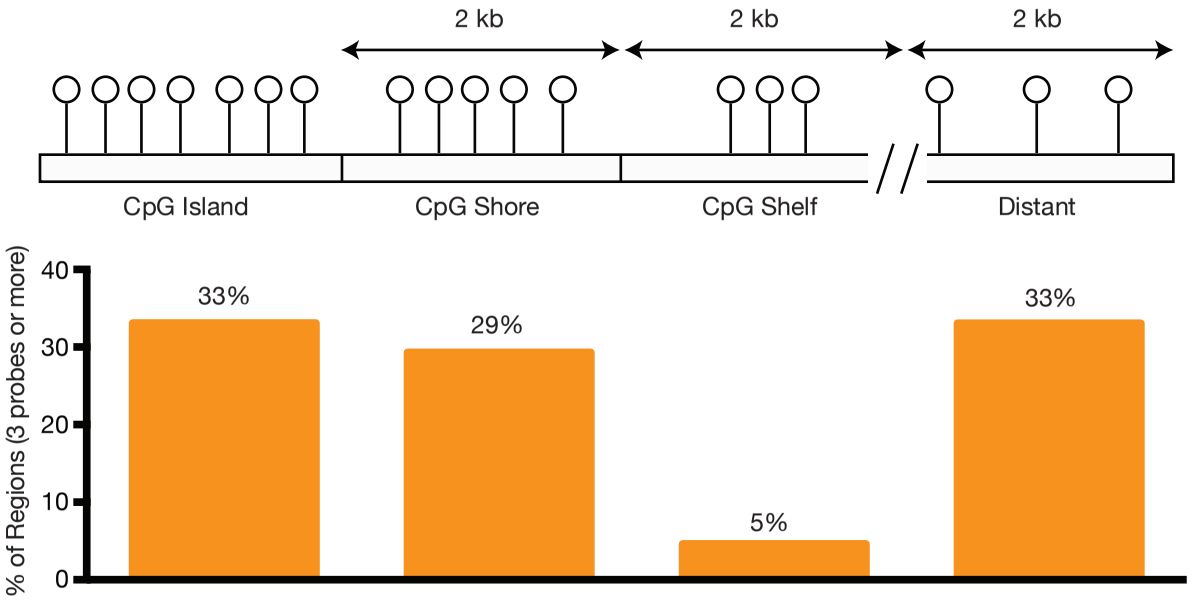


Figure 4.2) Distribution of number of regions (three probes or more) covered on the 450K array in relation to CpG Islands.

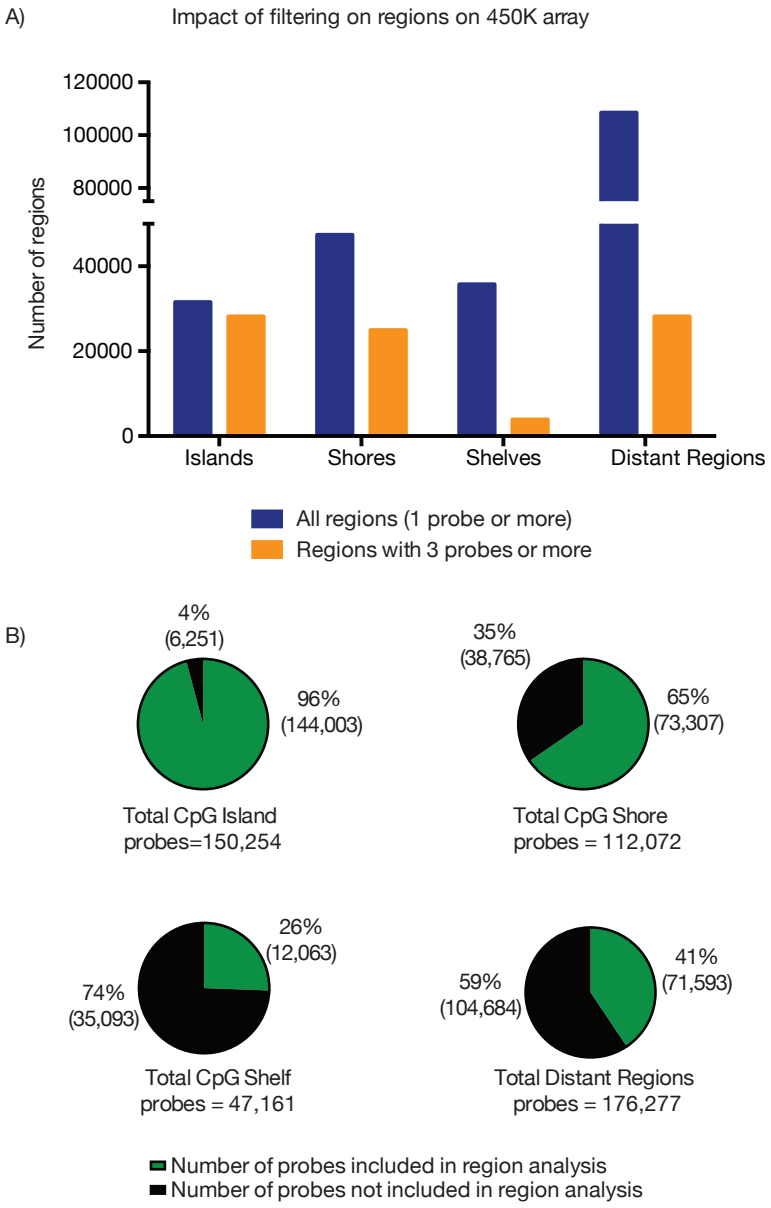


Figure 4.3) Impact of filtering for three probes or more in all regions of the 450K Illumina Methylation array. A) Histogram showing impact of filtering for regions containing at least three probes or more. B) shows total number of probes being considered for analysis in CpG Islands, Shores, Shelves and Distant Regions once they are parsed into regions.

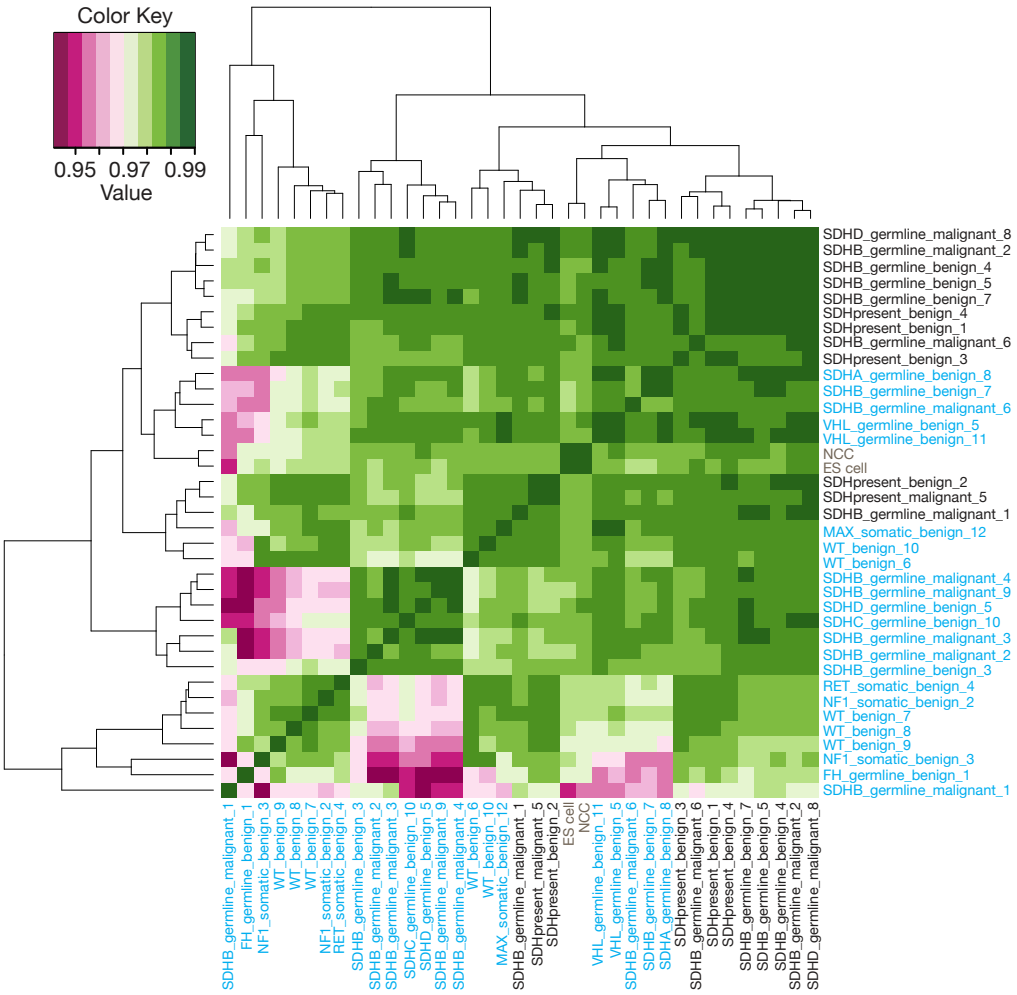


Figure 4.4) Hierarchical clustering dendrogram showing Pearson correlation of DNA methylation between SDH Deficient, SDH Present tumors from our study (in black) and Letouze *et al.* study (in blue). Controls are shown in grey. NCC = Neural Crest Cells, ES Cell = Embryonic Stem Cells. Numbers associated with each sample correspond to the numbered samples in Table 4.1.

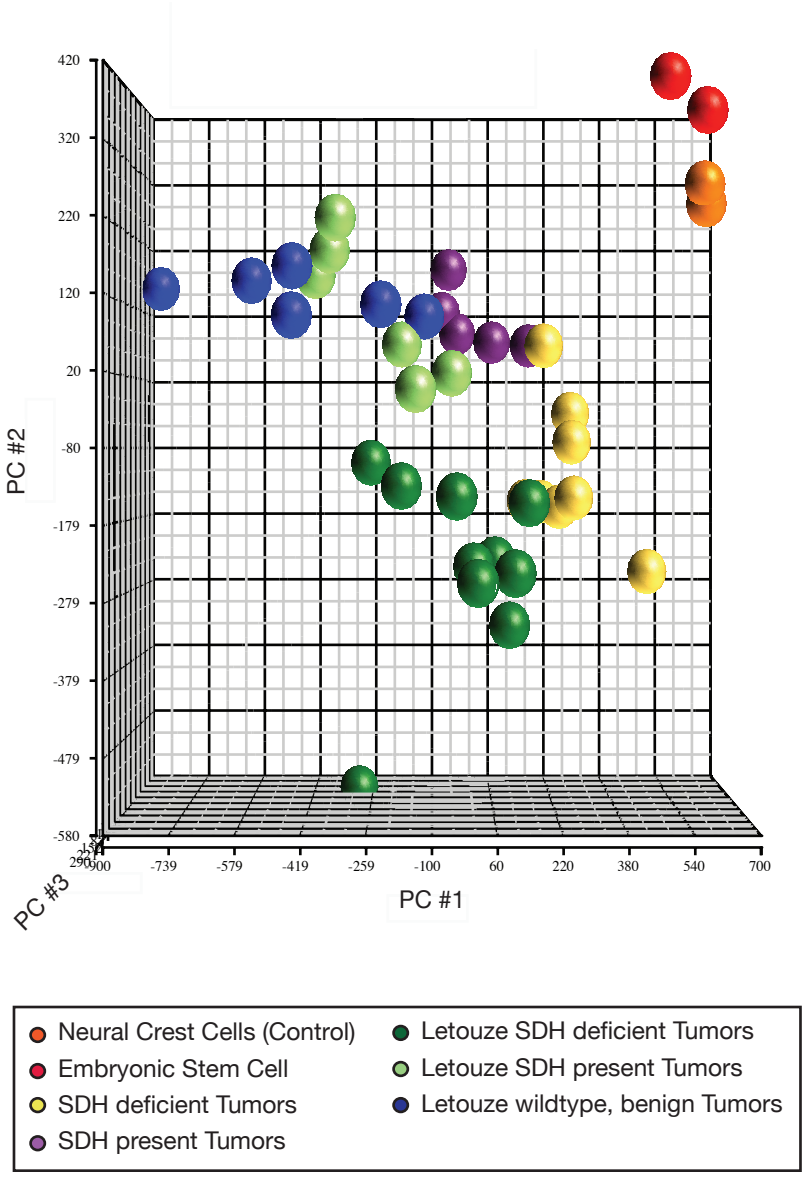


Figure 4.5) Principle component analysis plot of SDH deficient, SDH present PGL tumors and controls for all methylation probes in three dimensions. PC = Principle component (Letouze *et al.* SDH present tumors include tumors with mutations in VHL, NF1, RET and MAX. Letouze *et al.* benign tumors include wildtype, benign tumors. Refer to Table 4.1 for more information on tumors)

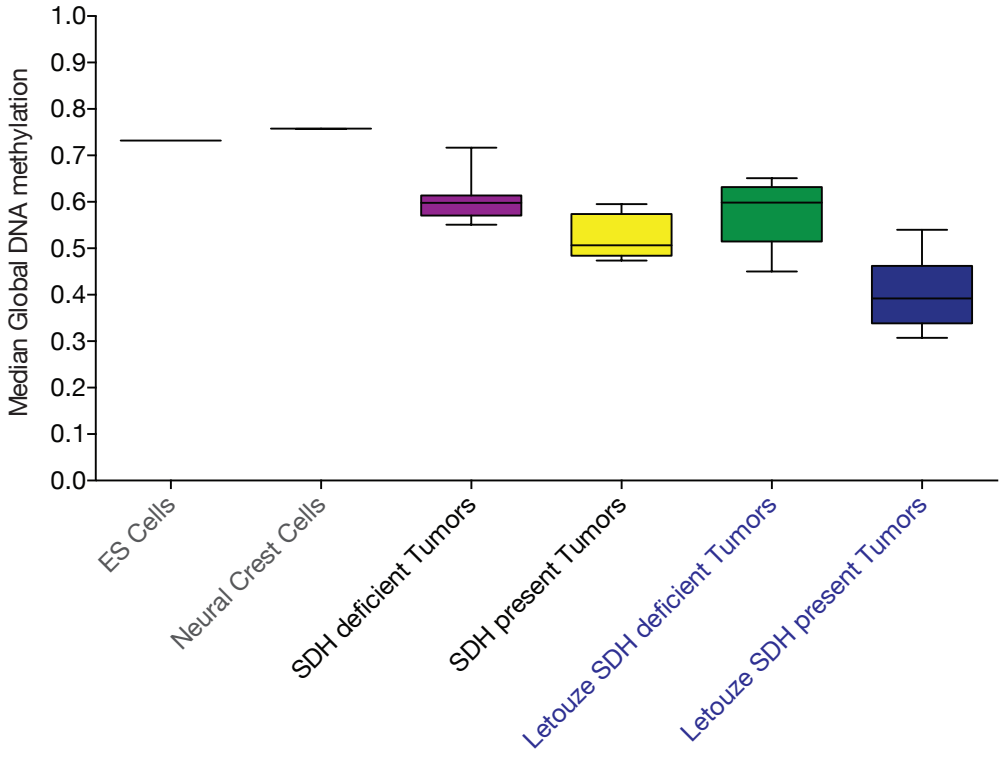


Figure 4.6) Box plot of median global DNA methylation for 485,579 probes on 450K methylation array for controls (in grey) SDH Deficient, SDH Present Paraganglioma tumors (in black), and Letouze *et al.* PGLs (in blue). (NCCs = neural crest cells, ES cells = Embryonic Stem Cells)

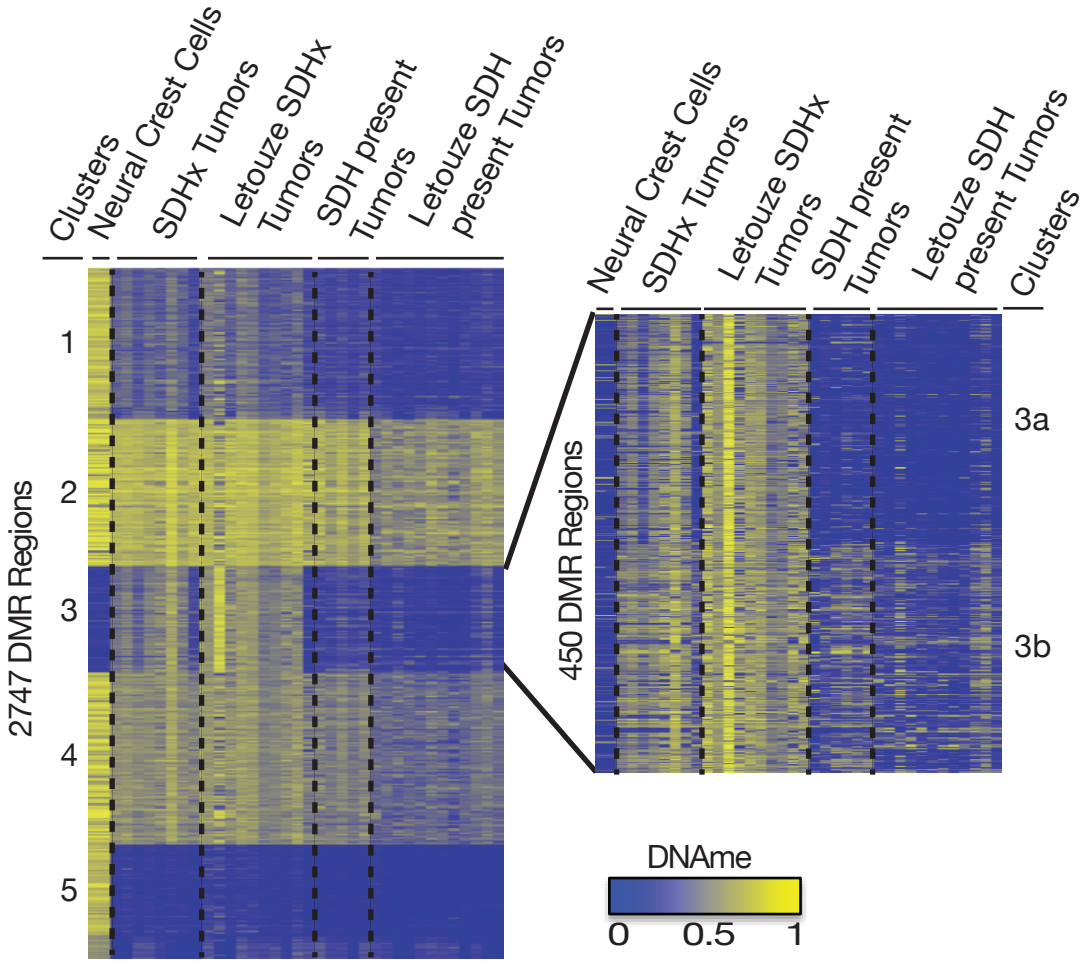


Figure 4.7) Five distinctive clusters (k-means) in regard to DNA methylation at CpG Islands when comparing PGLs to NCCs. k-means clustering (k = 5) of mean fraction CG methylation, p-value with FDR < 0.05, change in fraction methylation >=15%. Clusters 1, 2, 4 and 5 lose DNA methylation in tumors compared to NCCs. Cluster 3, gains DNA methylation in tumors and was further partitioned in two distinct clusters using k-means (k = 2). (DMR = Differentially methylated regions, SDHx = SDH Deficient tumors, PGLs = Paragangliomas, NCCs = Neural Crest Cells, Letouze *et al.* SDH Present tumors include tumors with mutations in VHL, NF1, RET MAX and wildtype tumors. Refer to Table 4.1 for more sample information)

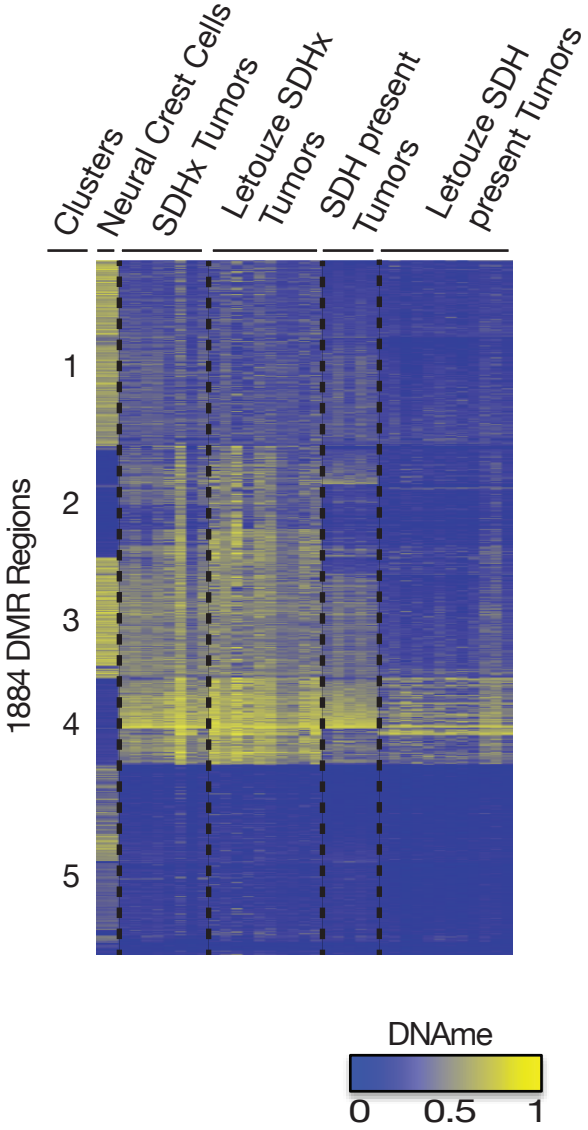


Figure 4.8) Five distinctive clusters (k-means) in regard to DNA methylation at CpG Shores when comparing PGLs to NCCs. k-means clustering (k = 5) of mean fraction CG methylation, p-value with FDR < 0.05, change in fraction methylation >=15%. Clusters 1,3, and 5 gain DNA methylation in tumors, whereas clusters 2 and 4 lose DNA methylation in tumors compared to NCCs. (DMR = Differentially methylated regions, SDHx = SDH Deficient tumors, PGLs = Paragangliomas, NCCs = Neural Crest Cells).

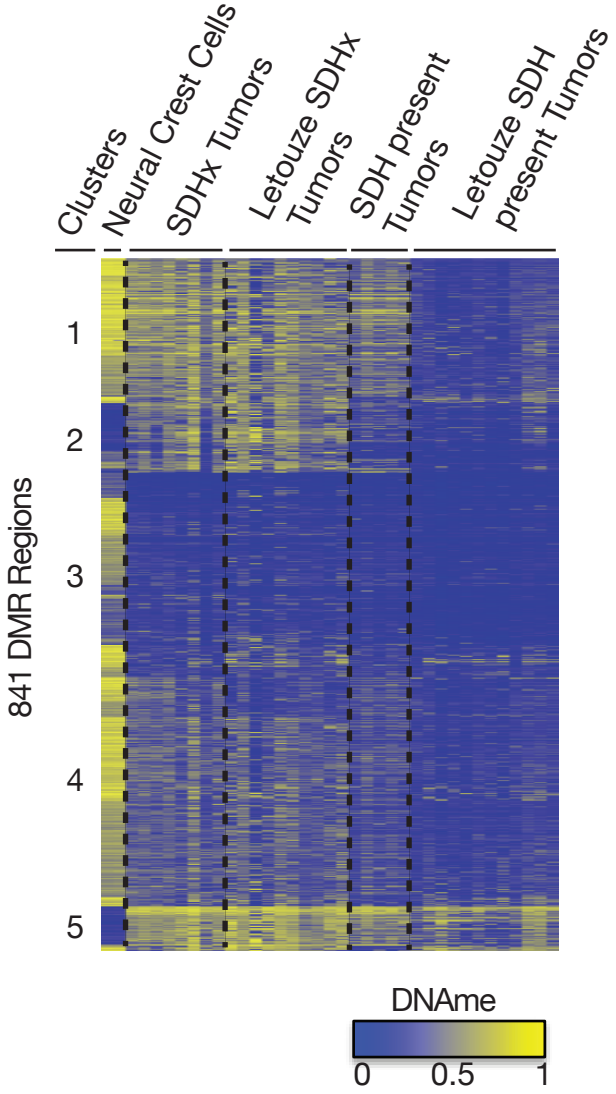


Figure 4.9) Five distinctive clusters (k-means) in regard to DNA methylation at Distant Regions when comparing PGLs to NCCs. k-means clustering (k = 5) of mean fraction CG methylation, p-value with FDR < 0.05, change in fraction methylation >=15%. Clusters 2, and 5 gain DNA methylation in tumors, whereas clusters 1,3 and 4 lose DNA methylation in tumors compared to NCCs. (DMR = Differentially methylated regions, SDHx = SDH Deficient tumors, PGLs = Paragangliomas, NCCs = Neural Crest Cells).

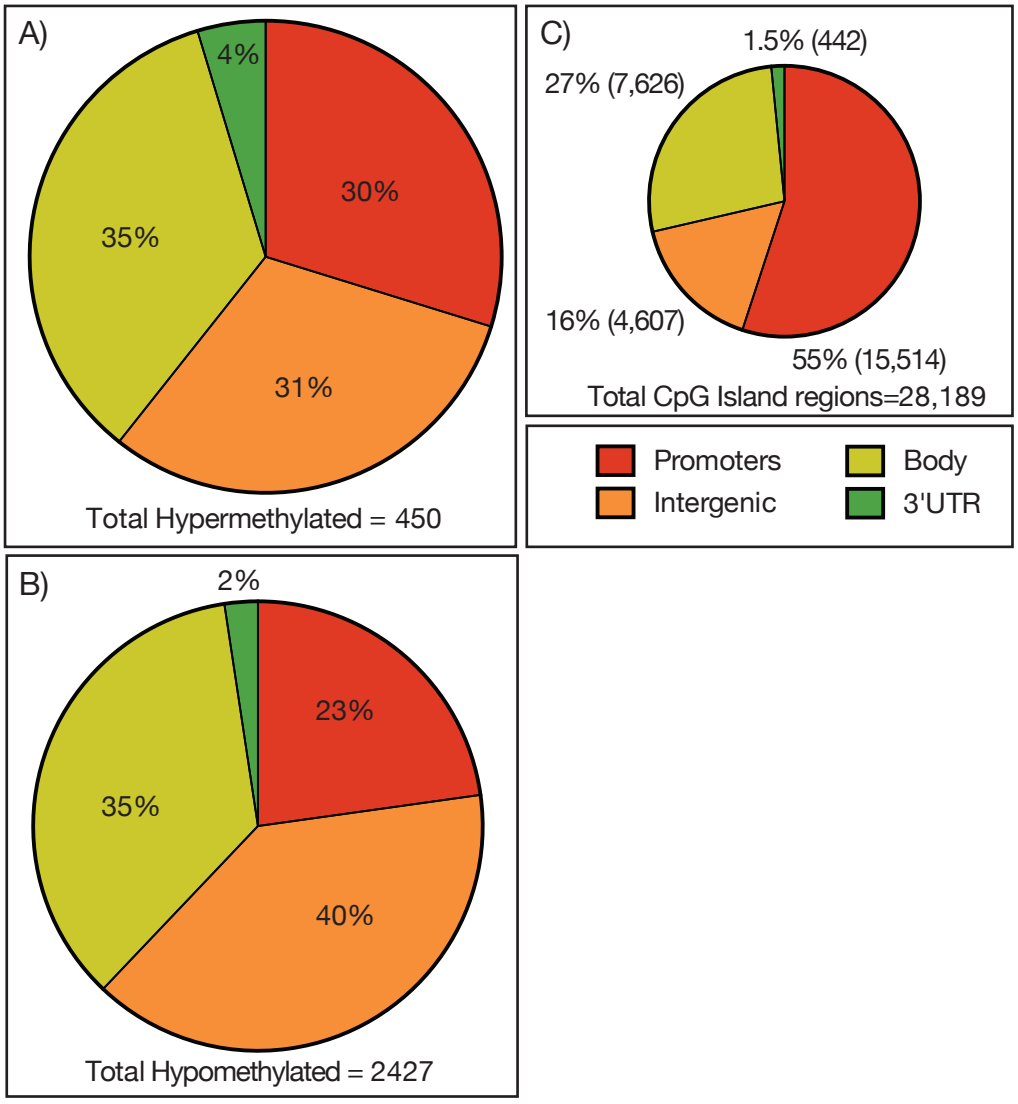


Figure 4.10) Genomic distribution of differentially methylated regions in CpG Islands. A) represents genomic distribution of cluster 3 DMRs in CpG Islands (gaining methylation; Figure 4.7). B) represents clusters 1,2,4 and 5 DMRs (losing methylation; Figure 4.7). C) shows distribution of total CpG Island regions (containing three probes or more) on 450K array; DMRs determined with p-value with FDR < 0.05, change in fraction methylation >= 15%.

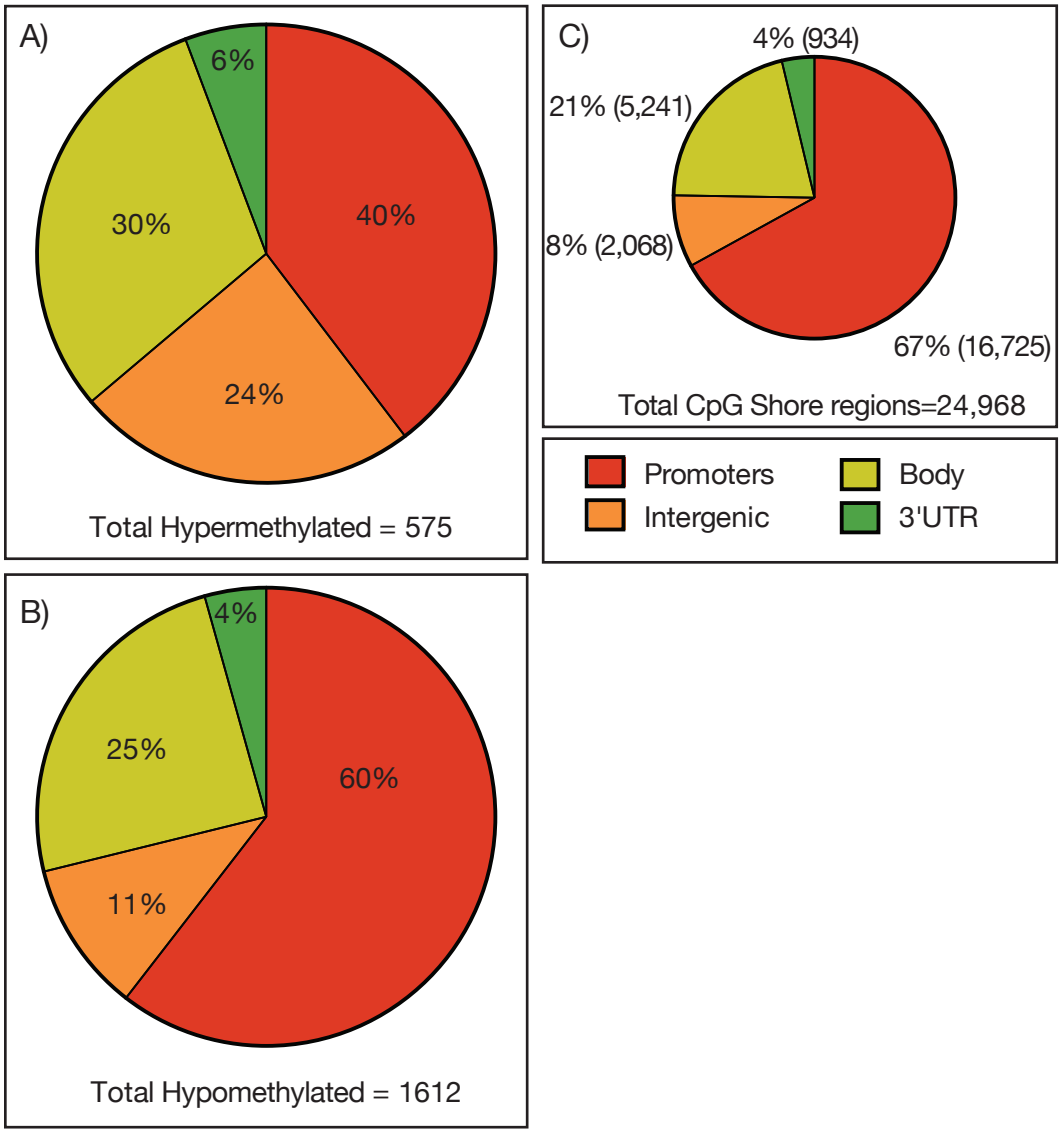


Figure 4.11) Genomic distribution of differentially methylated regions in CpG Shores. A) shows distribution of clusters 2 and 4 DMRs in CpG Shores (gaining DNA methylation, Figure 4.8). B) shows distribution of clusters 1, 3 and 5 (losing DNA methylation, Figure 4.8). C) represents all CpG Shore regions (three probes or more) on 450K Illumina Methylation Array; DMRs determined with p-value with FDR < 0.05, change in fraction methylation >= 15%.

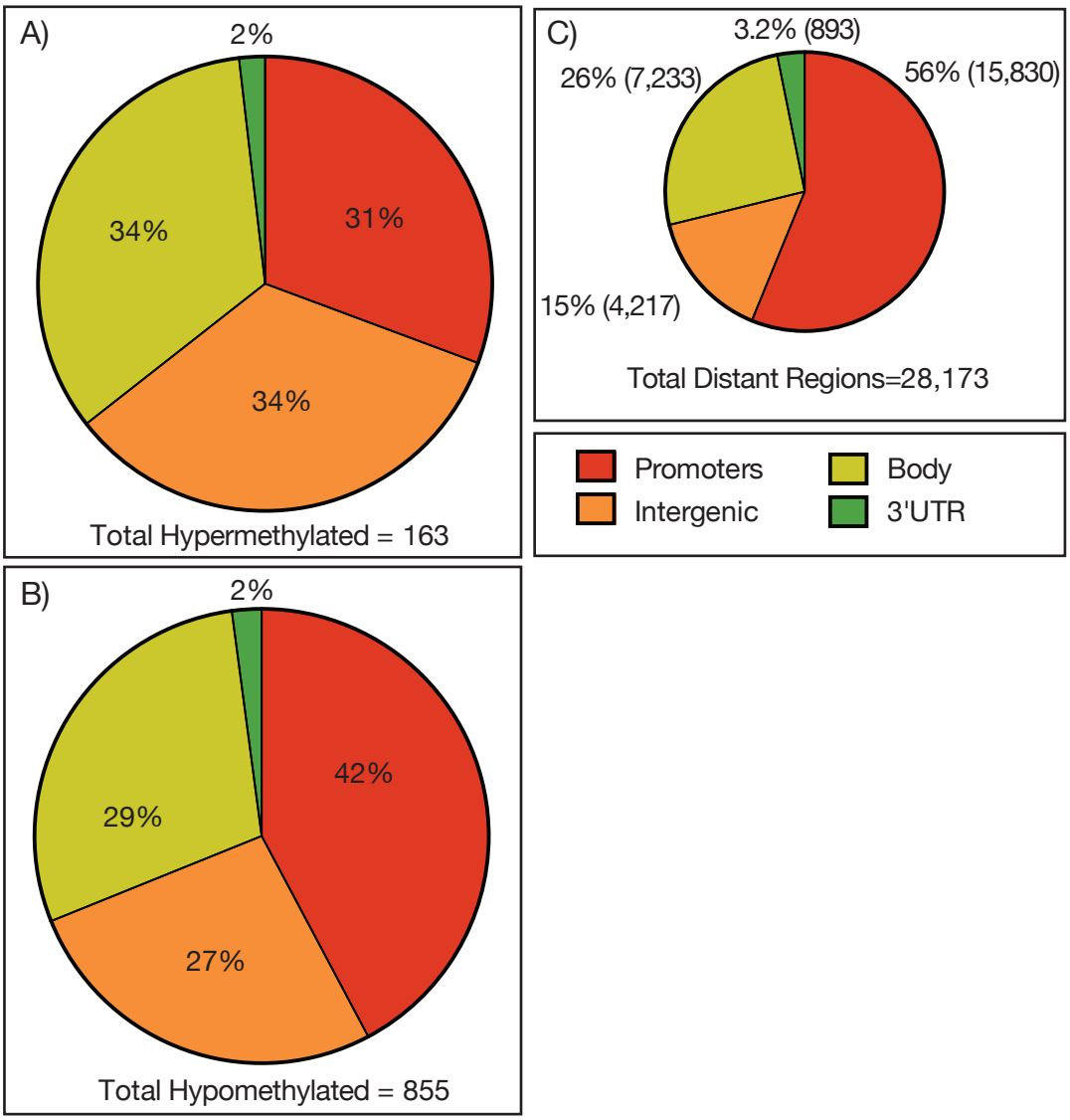


Figure 4.12) Genomic distribution of differentially methylated regions in Distant Regions. A) shows distribution of clusters 2 and 5 DMRs in Distant Regions (gaining DNA methylation, Figure 4.9) and B) represents clusters 1, 3 and 4 (losing DNA methylation, Figure 4.9). C) represents all Distant Regions (three probes or more) on 450K Illumina Methylation Array. DMRs determined with p-value with FDR < 0.05, change in fraction methylation >= 15%.

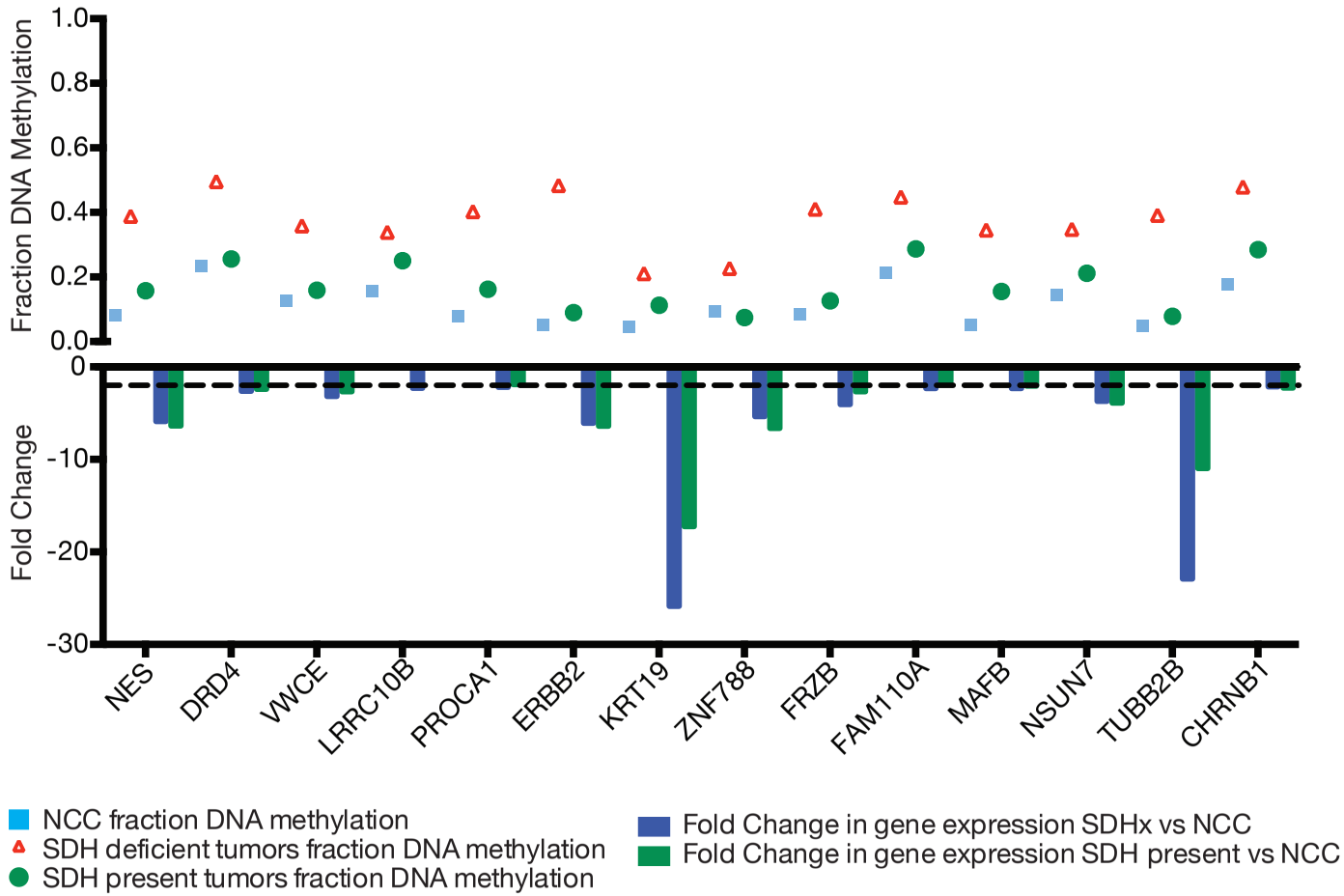


Figure 4.13) Genes from cluster 3a (Figure 4.7) that lose expression with a gain in DNA methylation in their CpG Island promoters. Remarkably, SDH Present tumors follow similar trends in changes in gene expression regardless of DNA methylation changes. Transcription changes based on p-value with FDR < 0.05, fold change >= 2 in at least one of the tumor groups. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.

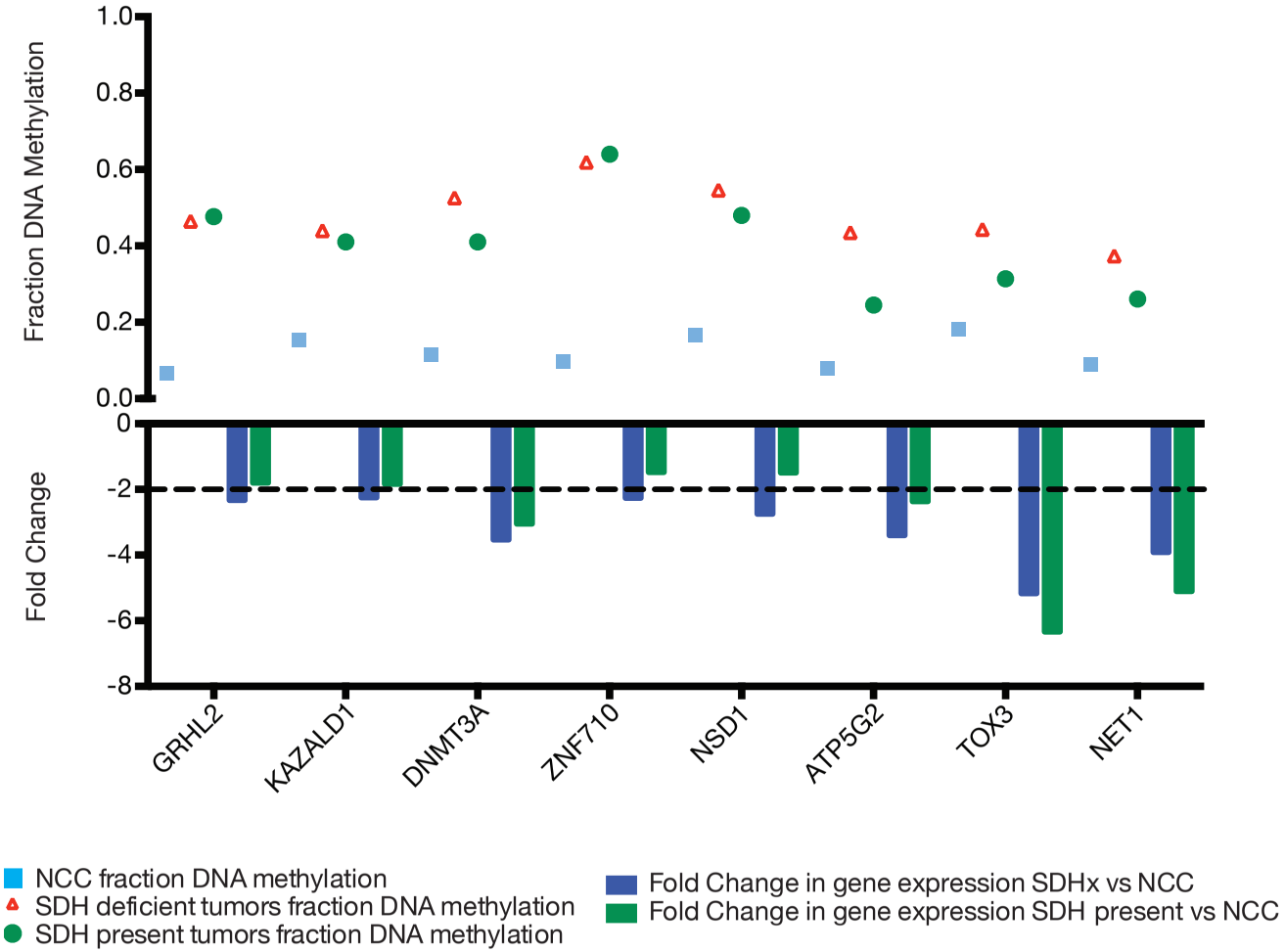


Figure 4.14) Genes from cluster 3b (Figure 4.7) that lose gene expression with a gain in DNA methylation in their CpG Island promoters in both SDH Deficient and SDH Present PGLs. Transcription changes based on p-value with FDR < 0.05, fold change >= 2 in at least one of the tumor groups. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.

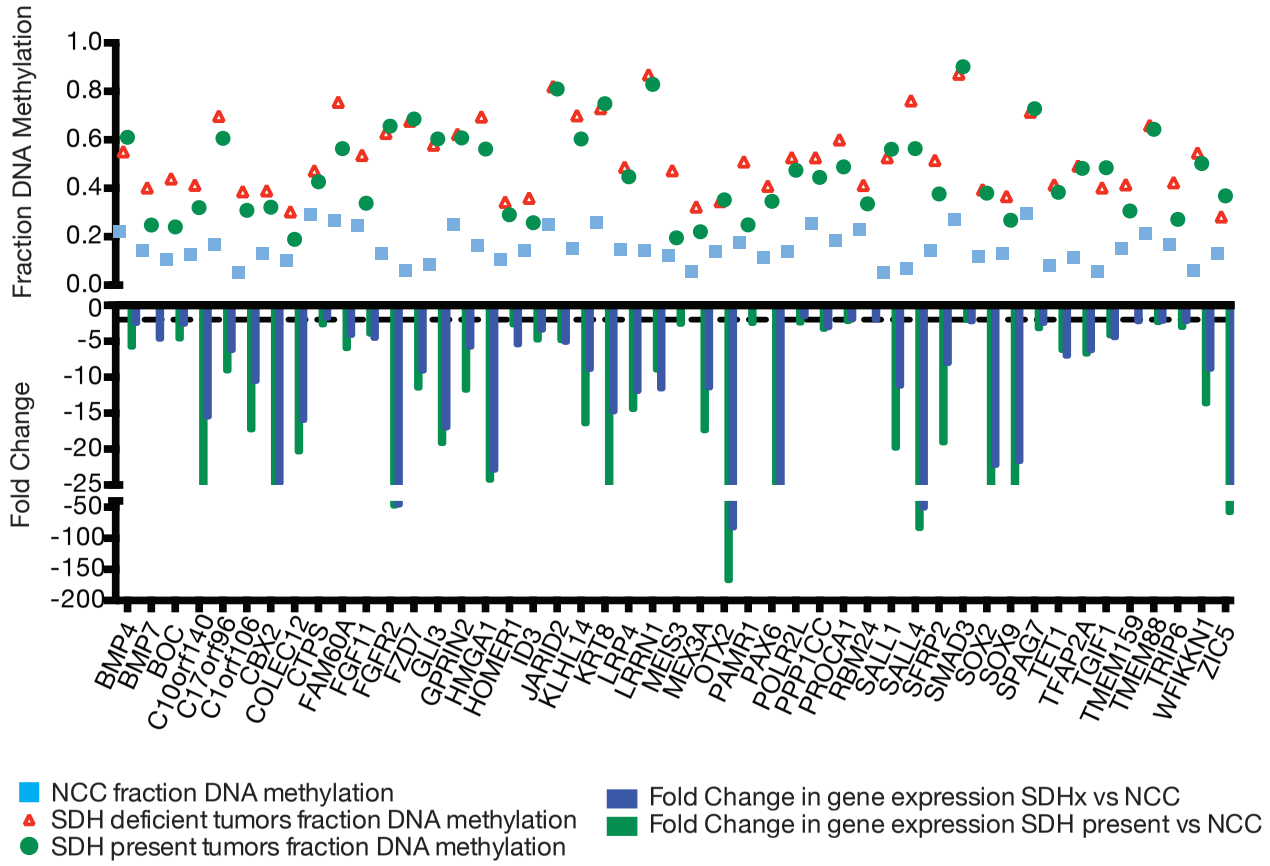


Figure 4.15) Genes that lose gene expression with a gain of DNA methylation in their CpG Shore promoters in SDHx tumors and SDH Present tumors compared to NCCs (clusters 2 and 4; Figure 4.8). Transcription changes based on p-value with FDR < 0.05, fold change >= 2 in at least one of the tumor groups. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.

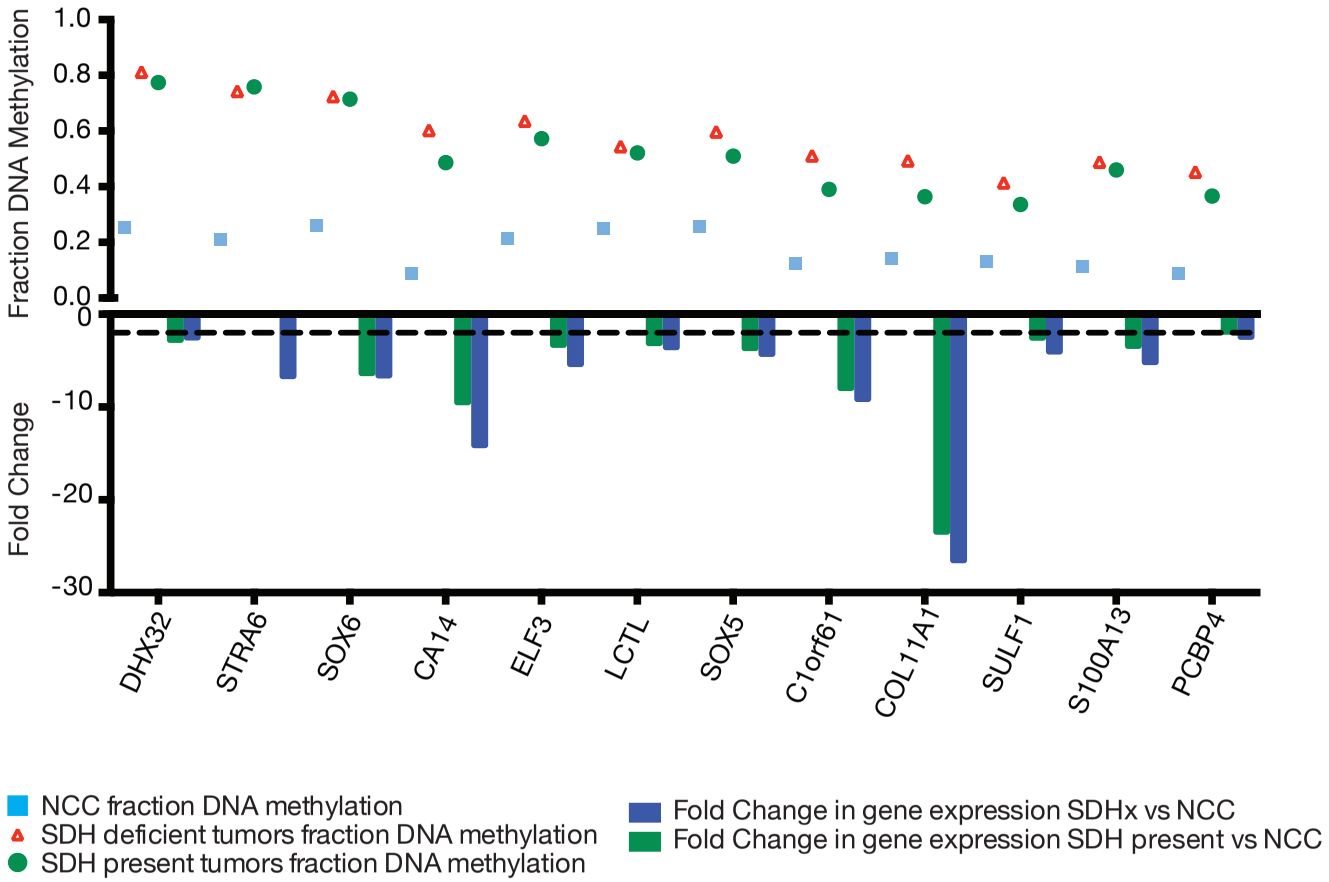


Figure 4.16) Genes that lose gene expression with a gain of DNA methylation in the promoters of Distant Regions in both SDHx tumors and SDH Present tumors compared to NCCs (from clusters 2 and 5; Figure 4.9). Interestingly, SDH present tumors follow similar trends in changes in gene expression regardless of DNA methylation changes. Transcription changes based on p-value with FDR < 0.05, fold change >= 2 in at least one of the tumor groups. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.

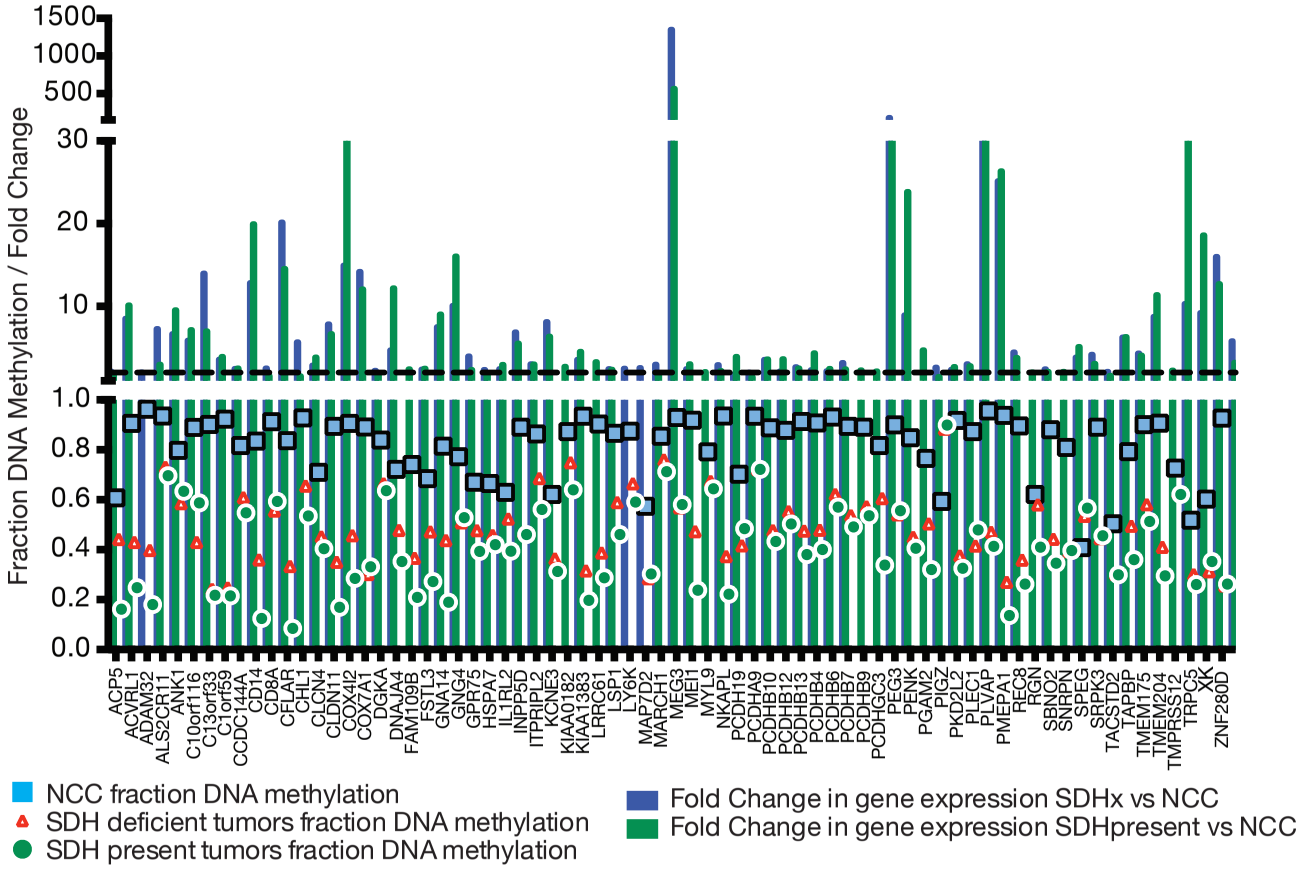


Figure 4.17) Graph shows fold change expression of genes significantly losing DNA methylation in their promoters from clusters 1, 2, 4 and 5 (Figure 4.7) in their CpG Island promoters in both SDHx and SDH Present tumors compared to NCCs. Transcription changes based on p-value with FDR < 0.05, fold change >= 2 in at least one of the tumor groups. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.

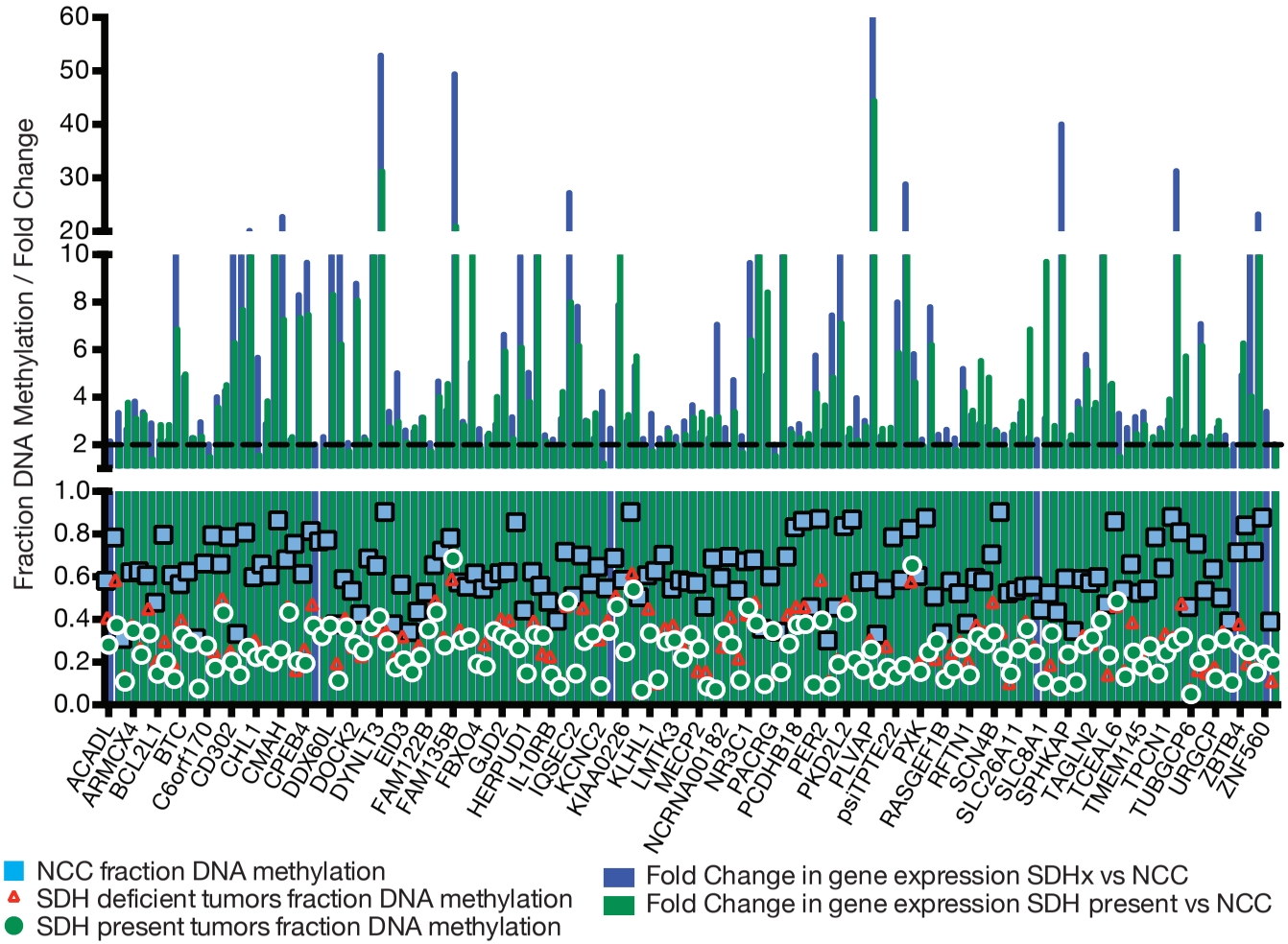


Figure 4.18) Genes that lose DNA methylation in their CpG Shore promoters (clusters 1, 3 and 5; Figure 4.8) and increase in transcription. Transcription changes based on p-value with FDR < 0.05, fold change >= 2 in at least one of the tumor groups. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.

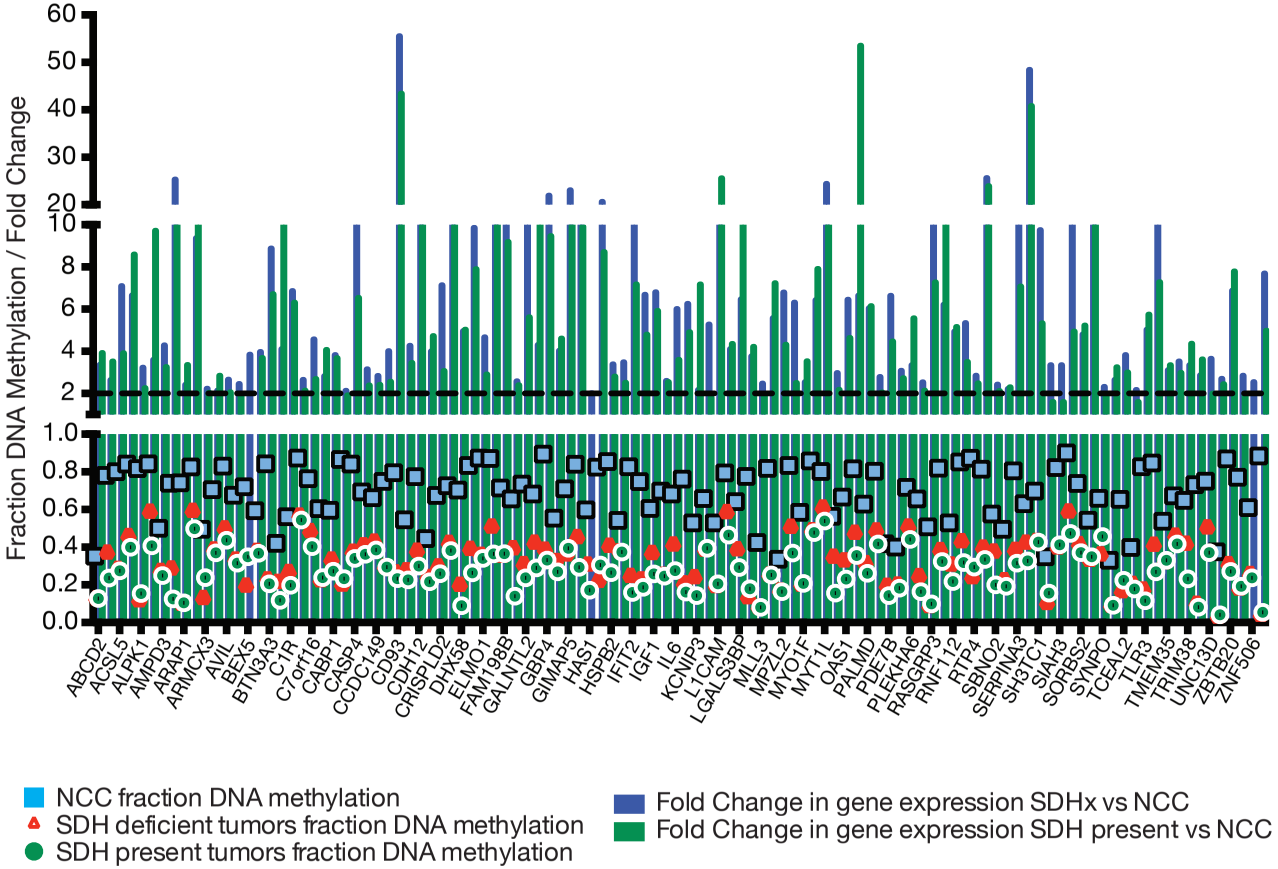


Figure 4.19) Genes that lose DNA methylation in their Distant Region promoters (clusters 1, 3 and 4; Figure 4.9) and increase in transcription. Transcription changes based on p-value with FDR < 0.05, fold change >= 2 in at least one of the tumor groups. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.

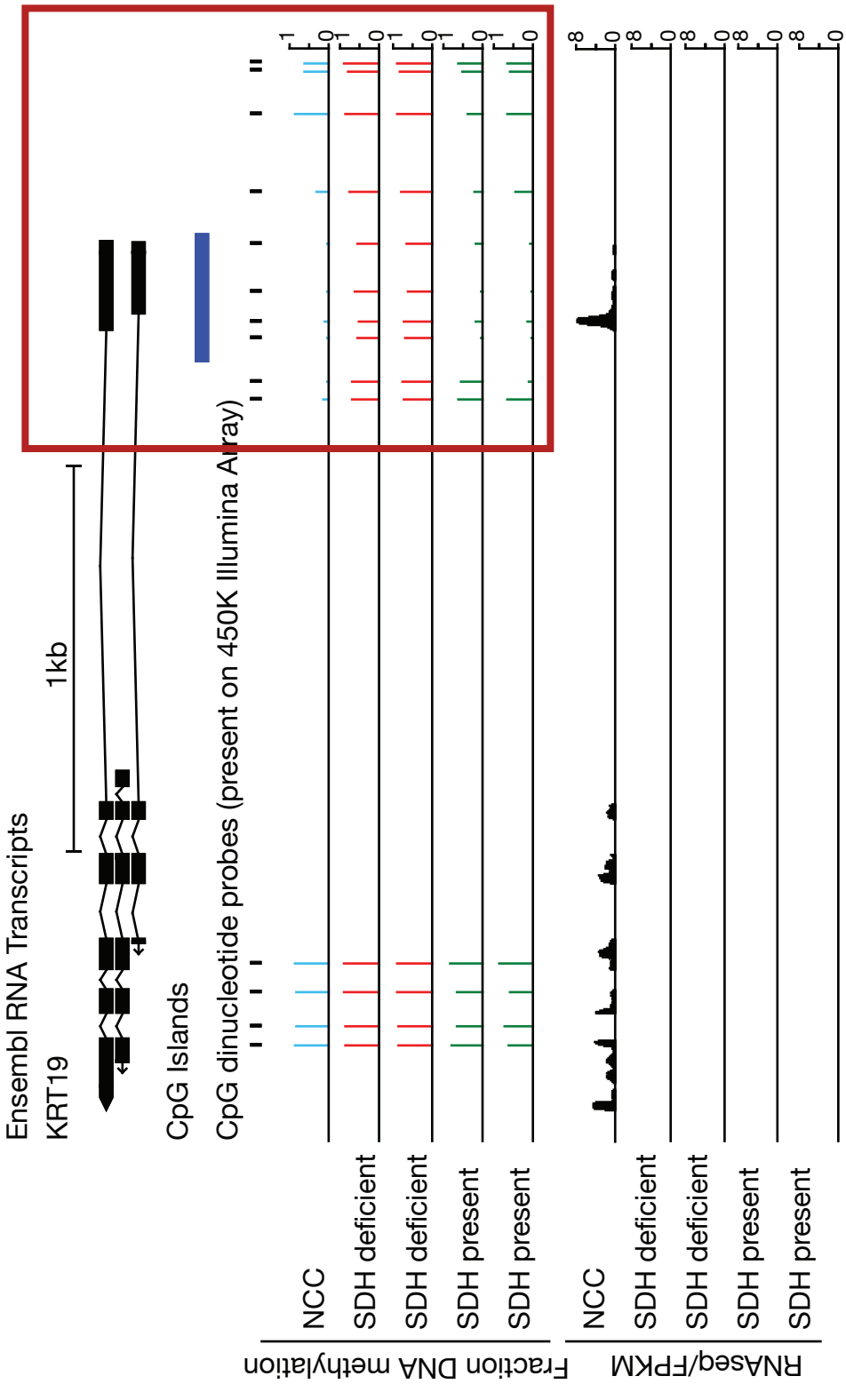


Figure 4.20) Genome snapshot of KRT19. CpG Island in the promoter region gains DNA methylation exclusively in SDH Deficient tumors compared to NCCs. Gene is repressed in both SDH Deficient and Present tumors.

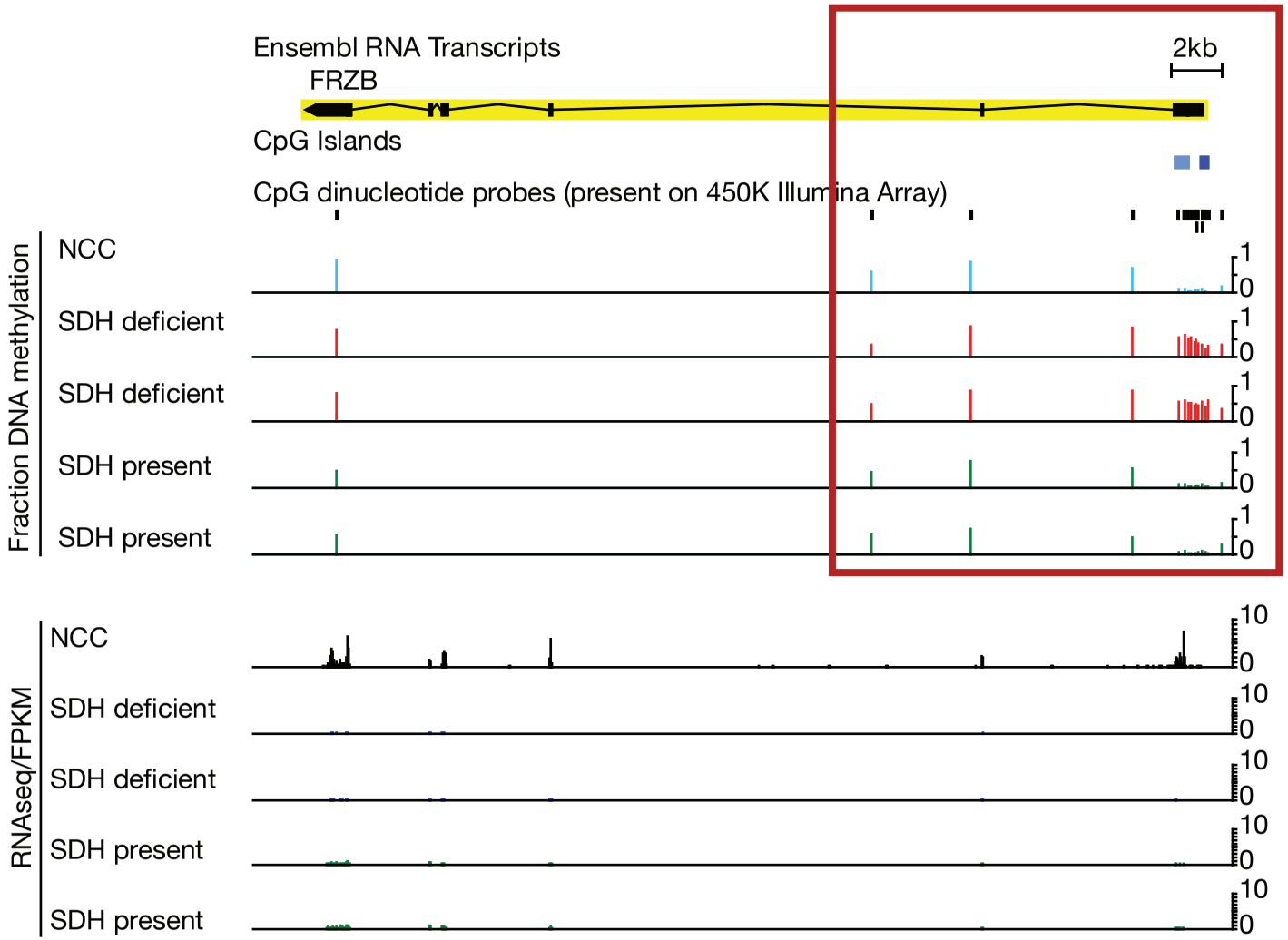


Figure 4.21) Genome snapshot of FRZB. CpG Island in the promoter region gains DNA methylation exclusively in SDH Deficient tumors compared to NCCs. Gene is repressed in both SDH Deficient and Present tumors.

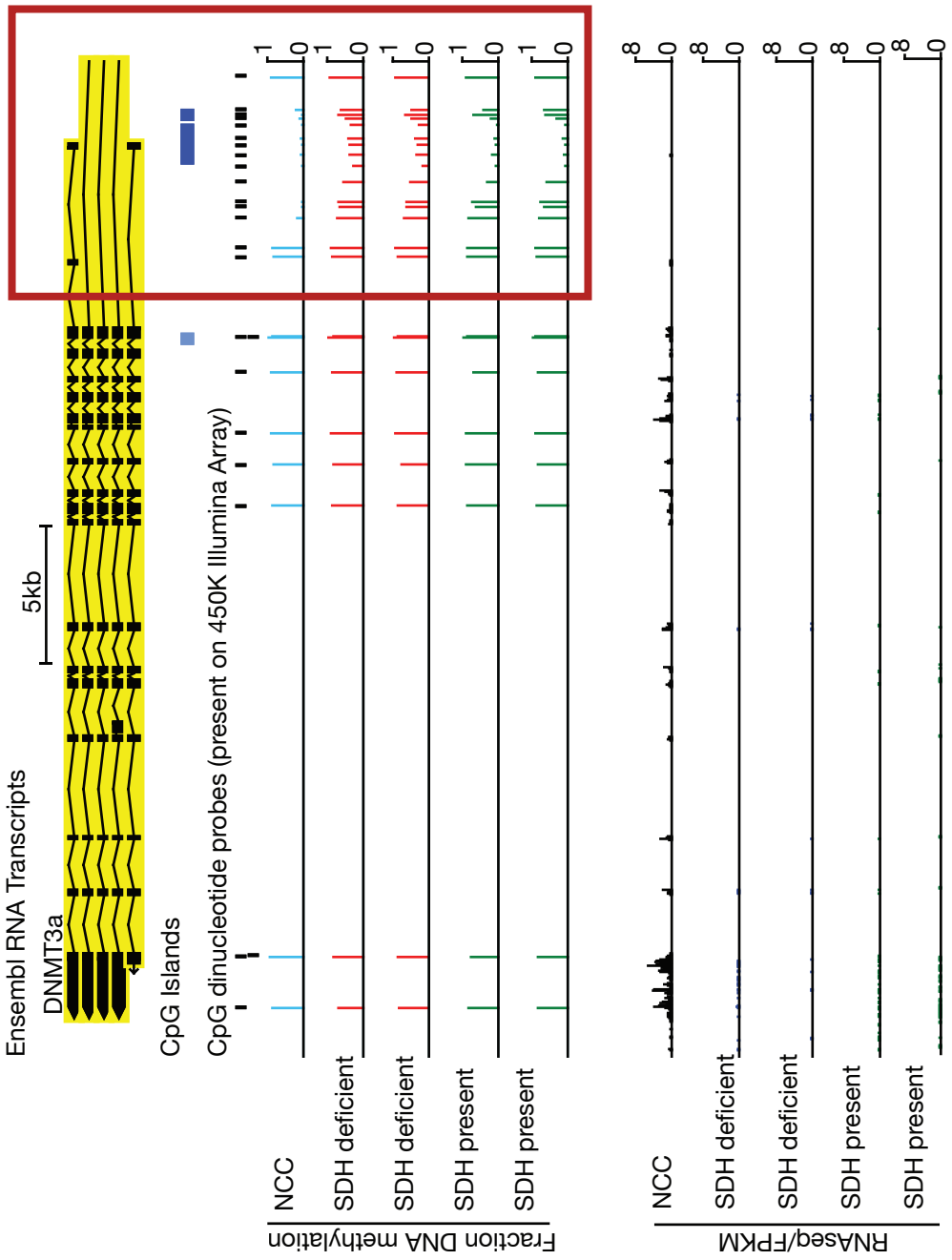


Figure 4.22) Genome snapshot of DNMT3a. CpG Island in the promoter region gains DNA methylation in both SDH Deficient and Present PGLs compared to NCCs. Gene is repressed in both SDH Deficient and Present tumors.

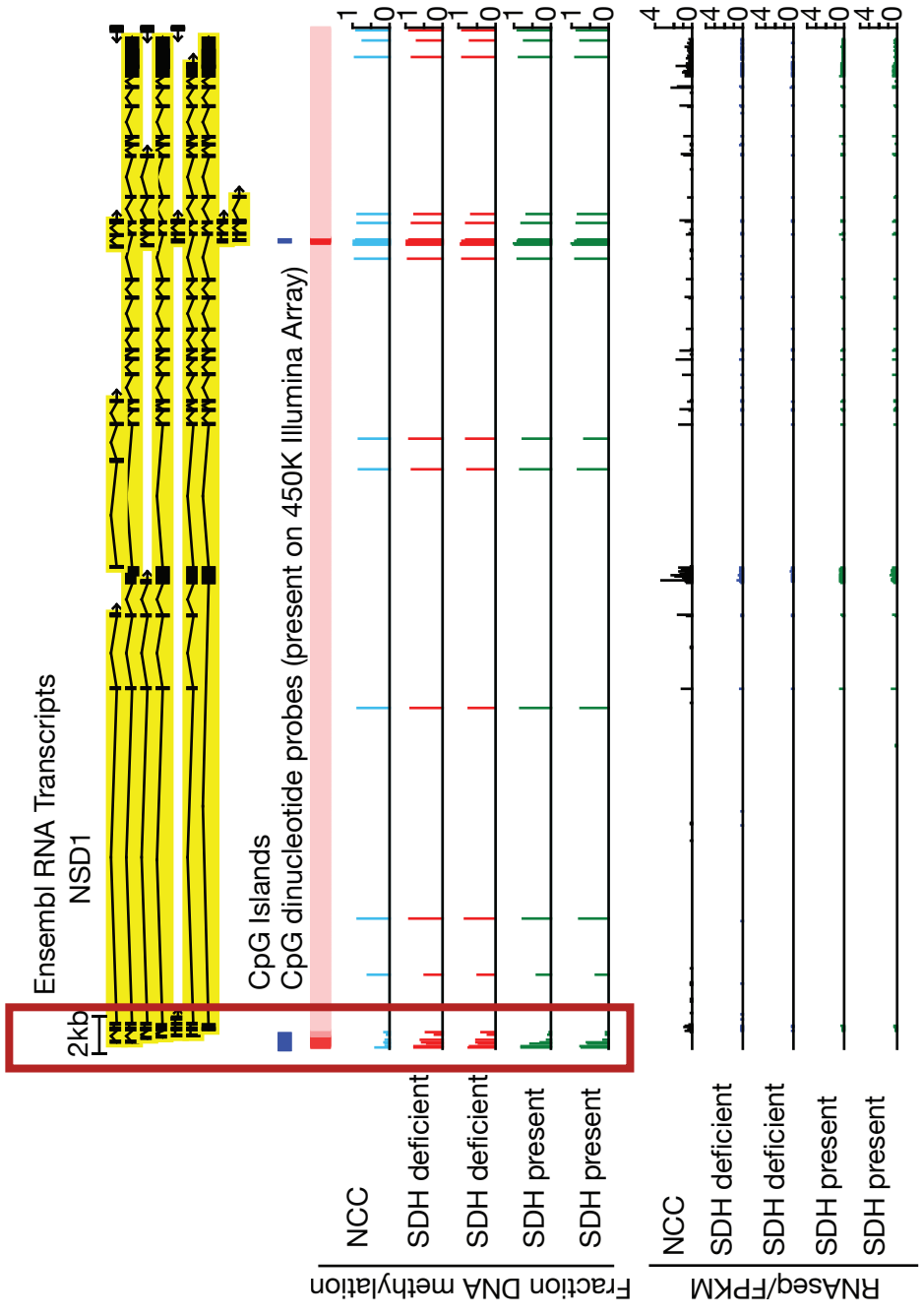


Figure 4.23) Genome snapshot of NSD1. CpG Island in the promoter region gains DNA methylation in both SDH Deficient and Present PGLs compared to NCCs. Gene is repressed in both SDH Deficient and Present tumors.

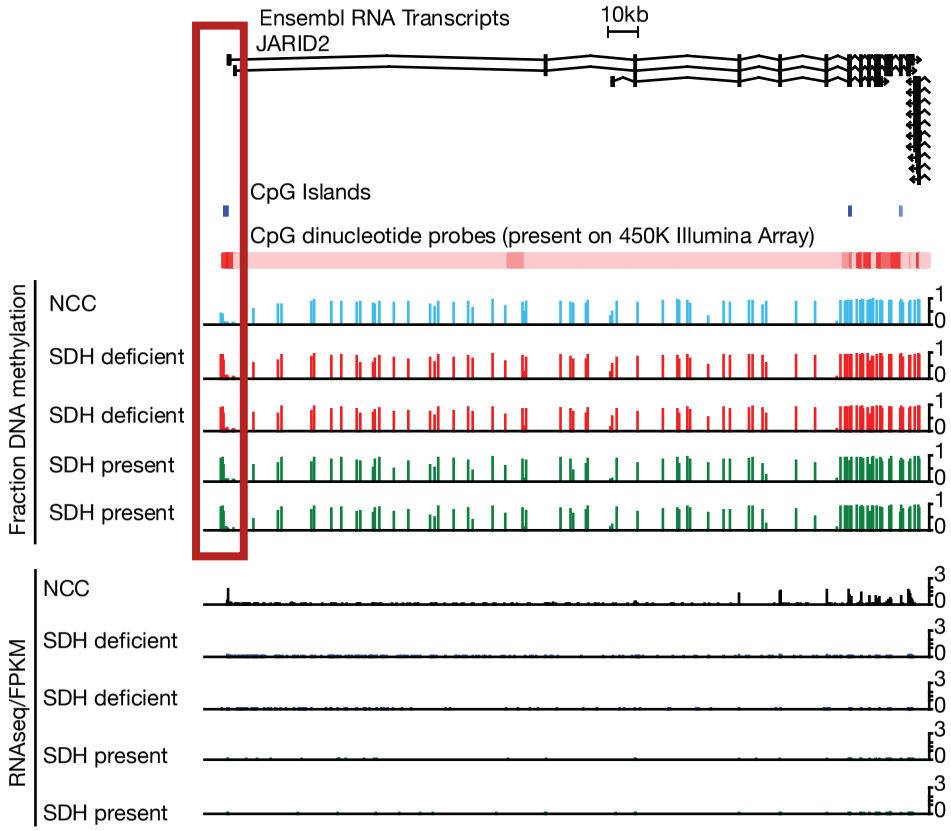


Figure 4.24) Genome snapshot of JARID2. CpG Shore in the promoter region gains DNA methylation in both SDH Deficient and Present PGLs compared to NCCs. Gene is repressed in both SDH Deficient and Present tumors.

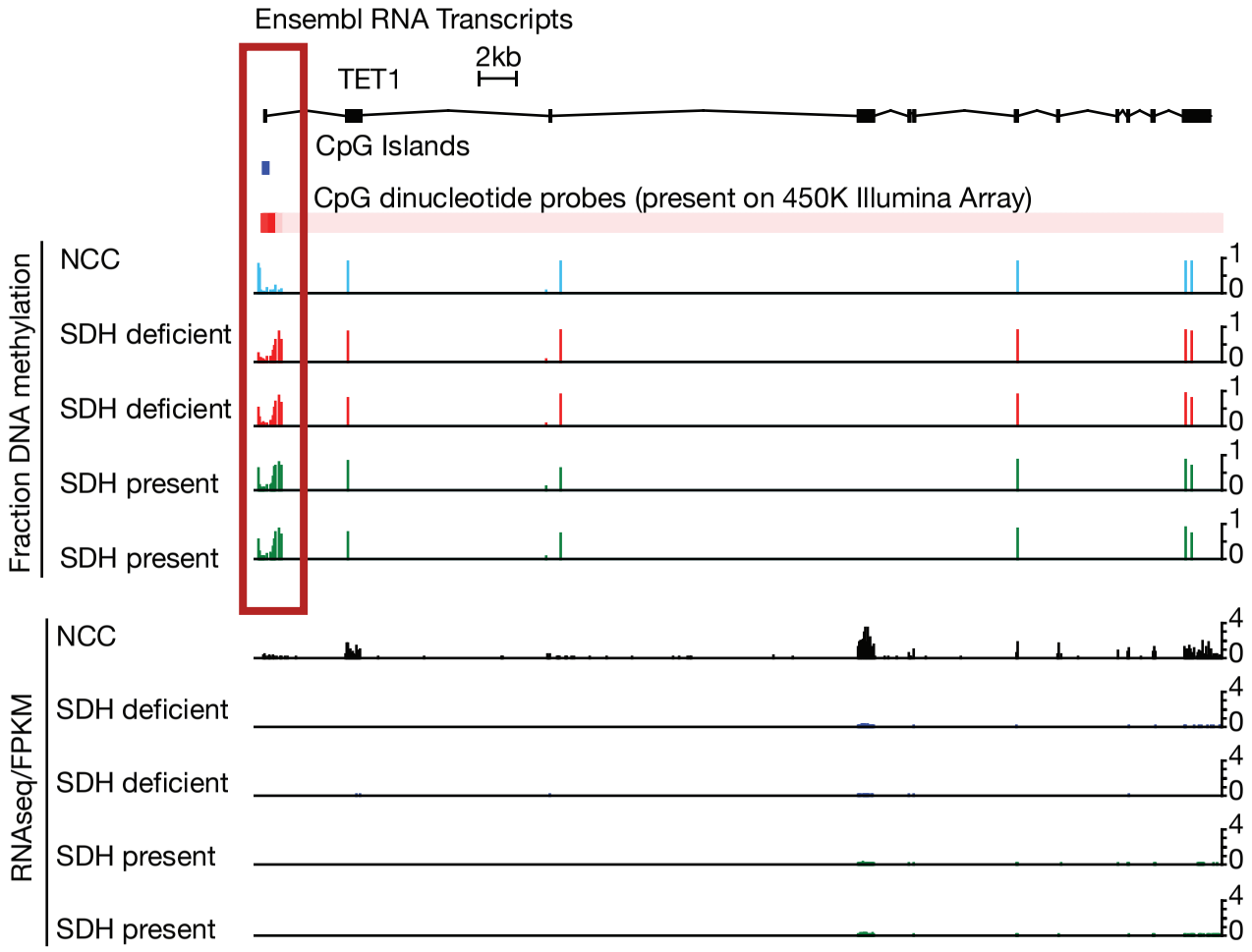


Figure 4.25) Genome snapshot of TET1. CpG Shore in the promoter region gains DNA methylation in both SDH Deficient and Present PGLs compared to NCCs. Gene is repressed in both SDH Deficient and Present tumors.

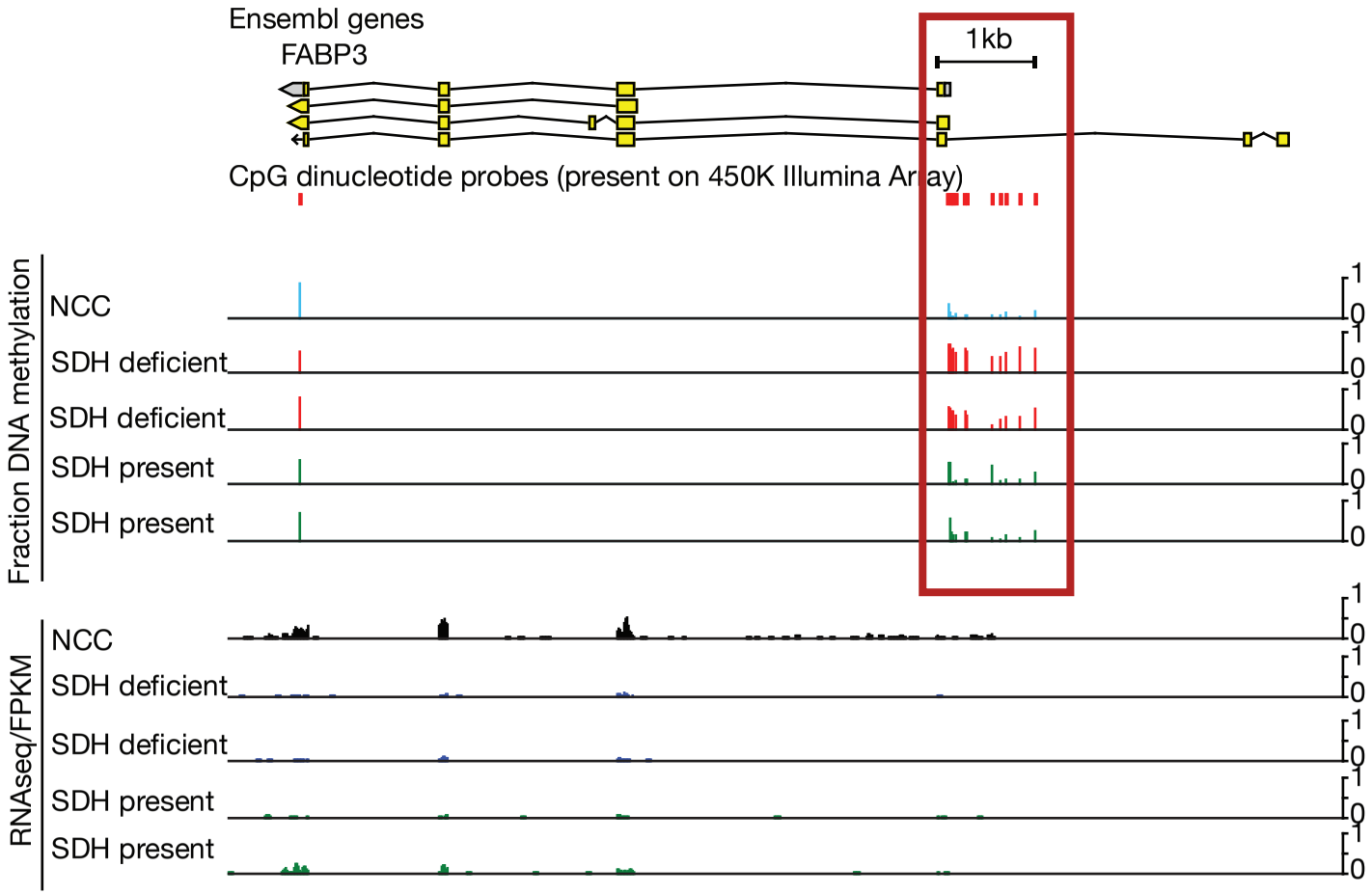
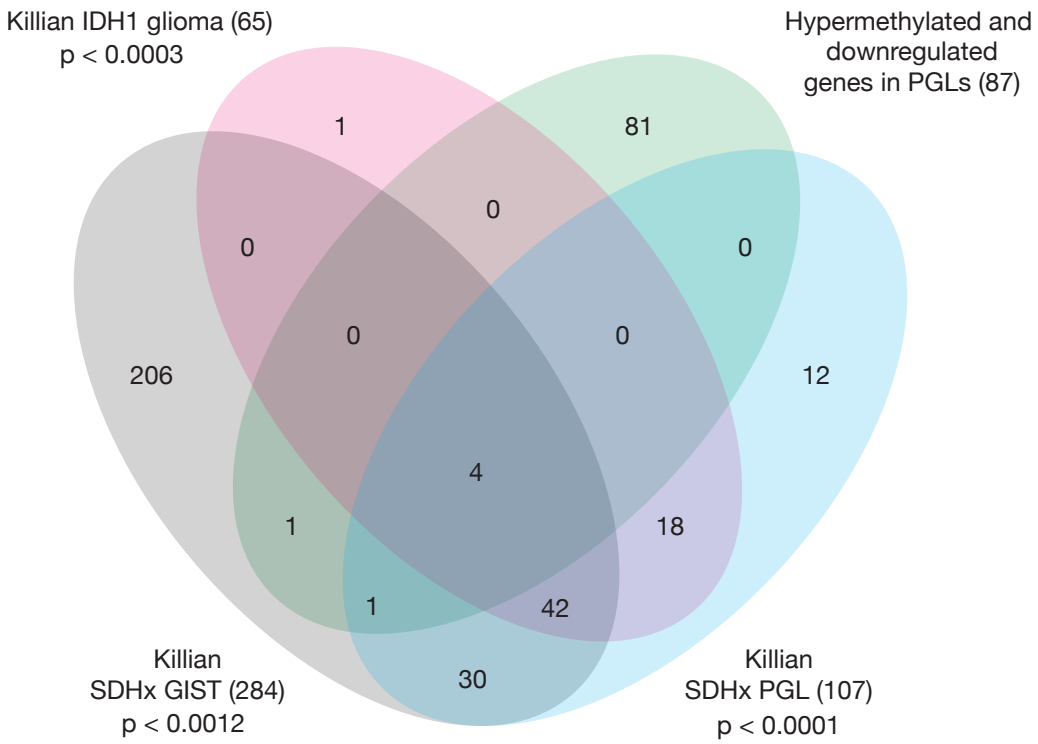


Figure 4.26) Genome snapshot of FABP3. Promoter region (not associated with CpG Island) gains DNA methylation in both SDH Deficient and Present PGLs compared to NCCs. Gene is repressed in both SDH Deficient and Present tumors.



4 genes in common: BMP4, FABP3, FRZB, TRIP6

Figure 4.27) Venn diagram showing genes that are hypermethylated and downregulated in PGLs and those that are hypermethylated in IDH1 gliomas compared to glia, SDHx GISTs compared to muscularis and SDHx PGLs compared to adrenal (Killian *et al.*). p-values indicated are from pairwise comparisons of each list to the genes hypermethylated in PGLs from our study.

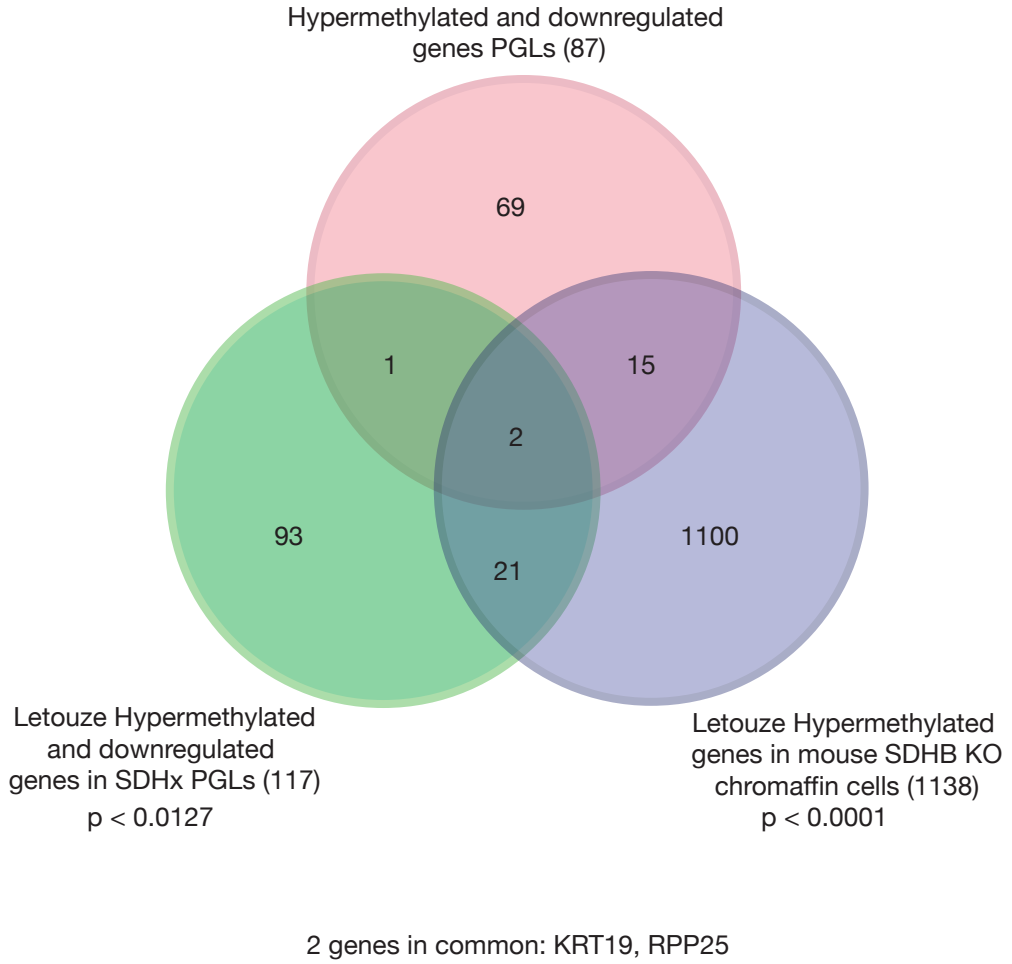


Figure 4.28) Venn diagram showing genes that are hypermethylated and downregulated in PGLs and those that are hypermethylated and downregulated in SDHx PGLs compared to SDH Present PGLs and SDHB knockout mouse chromaffin cells compared to wildtype (Letouze *et al.*). p-values indicated are from pairwise comparisons of each list to the genes hypermethylated in PGLs from our study.

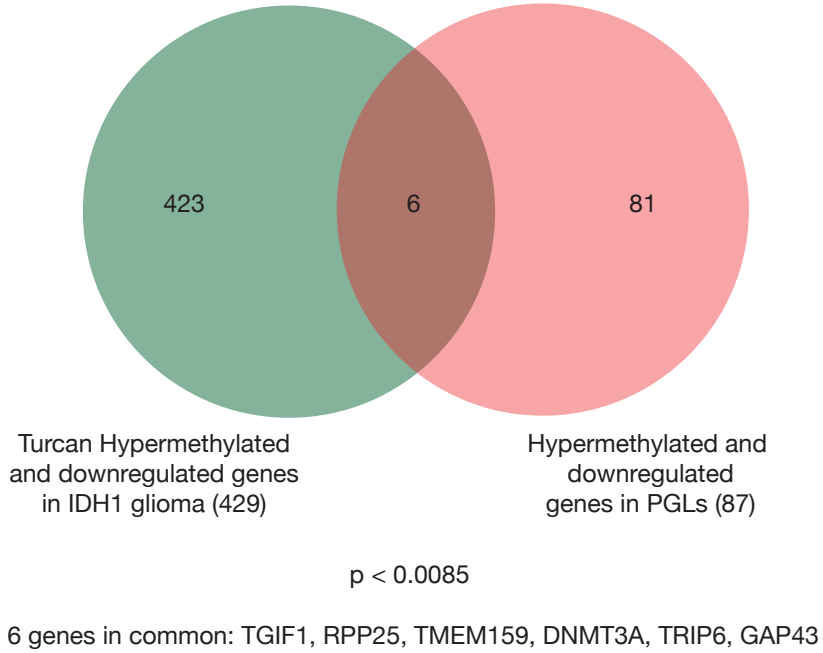
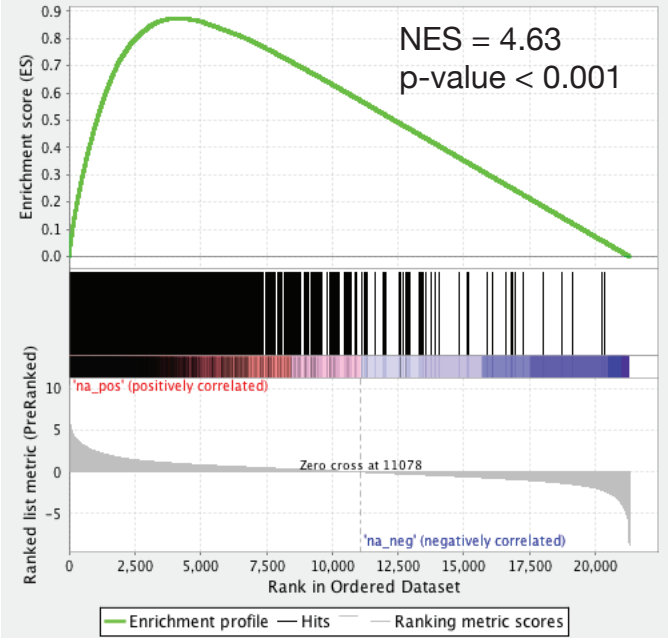


Figure 4.29) Venn diagram showing genes that are hypermethylated and downregulated in PGLs and those that are hypermethylated and downregulated in IDH1 gliomas (Turcan *et al.*). p-values indicated are from pairwise comparisons of each list to the genes hypermethylated in PGLs from our study.



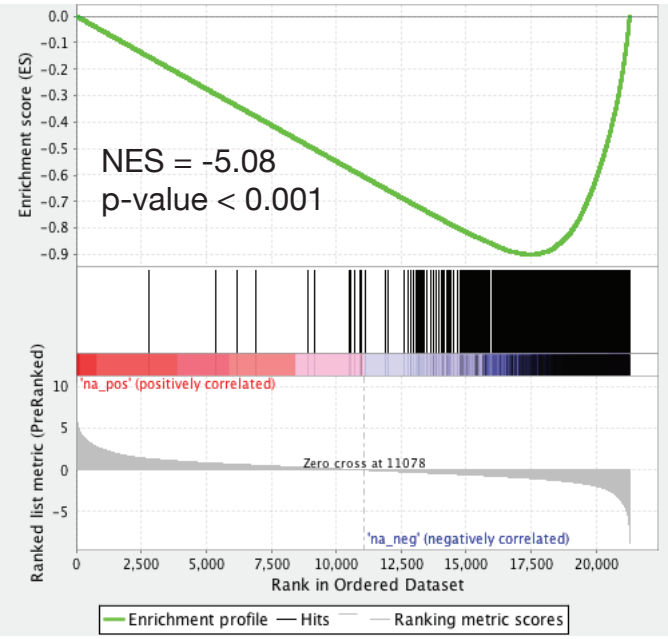


Figure 4.30) Gene set enrichment analysis (GSEA) using genes regulated in SDH Deficient tumors (RNA-seq) as the rank-ordered data set and the genes upregulated and downregulated in SDH Present tumors. (RNA-seq, FDR < 0.001, fold change >= 2 for upregulated genes and fold changes <= -2 for downregulated genes).

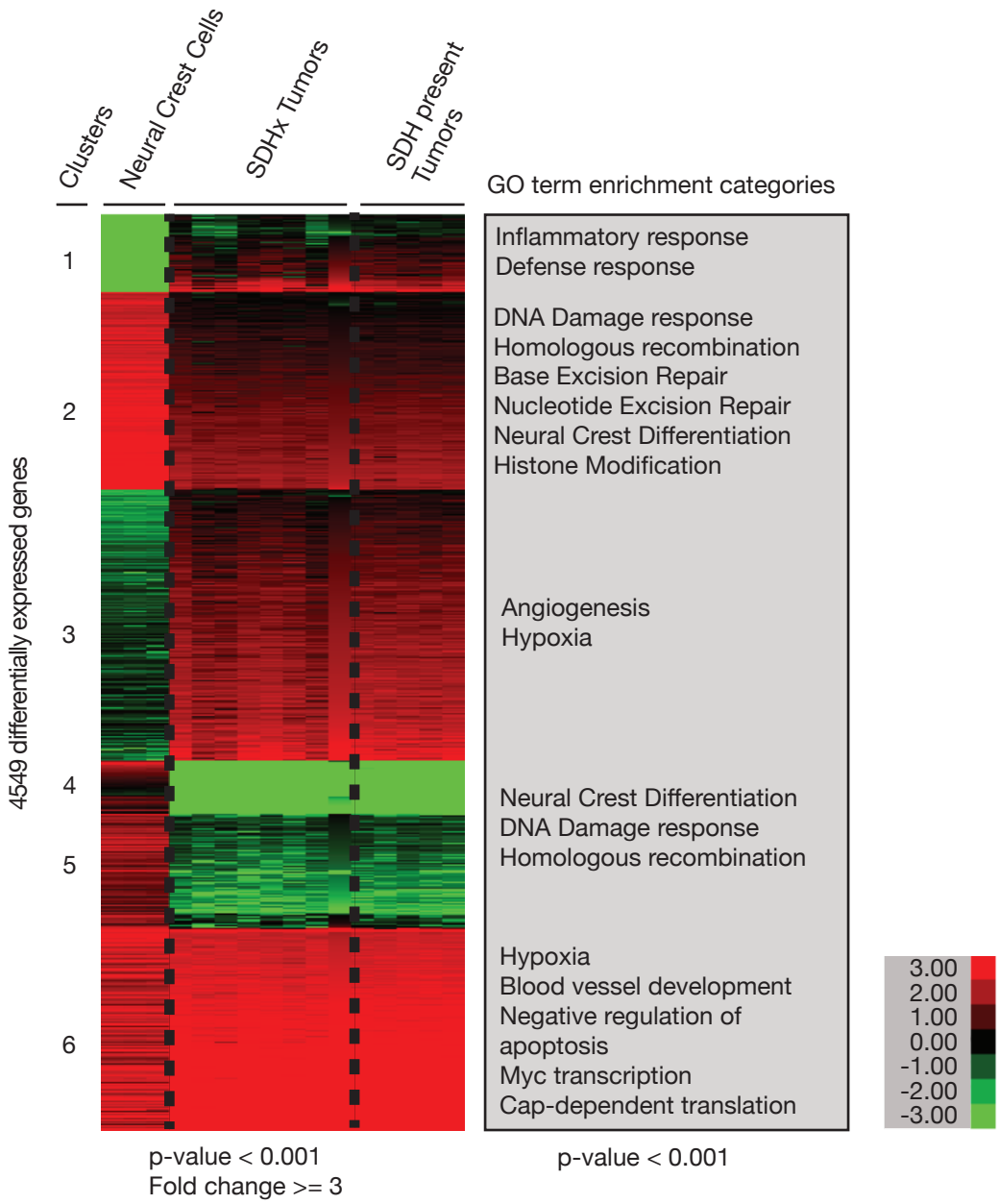


Figure 4.31) Six distinct clusters (k-means) of all genes differentially expressed in SDH Deficient and SDH Present as compared to NCCs. k-means clustering (k = 6) of log2 transformed FPKM values from RNA-seq data. Genes were filtered on an FDR < 0.001 and a fold change > = 2 when compared to NCCs. Right panel shows enriched Go-terms for different clusters with permuted p-value < 0.05.

Table 4.1

Clinical Characteristics of PGL Tumors

SDH Deficient tumors								
No	Name	Genotype	Germline / Somatic status	Location of PGL	SDHB staining	Gender	Age	Malignant (Metastatic or Recurrent) / Benign
1	05-00327	SDHB	Germline	Heart	Negative	Female	59	Malignant (Metastatic)
2	05-00303	SDHB	Germline	Carotid Body	Negative	Male	36	Malignant (Metastatic)
3	09-08423	SDHB	Germline	Peritoneum	Negative	Male	38	Benign
4	09-09342	SDHB	Germline	Carotid Body	Negative	Male	62	Benign
5	10-07068	SDHB	Germline	Glomus	Negative	Female	47	Benign
6	10-08139	SDHB	Germline	Vas deferens	Negative	Male	38	Malignant (Metastatic)
7	10-02832	SDHB	Germline	Carotid Body	Negative	Male	31	Benign
8	11-09937	SDHD	Germline	Carotid Body	Negative	Female	49	Malignant (Recurrent)

Table 4.1 continued

SDH Present tumors								
No	Name	Genotype	Germline / Somatic status	Location of PGL	SDHB staining	Gender	Age	Malignant (Metastatic or Recurrent) / Benign
1	06-00577	ND	ND	Carotid Body	Positive	Female	80	Benign
2	10-06922	ND	ND	Glomus	Positive	Male	77	Benign
3	11-11668	ND	ND	Carotid Body	Positive	Female	75	Benign
4	12-10660	ND	ND	Carotid Body	Positive	Female	73	Benign
5	12-13613	ND	ND	Carotid Body	Positive	Female	52	Malignant (Metastatic)

ND : Not determined

Table 4.1 continued

Controls								
No	Name	Derived from	Protocol from:					
1	NCC	H9 ES cells	Bajpai et al, (2010)					
2	NCC	H9 ES cells						
Letouze SDH Deficient tumors								
No	Name	Genotype	Germline / Somatic Status	Location of Tumor	PGL/Pheo	Gender	Age	Malignant / Benign
1	HS_144	SDHB	Germline	Left Adrenal Gland	PGL	Male	28	Malignant
2	HS_082	SDHB	Germline	Abdominal PGL	PGL	Male	44	Malignant
3	HS_083	SDHB	Germline	Abdominal	PGL	Female	24	Malignant

Table 4.1 continued

No	Name	Genotype	Germline / Somatic Status	Location of Tumor	PGL/Pheo	Gender	Age	Malignant / Benign
4	HS_085	SDHB	Germline	abdominal node	PGL	Female	38	Malignant
5	HS_088	SDHD	Germline	Abdominal PGL	PGL	Female	31	Benign
6	HS_149	SDHB	Germline	Right Adrenal Gland	PGL	Male	59	Malignant
7	HS_191	SDHB	Germline	Abdominal PGL	PGL	Female	33	Benign
8	HS_186	SDHA	Germline	Left Adrenal Gland	PGL	Female	32	Benign
9	HS_201	SDHB	Germline	Abdominal PGL	PGL	Female	47	Malignant
10	HS_207	SDHC	Germline	Thoracic PGL	PGL	Female	16	Benign

Table 4.1 continued

Letouze SDH Present tumors								
No	Name	Genotype	Germline / Somatic Status	Location of Tumor	PGL/Pheo	Gender	Age	Malignant / Benign
1	HS_021	FH	Germline	Right adrenal gland	Pheo	Female	68	Benign
2	HS_033	NF1	Somatic	Left adrenal gland	Pheo	Female	65	Benign
3	HS_045	NF1	Somatic	Left adrenal gland	Pheo	Female	51	Benign
4	HS_060	RET	Somatic	Left adrenal gland	Pheo	Male	59	Benign
5	HS_111	VHL	Germline	Left adrenal gland	Pheo	Female	10	Benign
6	HS_129			Left adrenal	Pheo	Female	36	Benign

Table 4.1 continued

No	Name	Genotype	Germline / Somatic Status	Location of Tumor	PGL/Pheo	Gender	Age	Malignant / Benign
7	HS_164			Right adrenal gland	Pheo	Male	60	Benign
8	HS_165			Right adrenal gland	Pheo	Male	47	Benign
9	HS_168			Abdominal PGL	PGL	Female	47	Benign
10	HS_178			Left adrenal gland	Pheo	Female	37	Benign
11	HS_205	VHL	Germline	Right adrenal gland	Pheo	Male	27	Benign
12	HS_132	MAX	Somatic	Right adrenal gland	Pheo	Male	24	Benign

Table 4.2

Short List of Genes that Gain DNA Methylation in their CpG Island Promoters and Lose Gene Expression in SDHx Paragangliomas Compared to Control (cluster 3a), with a Characterized Role in Cancer.

Gene	Name	Chr	Role in Cancer / Function	Reference
DRD4	Dopamine Receptor D4	11	Epigenetically silenced in Medulloblastomas	Unland, *et al.,* 2012
KRT19	Keratin 19	17	Silencing of KRT19 results in increased cell proliferation, migration, invasion and survival in breast cancer cells	Ju, *et al.,* 2013
FRZB	Frizzled-Related Protein	2	Epigenetic silencing of tumor suppressor in non-small-cell lung cancer[1], ovarian cancer[2], colorectal cancer cells[3], hepatoma cells[4], breast cancer[5]	[1]Fukui, *et al.,* 2005, [2]Takada, *et al.,*2004, [3]Suzuki, *et al.,*2004, [4]Shih, *et al.,*2007, [5]Lo, *et al.,*2006
NSUN7	NOP2 / Sun Domain Family, Member 7	4	RNA methyltrasferase, mutations associated with infertility or sperm motility defects in male mice	Harris, *et al.,*2007

Table 4.3

Short List of Genes that Gain DNA Methylation in their CpG Island Promoters and Lose Gene Expression in SDHx and SDH Present Paragangliomas Compared to Control (cluster 3b), with a Characterized Role in Cancer.

Gene	Name	Chr	Role in Cancer	Reference
DNMT3A	DNA (Cytosine-5-)-Methyltransferase 3 Alpha	2	[1]Deletion of DNMT3a promotes lung tumor progression (but not initiation) [2]Mutations in DNMT3a found in AML and associated with poor patient prognosis	[1]Gao *et al.,* 2011 [2]Kandoth *et al.,* 2013
GRHL2	Grainyhead-Like Protein 2 Homolog	8	Downregulation in breast cancer shown to be crucial for EMT	Werner *et al.,* 2013
KAZALD1	Kazal-Type Serine Peptidase Inhibitor Domain 1	10	Epigenetic silencing is associated with malignant transformation / progression in gliomas.	Wang *et al.,* 2013
NSD1	Nuclear Receptor Binding SET Domain Protein 1	5	[1]Inactivation by DNA methylation in neuroblastoma and glioma. Hypermethylation associated with poor patient outcome. [2]Mutations in NSD1 found in HNSC	[1]Berdasco *et al.,* 2009 [2]Kandoth *et al.,* 2013

Table 4.3 continued

Gene	Name	Chr	Role in Cancer	Reference
ATP5G2	ATP synthase, H+ transporting, Mitochondrial Complex, Subunit C2	12	Epigenetically silenced in Renal Cell Carcinomas	Morris, *et al.,* 2011
TOX3	TOX high mobility group box family member 3	16	Epigenetically silenced in lung and breast tumors	Tessema, *et al.,* 2012

Table 4.4

Short List of Genes that Gain DNA Methylation in their CpG Shore Promoters and Lose Gene Expression in SDHx and SDH Present Paragangliomas Compared to Control, with a Characterized Role in Cancer.

Gene	Name	Chr	Role in Cancer	Reference
FAM60A	Family with Sequence Similarity 60, Member A	12	Loss of FAM60A leads to changes in cell morphology and increase in cell migration in 293T cells by changing expression of genes in TGFB signaling pathyway	Smith *et al*., 2012
FGFR2	Fibroblast Growth Factor Receptor	10	Acts as a tumor suppressor in gastric tumors, salivary adenocarcinomas, prostate cancer, astrocyte-derived tumors, bladder tumors, associating with poor prognosis. Epigenetically silenced in gastric cancer cell lines	Park *et al.,* 2007
KRT8	Keratin 8	12	Downregulation of KRT8 in epithelial cancer cells increases cell migration	Fortier *et al.,* 2013

Table 4.4 continued

Gene	Name	Chr	Role in Cancer	Reference
PAX6	Paired Box 6	11	PAX6 acts as a tumor suppressor in glioblastoma cells[1] and prostate cancer cells[2] Pax6 also suppresses invasiveness in glioblastoma cells[3]	[1]Zhou *et al.,* 2005, [2]Shyr *et al.,* 2010, [3] Mayer et al., 2006
SFRP2	Secreted Frizzled-Related Protein 2	4	Epigenetically silenced in prostate carcinomas[1], human breast cancer[2], esophageal squamous cell carcinoma[3], advanced colorectal tumors[4], hepatocellular carcinoma[5] and [6]advanced gastric cancer	[1]Kilinc, *et al.*, 2012, [2]Veeck, *et al.*, 2008, [3]Hao, *et al.,* 2012 [4]Takeda, *et al.,* 2011, [5]Shih, *et al.,* 2013, [6]Hiraki, *et al.,* 2011
SOX9	SRY (Sex Determining Region Y) – Box 9	17	Epigenetically silenced in high grade bladder cancer and is associated with poor patient prognosis	Aleman, *et al.*, 2008

Table 4.4 continued

Gene	Name	Chr	Role in Cancer	Reference
TRIP6	Thyroid Hormone Receptor Interacting Protein	7	Downregulation of TRIP6 in lung cancer cells induces actin cytoskeletal arrangements leading to increased cell motility and hence malignant phenotype	Guryanova, *et al*., 2005

Table 4.5

Short List of Genes, that Gain DNA Methylation in their Distant Region Promoters and Lose Gene Expression in SDHx and SDH Present Paragangliomas Compared to Control with a Characterized Role in Cancer.

Gene	Name	Chr	Role in Cancer / Function	Reference
FABP3	Fatty Acid Binding Protein 3, Muscle And Heart (Mammary-Derived Growth Inhibitor)	1	[1]Tumor suppressor in breast cancer [2]Overexpression promotes apoptosis in embryonic cancer cells	[1]Huynh *et al.,* 1995 [2]Song *et al*., 2012
SOX5	SRY (Sex Determining Region Y)-Box 5	12	Overexpression in human gliomas cells led to reduction in proliferation and clone formation	Tchougounova *et al.,* 2009
SOX6	SRY (Sex Determining Region Y)-Box 6	11	Tumor suppressor in esophageal squamous cell cancer	Qin *et al.,* 2010

Table 4.6

Short List of Genes, that Lose DNA Methylation in their CpG Island Promoters and Gain Gene Expression in SDHx and SDH Present Paragangliomas Compared to Control (clusters 1, 2, 4 and 5), with a Characterized Role in Cancer.

Gene	Name	Chr	Role in Cancer	Reference
ACP5	Acid Phosphatase 5, Tartrate Resistant	19	Overexpression observed in early stage melanomas and acts as a pro-invasion metastasis driver	Scott *et al.,* 2011
ACVRL1/ ALK1	Activin A Receptor Type II-Like 1	12	In melanoma model, ACVRL1 may play a role in stabilizing angiogenic vessels and contribute to resistance to anti-VEGF therapies. Overexpression observed in many human tumor types and in circulating ECs from patients with advanced cancers	Hu-Lowe *et al.,* 2011
CD14	Myeloid Cell-Specific Leucine-Rich Glycoprotein	5	Involved in chronic inflammation. Elevated levels observed in cholangiocarcinoma	Subimerb *et al.,* 2010

Table 4.6 continued

Gene	Name	Chr	Role in Cancer	Reference
CFLAR	CASP8 And FADD-Like Apoptosis Regulator	2	Apoptosis regulator protein that may function as a crucial link between cell survival and cell death pathways in cells. Acts as an inhbitor of TNFRSF6 mediated apoptosis	Fulda, 2013
CHL1	Cell Adhesion Molecule L1-Like	3	Re-expression of the gene on the edge of tumor mass might promote local invasive growth and enable further metastatic spread in ovary, colon and breast cancer. Potential novel specific biomarker in the early pathogenesis of two major histological types of renal cancer.	Senchenko *et al.,* 2011
DGKA	Diacylglycerol Kinase, Alpha	12	Implicated in VEGF mediated angiogenesis[1] and suppression of TNF-alpha induced apoptosis of human melanoma cells via NF-KB.[2] This kinase is found to be constitutively activated in nucleophosmin/anaplastic lymphoma kinase (NPM/ALK) fusion in malignant lymphomas, where inhibition of DGKA significantly reduced tumor growth[3] Found mutated in pancreatic cancers[4]	[1]Baldanzi *et al*., 2004 [2]Yanagisawa *et al*., 2007 [3]Bacchiocchi *et al*., 2005 [4]Carter *et al.,* 2010

Table 4.6 continued

Gene	Name	Chr	Role in Cancer	Reference
FSTL3	Follistatin-Like 3 (Secreted Glycoprotein)	19	FSTL3 increased angiogenesis but not tumor growth in renal carcinoma tumors	Krneta *et al.,* 2006
GNA14	Guanine Nucleotide Binding Protein (G Protein), Alpha 14	9	Activation of TNF-α/TNFR1 signaling in the tumor microenvironment promotes gastric tumor development through induction of Noxo1 and Gna14, which contribute to maintaining the tumor cells in an undifferentiated state	Oshima *et al.*, 2013
LY6K	Lymphocyte Antigen 6 Complex, Locus K	8	LY6K is a cancer biomarker and a therapeutic target that induces invasion and metastasis in breast cancer	Kong *et al.*, 2012
PMEPA1	Prostate Transmembrane Protein, Androgen Induced 1	20	Involved in cell proliferation and overexpressed in many cancers such as in AR receptor negative prostate cancers and renal cell carcinomas	Liu *et al.,* 2011

Table 4.6 continued

Gene	Name	Chr	Role in Cancer	Reference
REC8	REC8 Meiotic Recombination Protein	14	REC8 is overexpressed in melanoma and may act as a driver of genomic instability	Lindsey *et al.*, 2013
SRPK3	SRSF Protein Kinase 3	X	Up-regulated in breast and colonic tumors where its expression increases coordinately with tumor grade. Inhibition of SRPK3 both increased apoptotic potential and enhanced cell killing after treatment with gemcitabine and cisplatin	Hayes *et al.*, 2007
TACSTD 2	Tumor-Associated Calcium Signal Transducer 2	1	Upregulation of this gene stimulates cancer growth; significantly upregulated TACSTD2 and Cyclin D1 correlate with poor prognosis of invasive ductal breast cancer	Trerotola *et al.*, 2013

Table 4.7

Candidate List Filtered from Whole Exome Sequencing of Nonsynomous Mutations in Epigenetic Enzymes in Paragangliomas

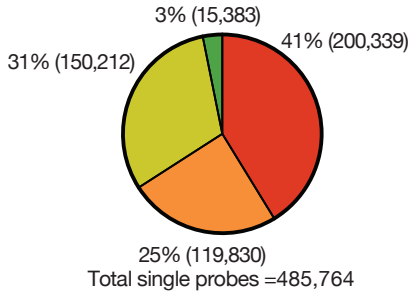
Gene	Tumor Identifier	Genome Position	cDNA change	SNV type	Protein change
YY1	11-09937	chr14:100705787	c.207_209del	Non-frameshift deletion	p.69_70del
KDM6B	11-09937	chr17:7751858	c.159_164del	Non-frameshift deletion	p.53_55del
	10-06992	chr17:7751858	c.159_164del	Non-frameshift deletion	p.53_55del
	11-09937	chr17:7750177	c.753_758del	Non-frameshift deletion	p.251_253del
	10-06992	chr17:7750177	c.753_758del	Non-frameshift deletion	p.251_253del
	12-13613	chr17:7750177	c.753_758ins	Non-frameshift insertion	p.251_253ins
	06-00577	chr17:7751162	c.T1556C	Nonsynomous SNV	p.L519P
	10-06992	chr17:7750532	c.G1019C	Nonsynomous SNV	p.R340P
	12-10660	chr17:7752292	c.A592C	Nonsynomous SNV	p.T198P
DNMT1	11-09937	chr19:10288010	c.A479G	Nonsynomous SNV	p.K160R
SETD1B	06-00577	chr12:122247909	c.C1058T	Nonsynomous SNV	p.P353L
	06-00577	chr12:122260993	c.T4379C	Nonsynomous SNV	p.L1460P

Table 4.7 continued

Gene	Tumor Identifier	Genome Position	cDNA change	SNV type	Protein change
	11-09937	chr12:122261001	c.G4387C	Nonsynomous SNV	p.A1463P
NSD1	06-00577	chr5:176721945	c.C7576T	Nonsynomous SNV	p.P2526S
KDM5C	10-06992	chrX:53223581	c.G3778T	Nonsynomous SNV	p.A1260S
JARID2	10-06992	chr6:15374477	c.G175C	Nonsynomous SNV	p.V59L
	06-00577	chr6:15496721	c.T1151G	Nonsynomous SNV	p.V384G
PRDM2	10-06992	chr1:14105121	c.831_832insGAA	Nonsynomous SNV	p.D277delinsDE
	12-10060	chr1:14106145	c.C1252T	Nonsynomous SNV	p.P418S
MTHFD1	11-11668	chr14:64855048	c.72delC	frameshift deletion	p.V24fs
	10-06992	chr14:64854996	c.19_20insG	frameshift insertion	p.R7fs
JMJD4	12-10660	chr1:227922874	c.C239G	Nonsynomous SNV	p.S80W
MBD5	12-10660	chr2:149226111	c.G599A	Nonsynomous SNV	p.R200Q
	12-10660	chr2:149241305	c.C220A	Nonsynomous SNV	p.L74I
SETD1A	12-13613	chr16:30976421	c.1359_1361del	Non-frameshift deletion	p.453_454del
YY1AP1	12-13613	chr1:155630245	c.A1363G	Nonsynomous SNV	p.M455V
MLL4	11-09937	chr19:36224005	c.A6546C	Nonsynomous SNV	p.K2182N
	06-00577	chr19:36224005	c.A6546C	Nonsynomous SNV	p.K2182N

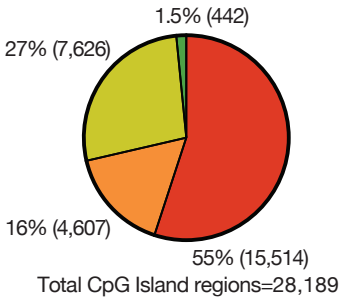
Genomic Location	Single CpGs
Promoter	200,339
Intergenic	119,830
Body	150,212
3'UTR	15,383

Subgroup	CpGs
TSS200	62,525
TSS1500	77,379
5'UTR	49,525
1st Exon	10,810



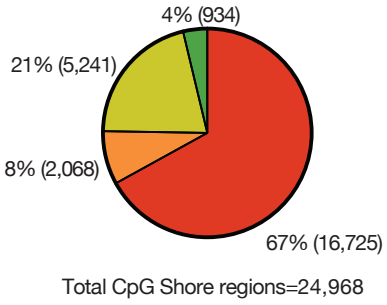
Genomic Location	CpG Island Regions (3 probes or more)
Promoter	15,514
Intergenic	4,607
Body	7,626
3'UTR	442

Subgroup	CpG Island Regions (3 probes or more)
TSS200	4,552
TSS1500	3,235
5'UTR	4,188
1st Exon	3,539



Genomic Location	CpG Shore Regions (3 probes or more)
Promoter	16,725
Intergenic	2,068
Body	5,241
3'UTR	934

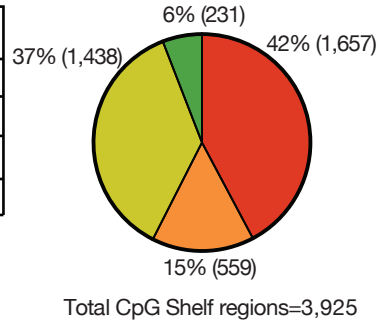
Subgroup	CpG Shore Regions (3 probes or more)
TSS200	4,324
TSS1500	7,860
5'UTR	2,673
1st Exon	1,868



Supplementary Figure 4.1) All regulatory regions are represented in the three probe analysis. Description of 450K DNA methylation array in terms of distribution of probes based on functional genomic locations, promoters (including TSS200, TSS1500, 5'UTR and 1st Exon), Body, 3'UTR and intergenic regions. The first pie-chart shows distribution of single probes on the array. The next four pi-charts show the distribution of the probes in regions, containing atleast three probes or more. These regions are defined in relation to CpG Islands namely, CpG Islands, CpG Shores, CpG Shelves and Distant Regions.

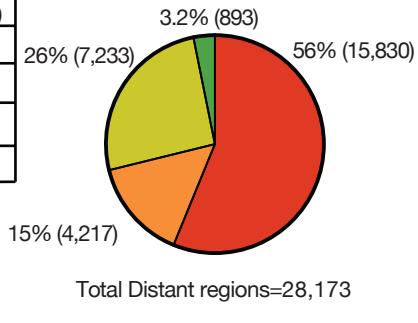
Genomic Location	CpG Shelf Regions (3 probes or more)
Promoter	1,657
Intergenic	559
Body	1,438
3'UTR	231

Subgroup	CpG Shelf Regions (3 probes or more)
TSS200	403
TSS1500	550
5'UTR	401
1st Exon	303

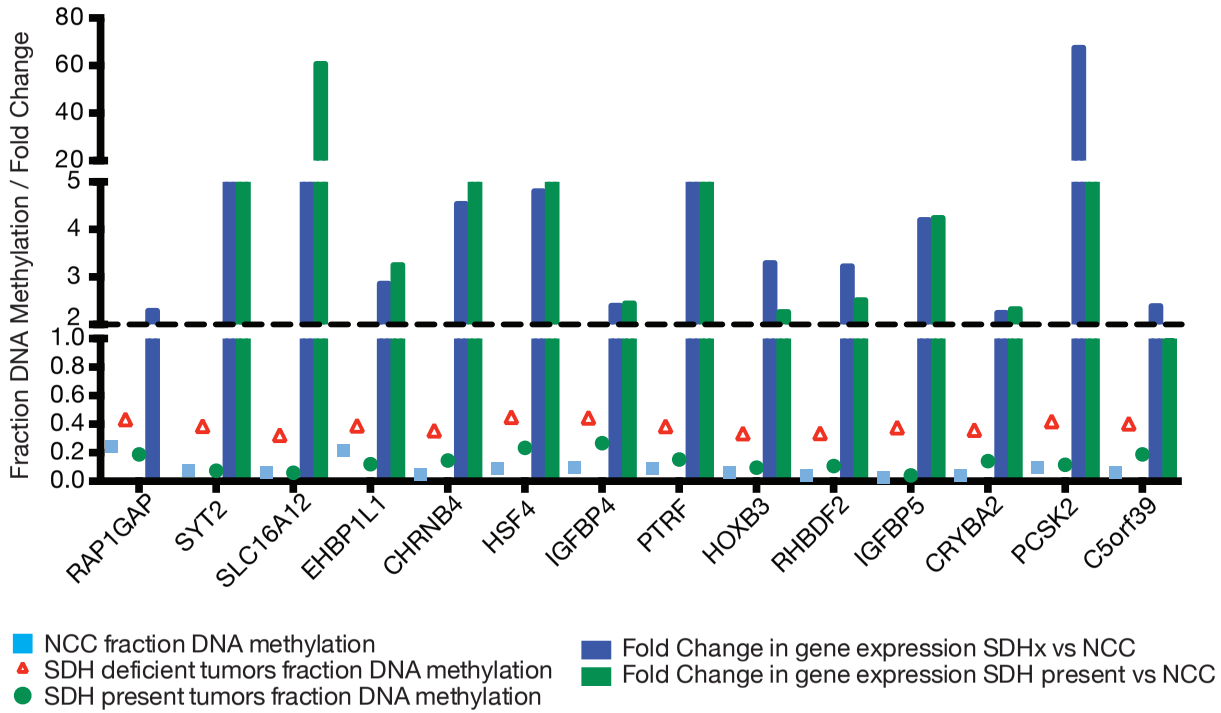


Genomic Location	Distant Regions (3 probes or more)
Promoter	15,830
Intergenic	4,217
Body	7,233
3'UTR	893

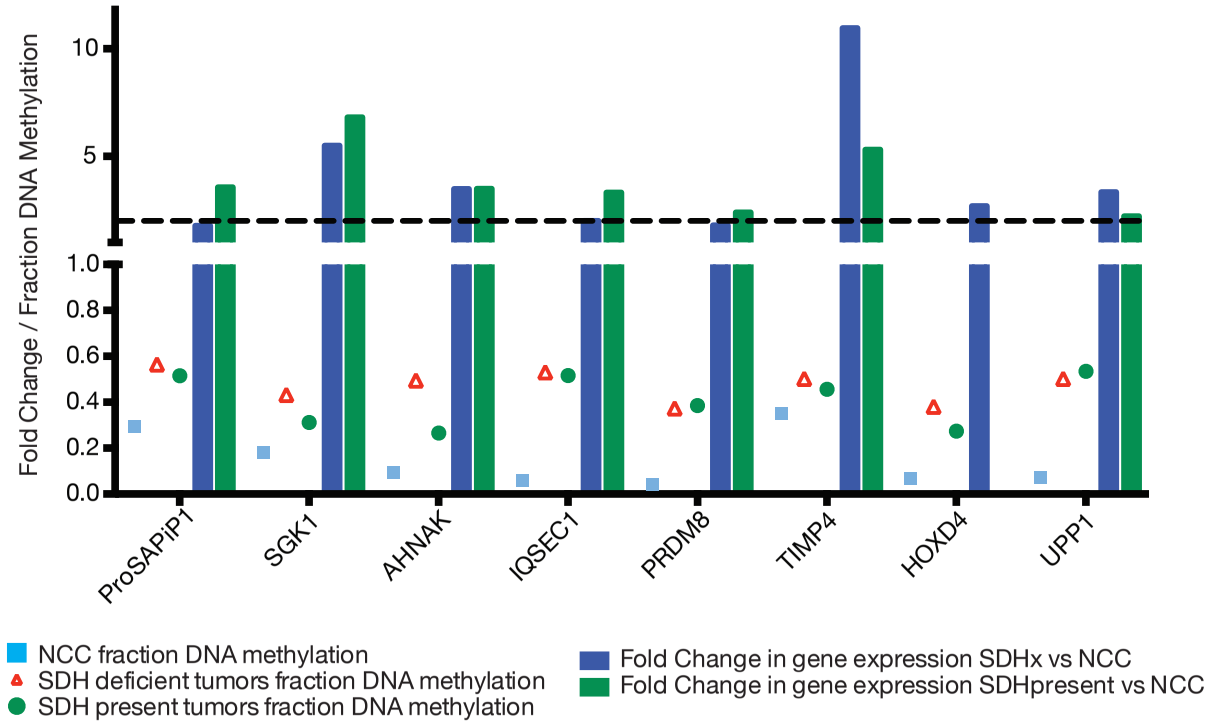
Subgroup	Distant Regions (3 probes or more)
TSS200	4,166
TSS1500	4,958
5'UTR	3,482
1st Exon	3,224



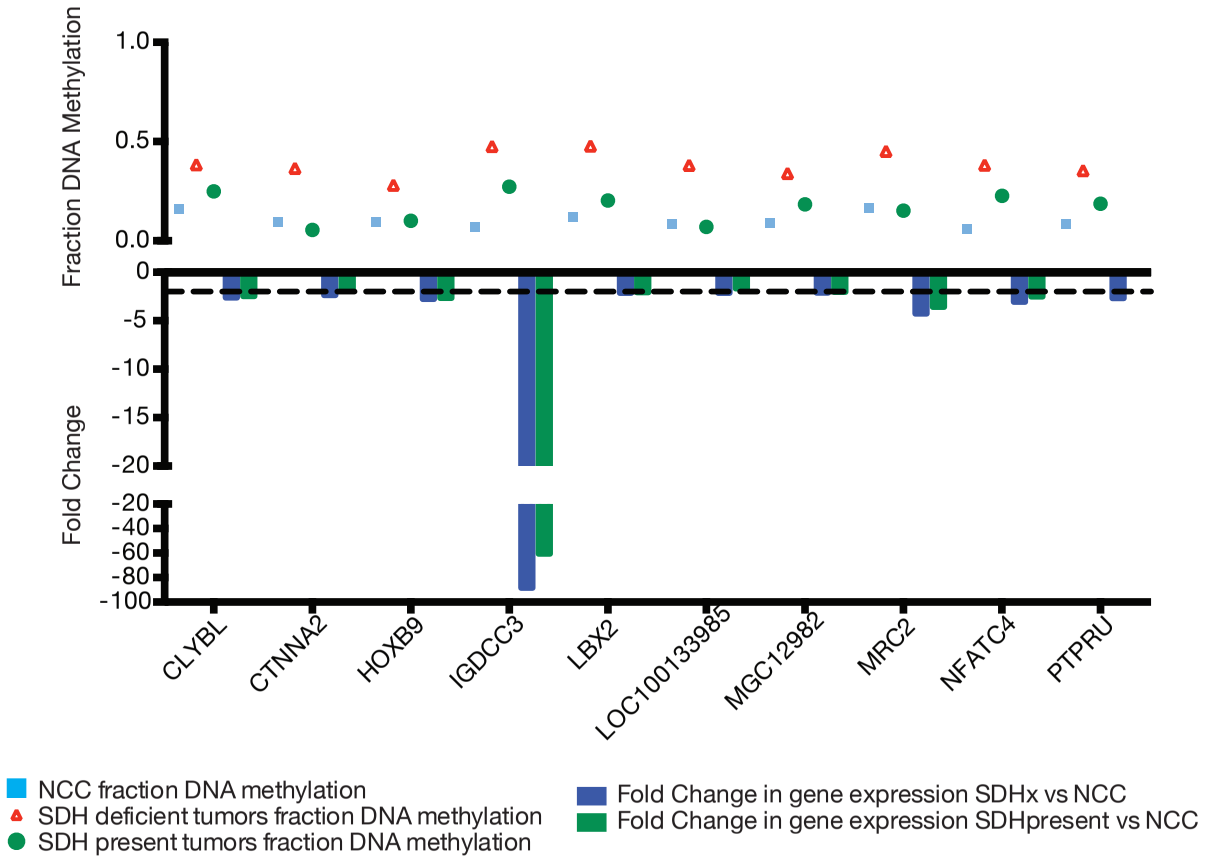
Supplementary Figure 4.1 continued



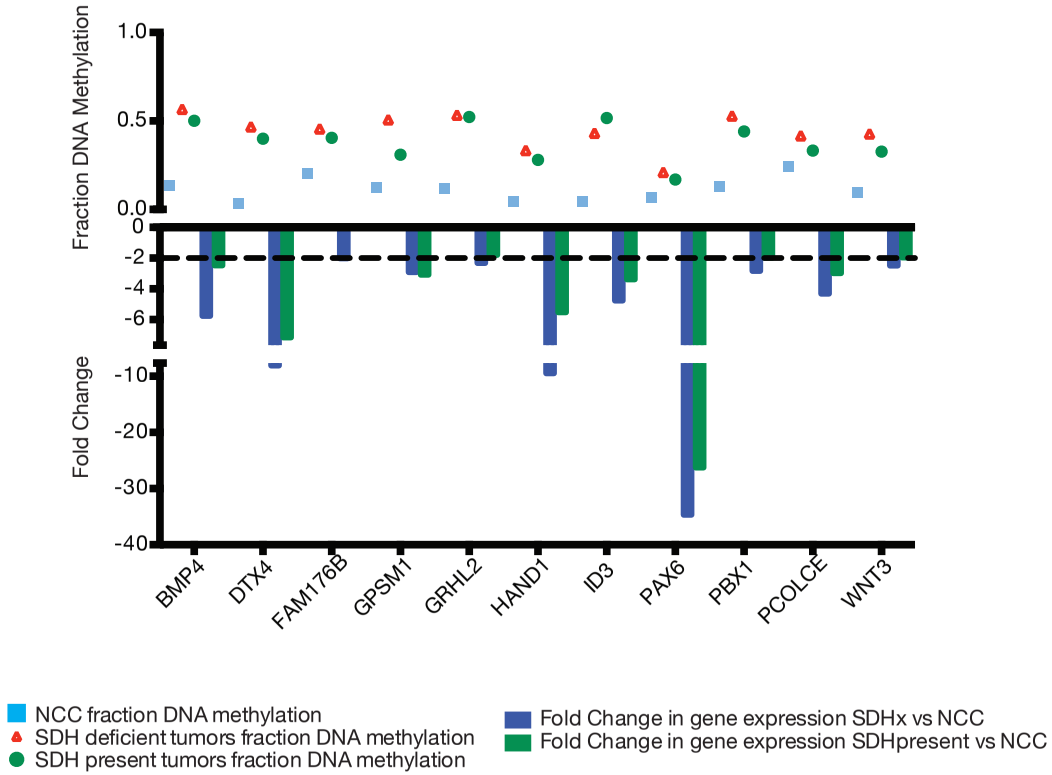
Supplementary Figure 4.2) CpG Islands gaining gene expression with a gain in DNA methylation in their promoters and nonpromoters. Genes from cluster 3a that gain gene expression even though they gain DNA methylation in their CpG Island promoters. SDH Present tumors follow similar trends in changes in gene expression regardless of DNA methylation changes. Transcription changes based on p-value with FDR < 0.05, fold change >= 2. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.



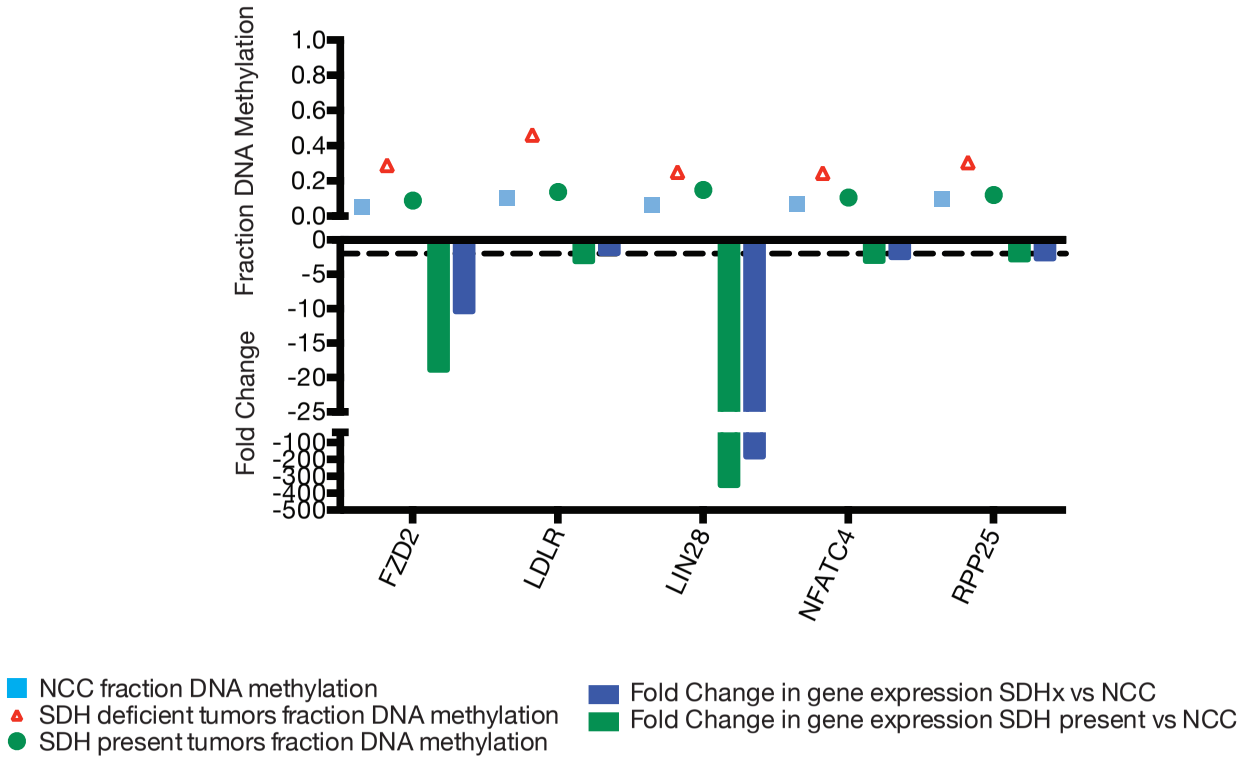
Supplementary Figure 4.3) Genes from cluster 3b that gain gene expression even though they gain DNA methylation in their CpG Island promoters. SDH Present tumors follow similar trends in changes in gene expression regardless of DNA methylation changes. Transcription changes based on p-value with FDR < 0.05, fold change >= 2. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.



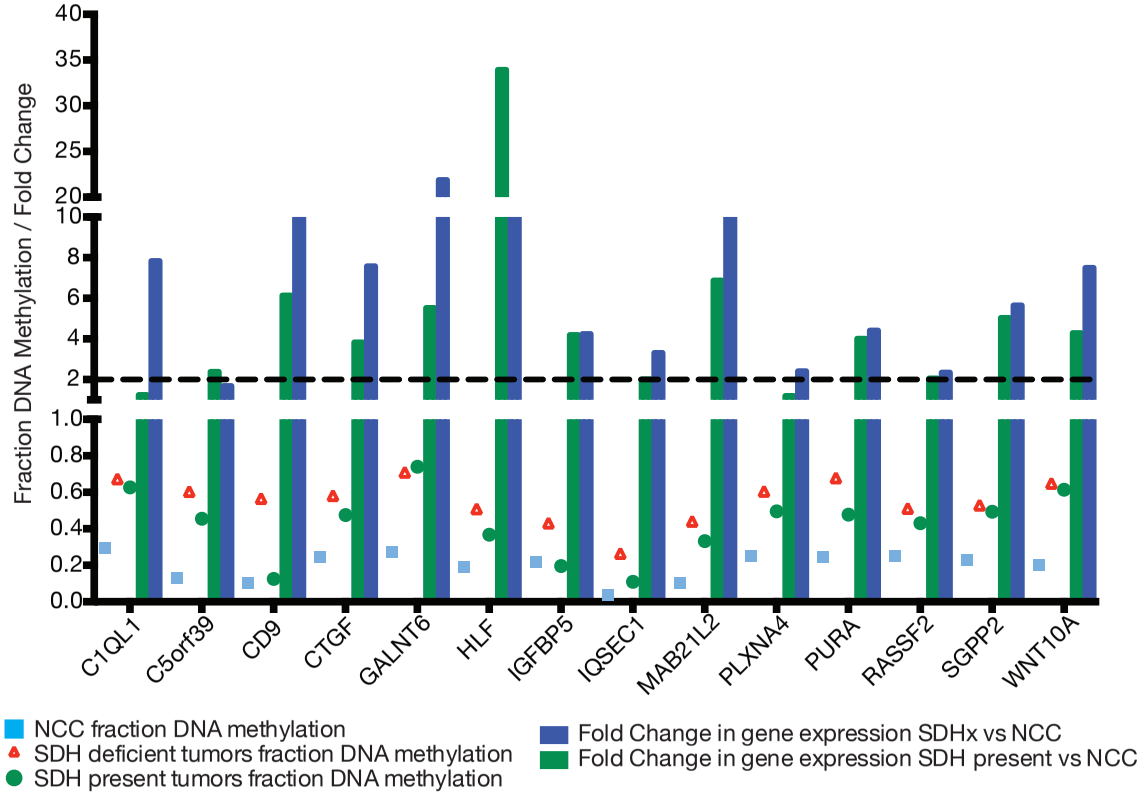
Supplementary Figure 4.4) Genes from cluster 3a that gain gene expression with a gain in DNA methylation in their CpG Island nonpromoters. SDH Present tumors follow similar trends in changes in gene expression regardless of DNA methylation changes. Transcription changes based on p-value with FDR < 0.05, fold change >= 2. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.



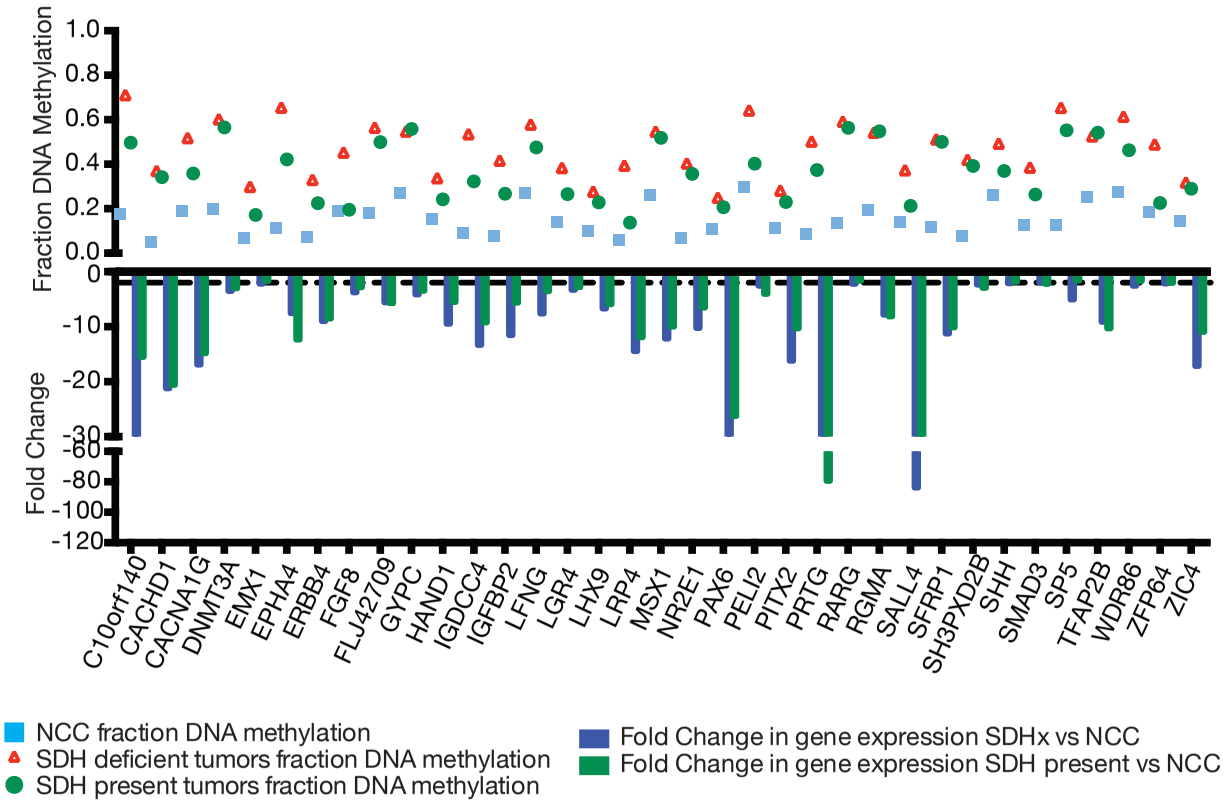
Supplementary Figure 4.5) Genes from cluster 3b (D) that gain gene expression with a gain in DNA methylation in their CpG Island nonpromoters. SDH Present tumors follow similar trends in changes in gene expression regardless of DNA methylation changes. Transcription changes based on p-value with FDR < 0.05, fold change >= 2. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.



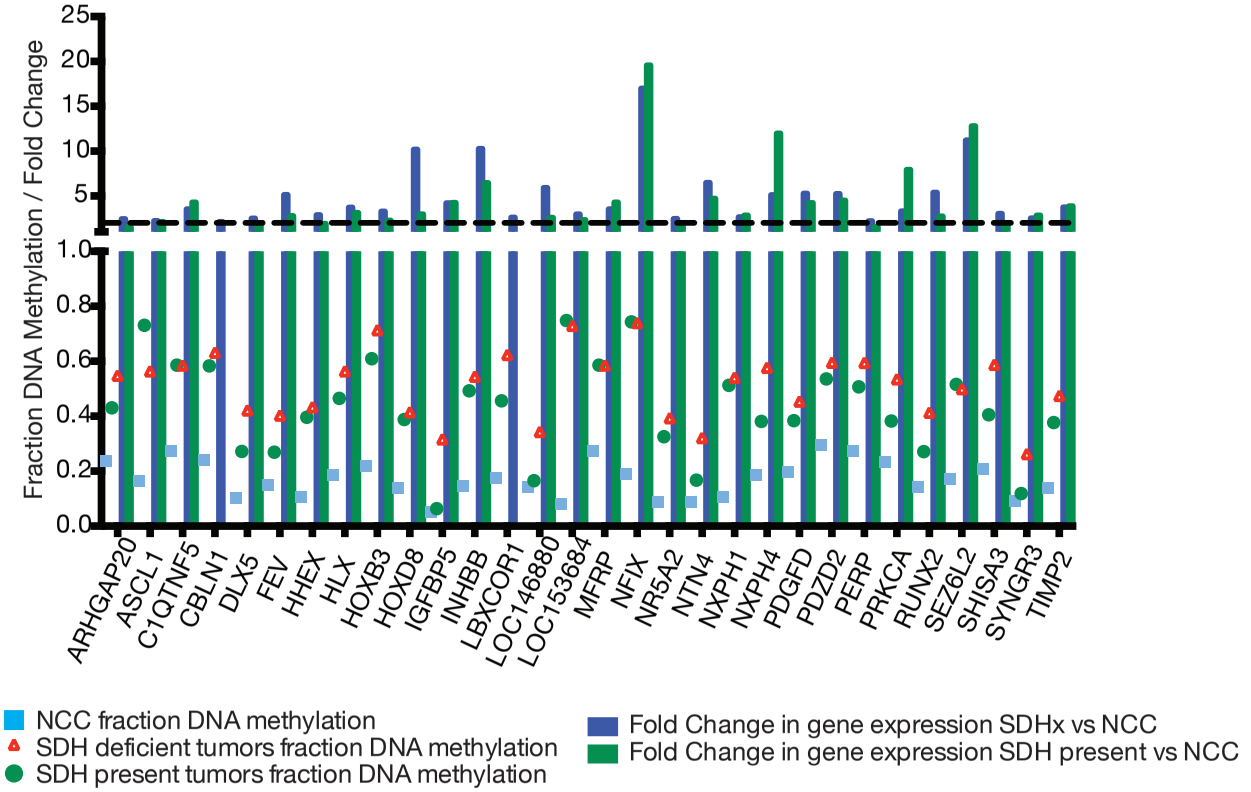
Supplementary Figure 4.6) CpG Shore promoters exclusively gaining DNA methylation in SDH Deficient tumors. Few genes gain DNA methylation in their CpG Shore promoters exclusively in SDHx tumors. Transcription changes based on p-value with FDR < 0.05, fold change >= 2 in at least one of the tumor groups. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.



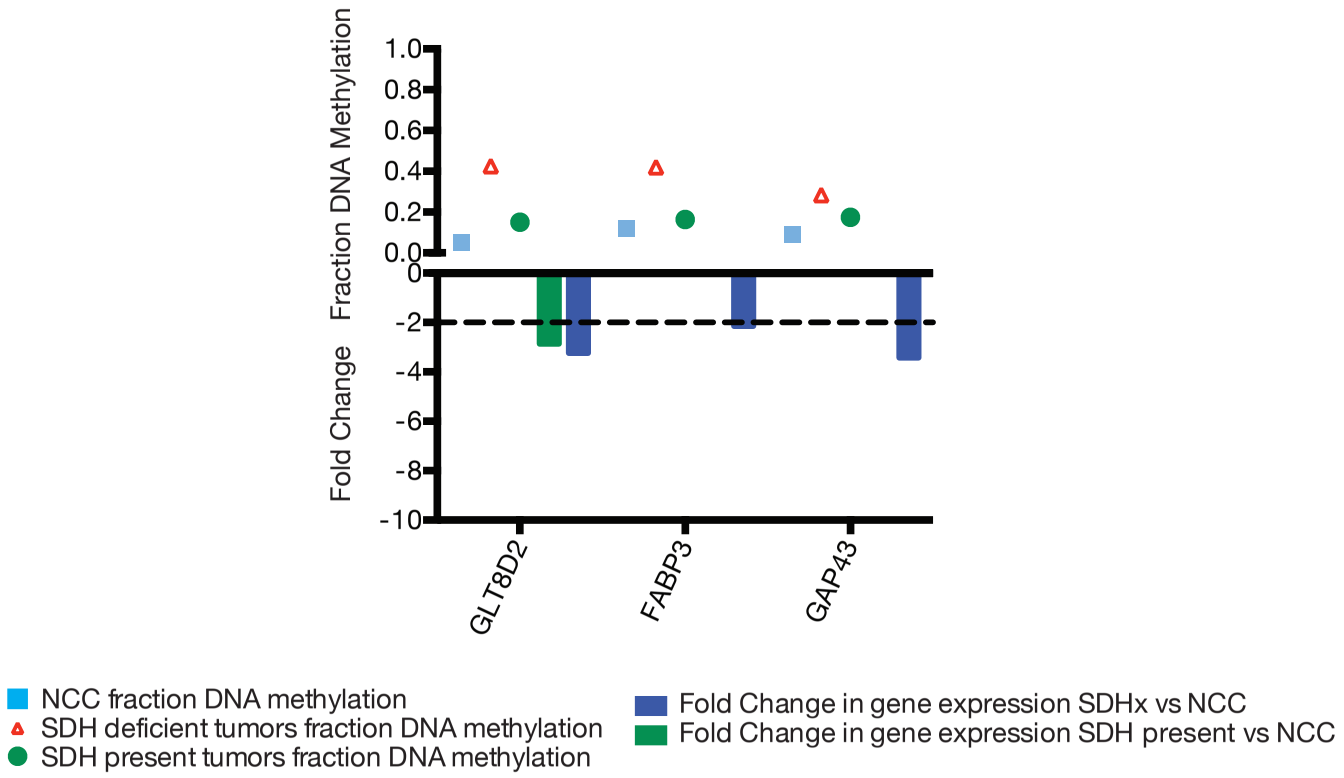
Supplementary Figure 4.7) Genes that gain gene expression even though they gain DNA methylation in their CpG shore promoters. SDH Present tumors follow similar trends in changes in gene expression regardless of DNA methylation changes. Transcription changes based on p-value with FDR < 0.05, fold change >= 2. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.



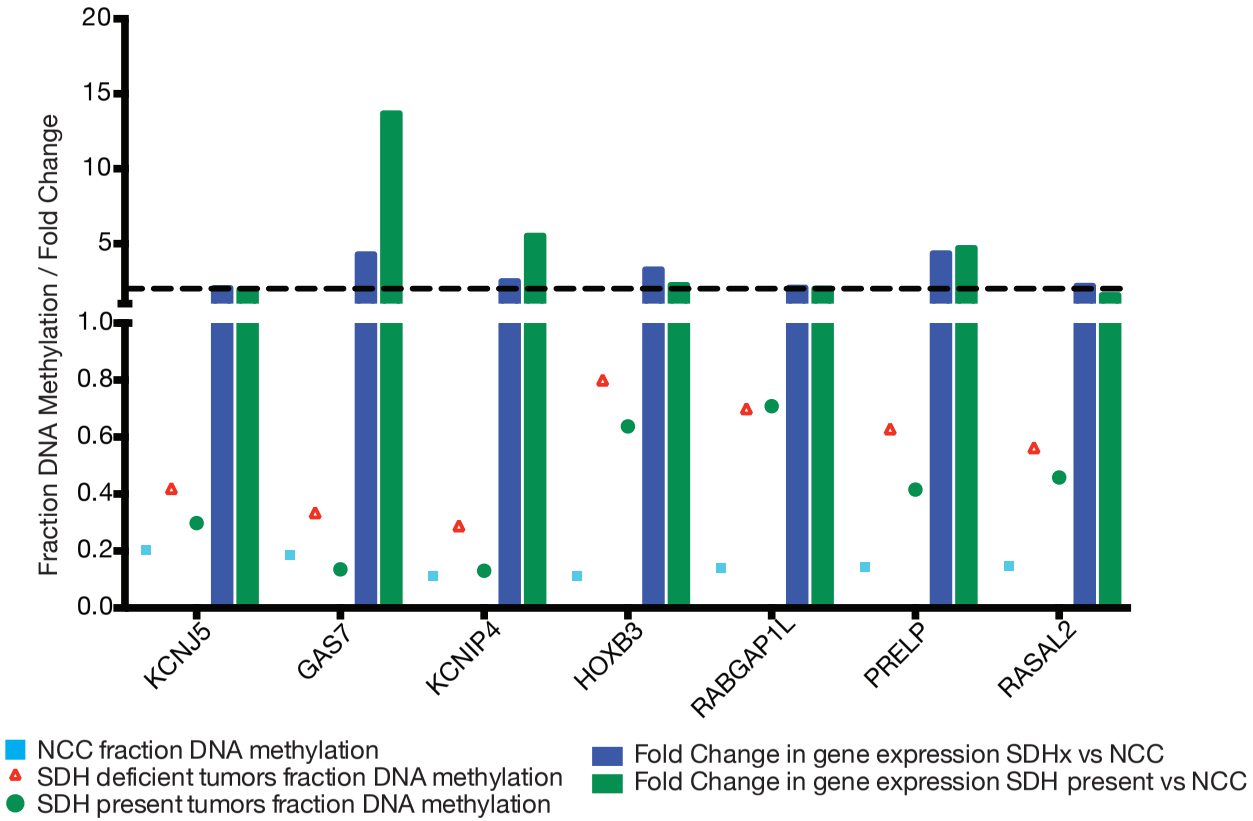
Supplementary Figure 4.8) Fold change expression of genes significantly methylated in their CpG Shore nonpromoters (body, 3'UTR) from clusters 1, 2 and 5 gaining DNA methylation in tumors vs NCCs. Graph shows genes that lose gene expression with gain of DNA methylation in their CpG Shore nonpromoters in SDHx tumors and SDH Present tumors compared to NCCs. Transcription changes based on p-value with FDR < 0.05, fold change >= 2. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.



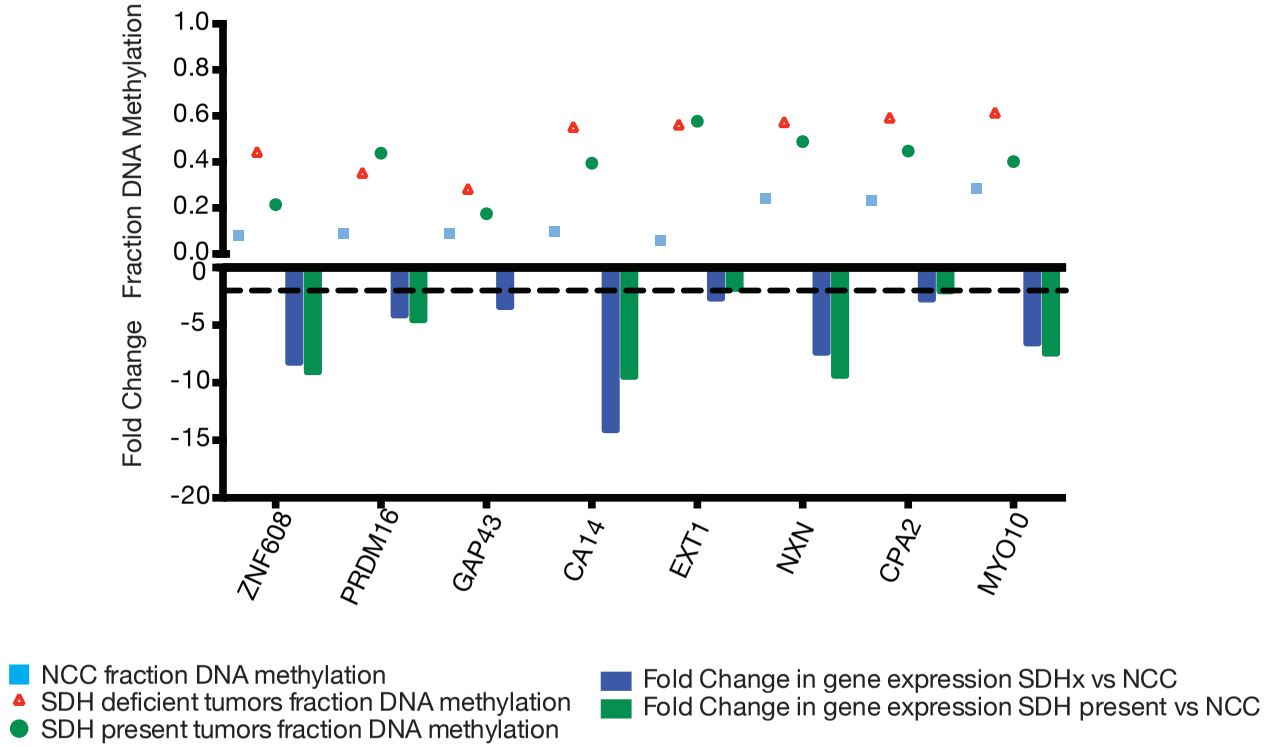
Supplementary Figure 4.9) Fold change expression of genes significantly methylated in their CpG Shore nonpromoters (body, 3'UTR) from clusters 1, 2 and 5 gaining DNA methylation in tumors vs NCCs. Graph shows genes that gain gene expression with a gain DNA methylation in their CpG Shore nonpromoters. SDH Present tumors follow similar trends in changes in gene expression regardless of DNA methylation changes.Transcription changes based on p-value with FDR < 0.05, fold change >= 2. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.



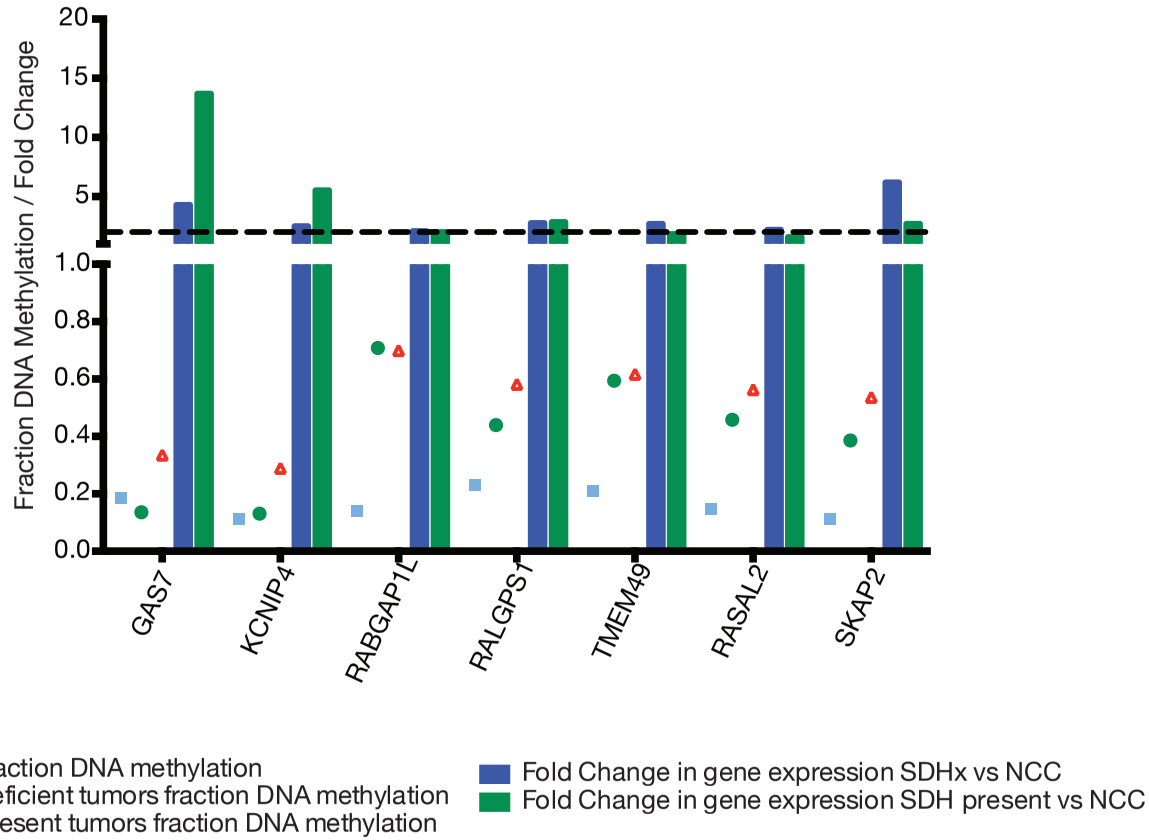
Supplementary Figure 4.10) Distant region methylation dynamics. Few genes gain DNA methylation in their Distant Region promoters exclusively in SDHx tumors. Transcription changes based on p-value with FDR < 0.05, fold change >= 2 in at least one of the tumor groups. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.



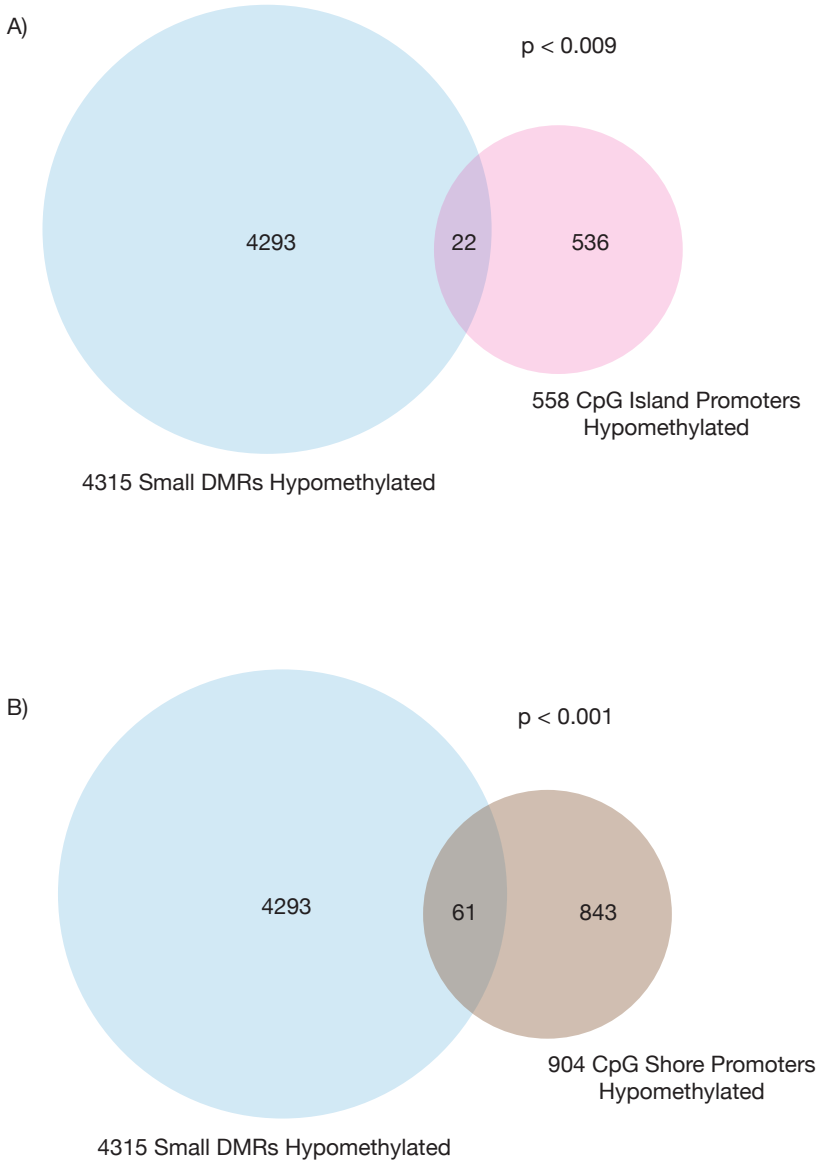
Supplementary Figure 4.11) Fold change expression of genes significantly gaining DNA methylation in the promoters of Distant Regions in tumors vs NCCs (from clusters 2 and 5). Graph shows genes that gain gene expression even though they gain DNA methylation in their promoters. SDH Present tumors follow similar trends in changes in gene expression regardless of DNA methylation changes. Transcription changes based on p-value with FDR < 0.05, fold change >= 2. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.



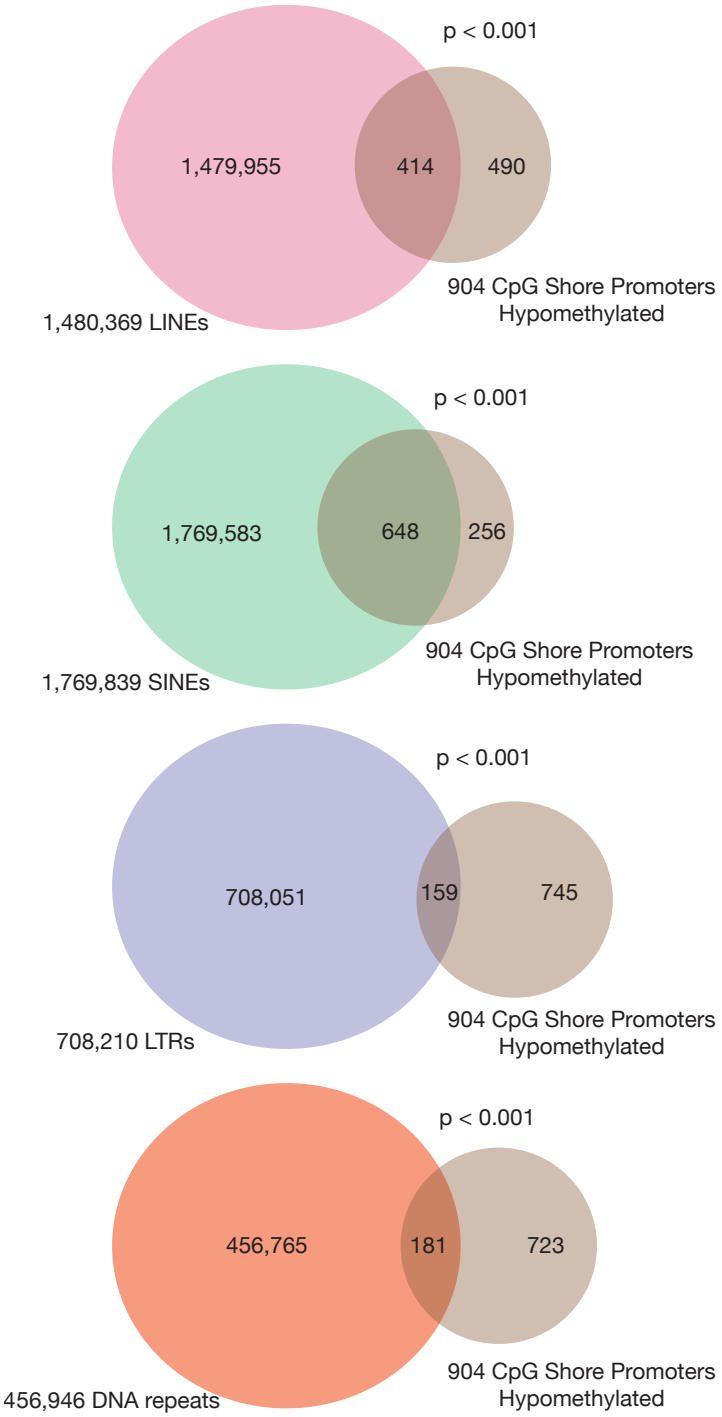
Supplementary Figure 4.12) Fold change expression of genes significantly gaining DNA methylation in their non-promoters (body, 3'UTR) in tumors vs NCCs (from clusters 2 and 5) that lose gene expression. Graph shows genes that lose gene expression with gain of DNA methylation in their nonpromoters in SDHx tumors and SDH Present tumors compared to NCCs. Transcription changes based on p-value with FDR < 0.05, fold change >= 2. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.



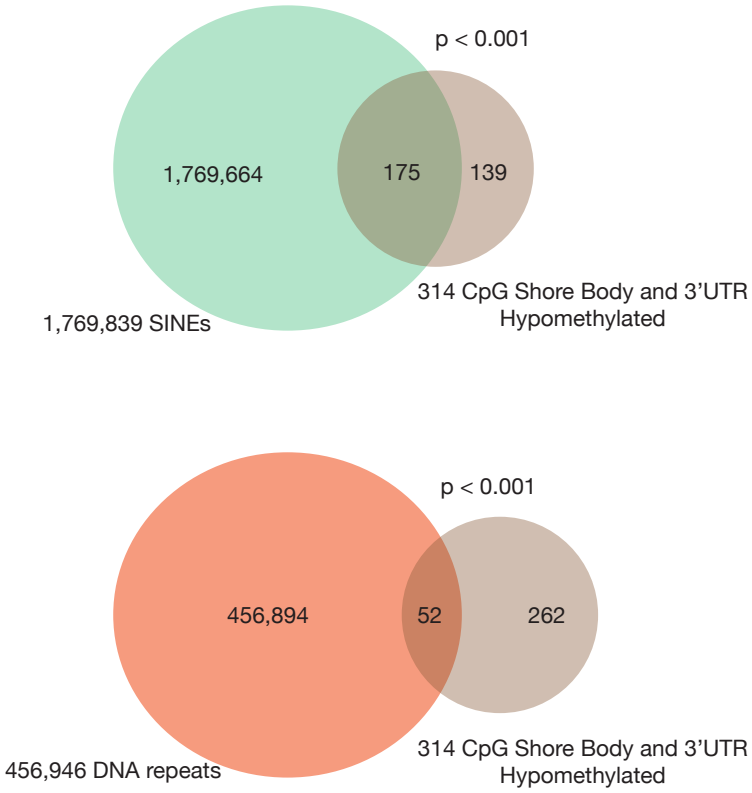
Supplementary Figure 4.13) Fold change expression of genes significantly gaining DNA methylation in their nonpromoters (body, 3'UTR) in tumors vs NCCs (from clusters 2 and 5) that gain gene expressoin. Graph shows genes that gain gene expression with a gain in DNA methylation in their nonpromoters. SDH Present tumors follow similar trends in changes in gene expression regardless of DNA methylation changes. Transcription changes based on p-value with FDR < 0.05, fold change >= 2. Dotted line represents 2 fold change. DNA methylation changes based on p-value with FDR < 0.05, change in fraction methylation >=15%.



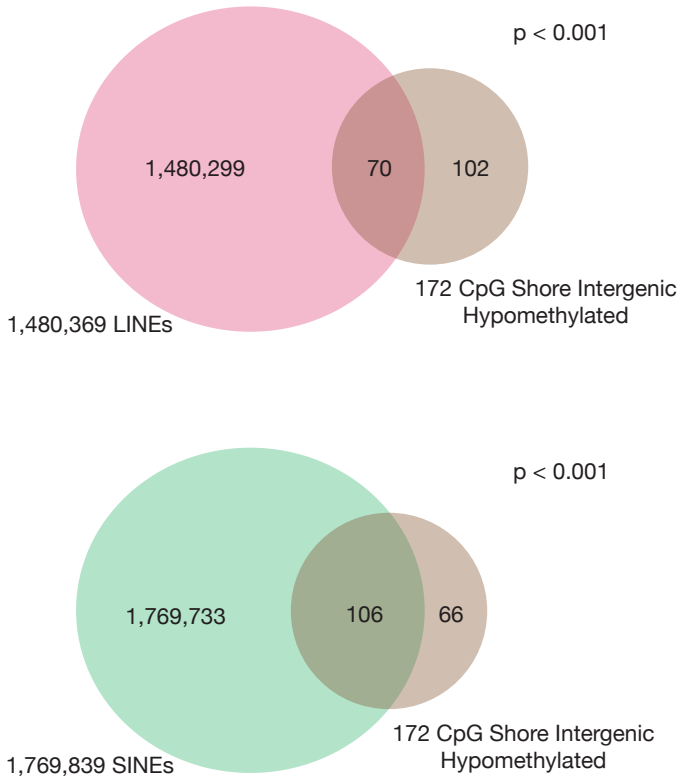
Supplementary Figure 4.14: Overlap of hypomethylated CpG Shore and Islands with hypomethylated CpG Shore DMRs from colon cancer.
A) diagram shows overlap of small DMRs that lose DNA methylation from Colon Cancers as published by Hansen *et al.*, (2011) and hypomethylated CpG Island promoters from SDH Deficient and SDH Present PGLs. Overlap of 22 regions was significant with a p-value ≤ 0.009. B) diagram shows overlap of hypomethylated small DMRs from colon cancer and hypomethylated CpG Shore promoters from SDH Deficient and SDH Present PGLs. Overlap of 61 regions was significant with a p-value < 0.001.



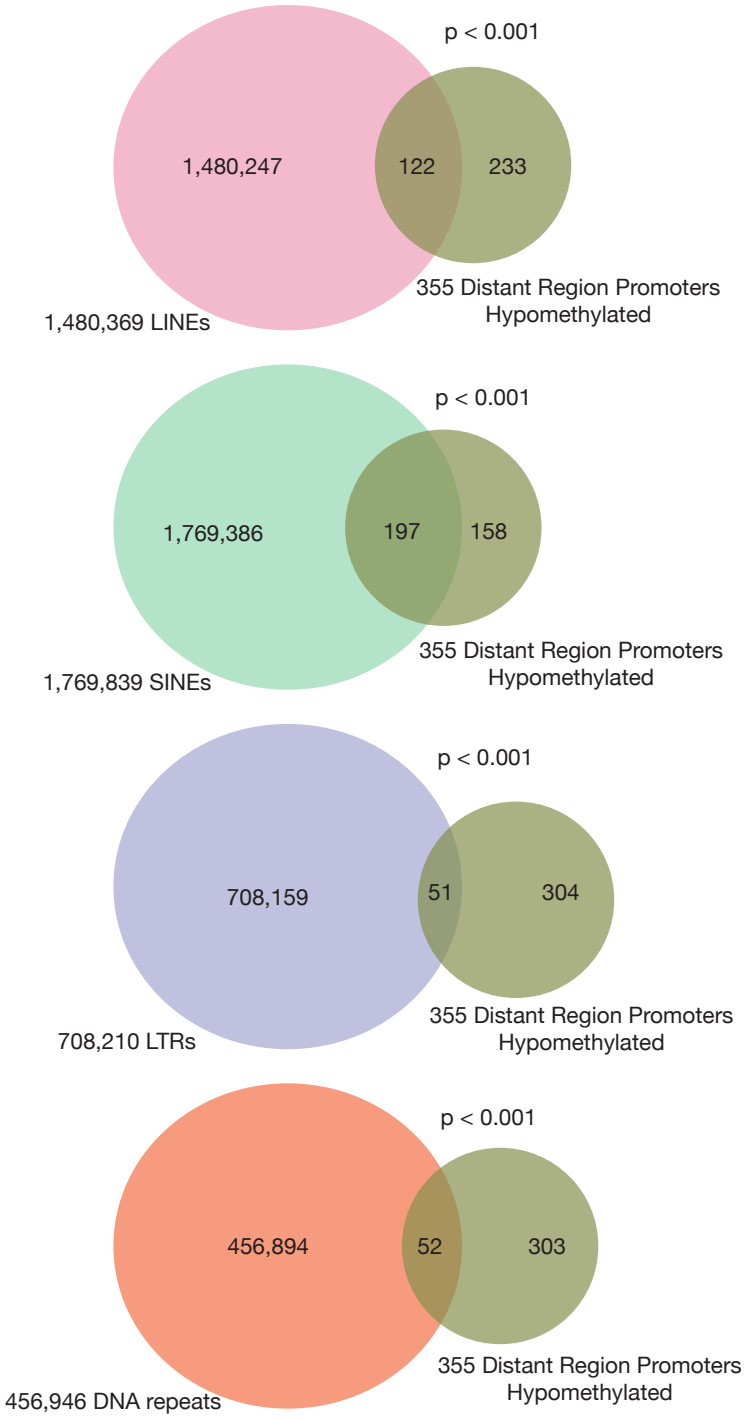
Supplementary Figure 4.15: Hypomethylated CpG Shore in promoters overlap with several repeat elements.



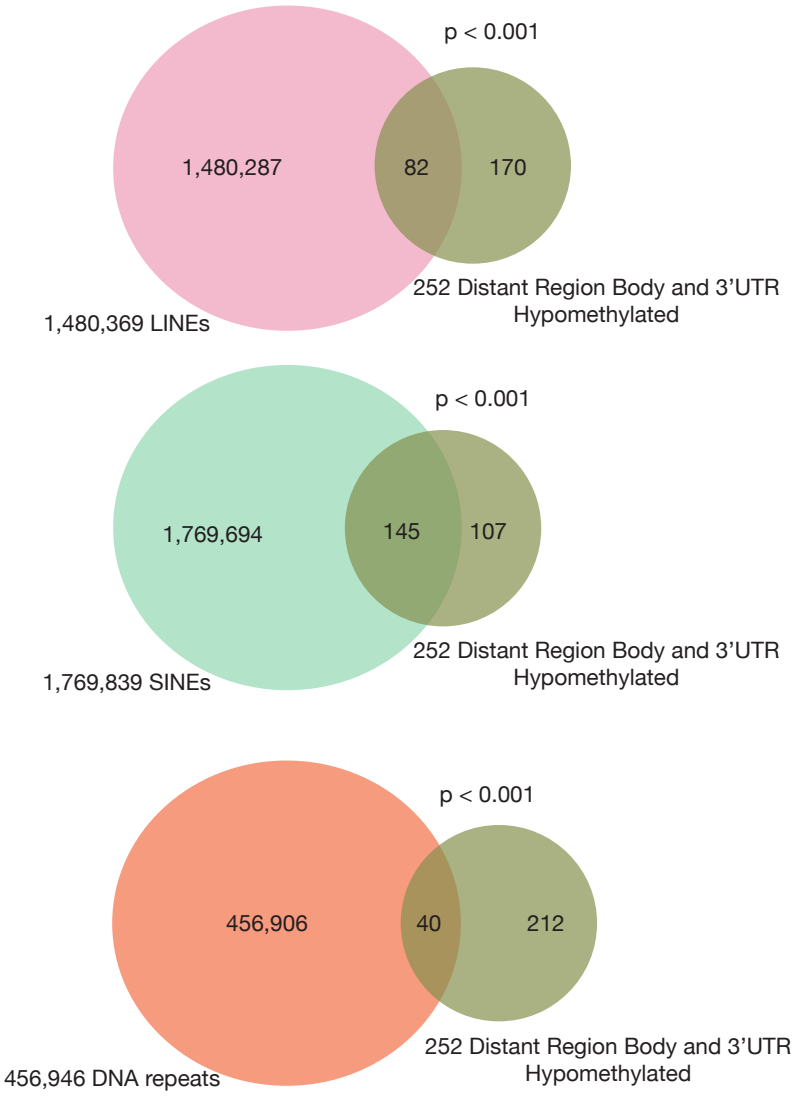
Supplementary Figure 4.16: Hypomethylated CpG Shore in body and 3'UTR overlap with several repeat elements.



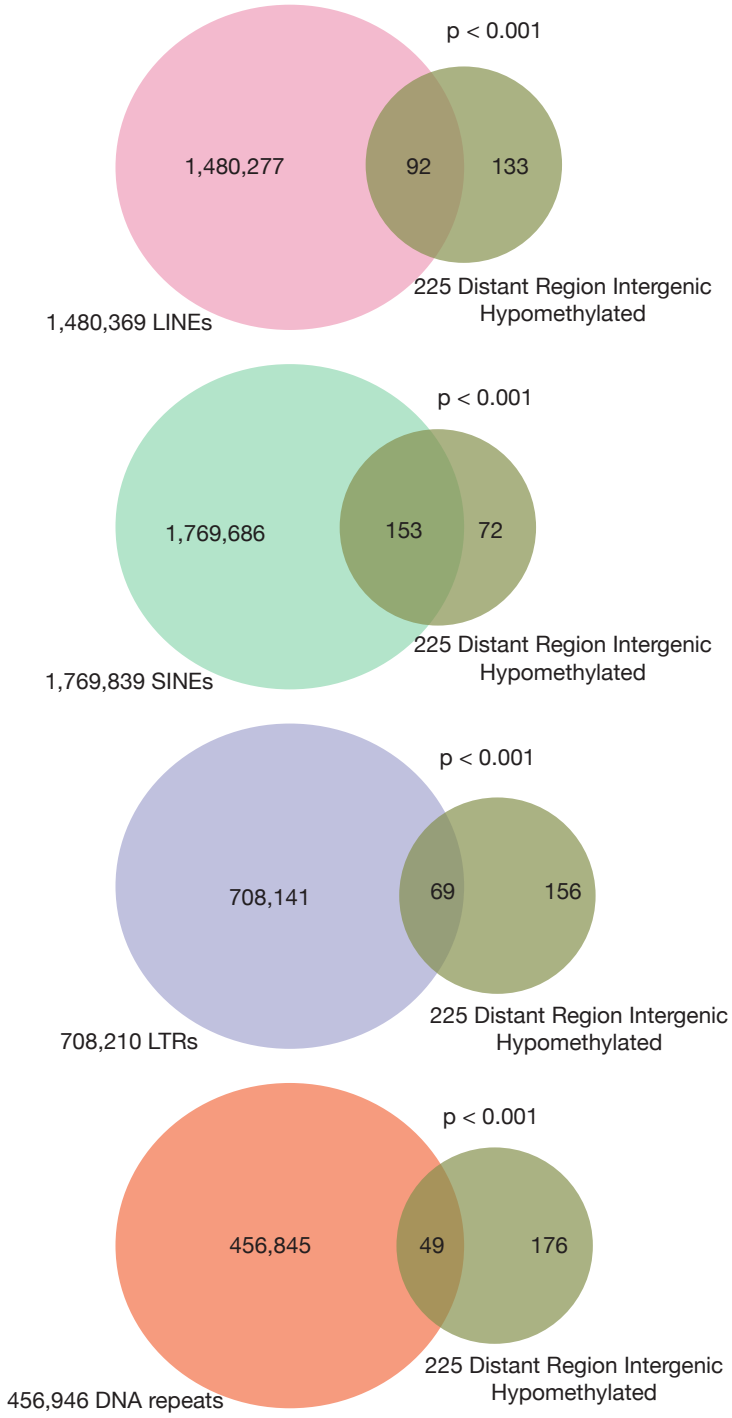
Supplementary Figure 4.17: Hypomethylated CpG Shore in intergenic regions overlap with several repeat elements.



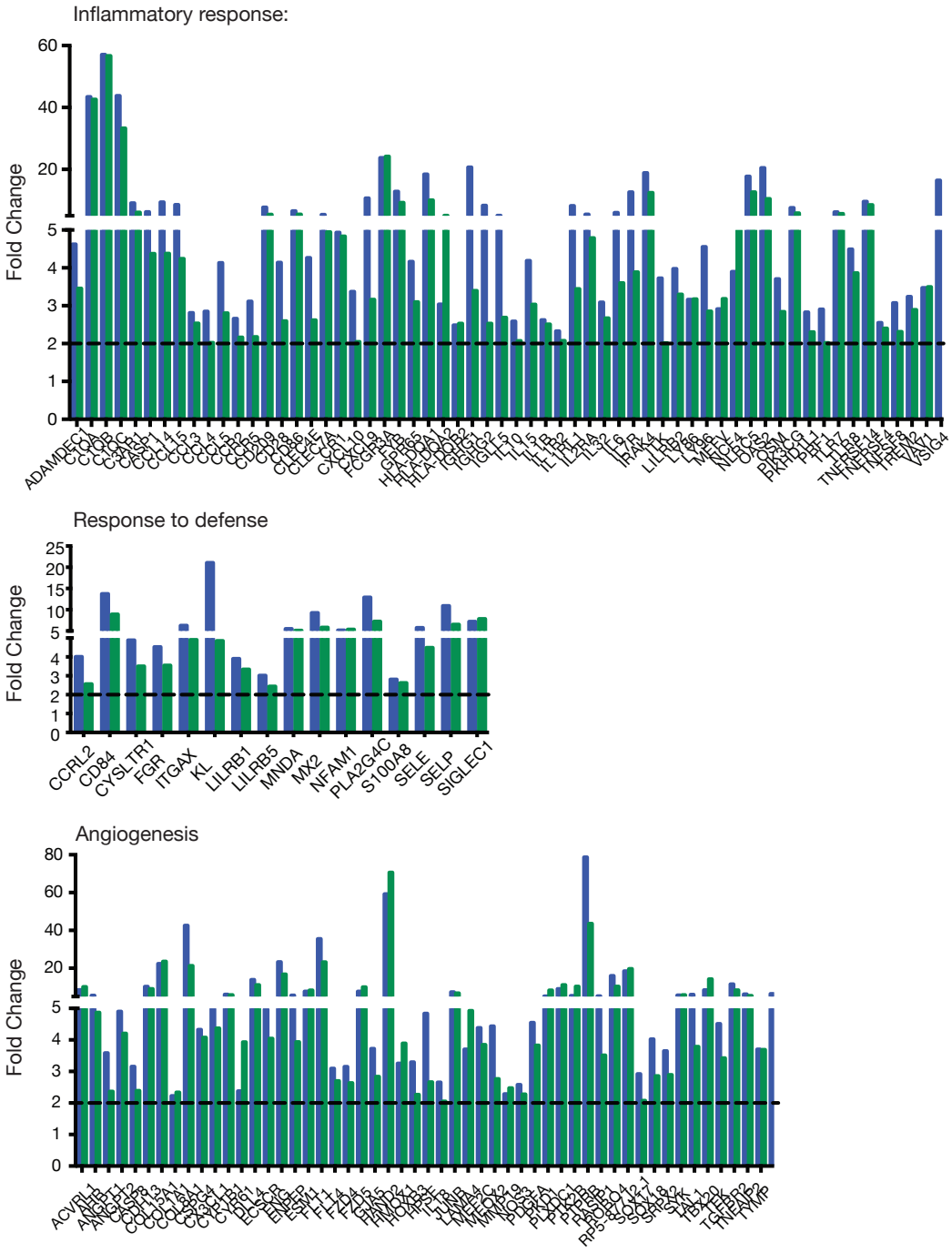
Supplementary Figure 4.18: Hypomethylated Distant Region (not associated with a CpG Islands) promoters overlap with several repeat elements.



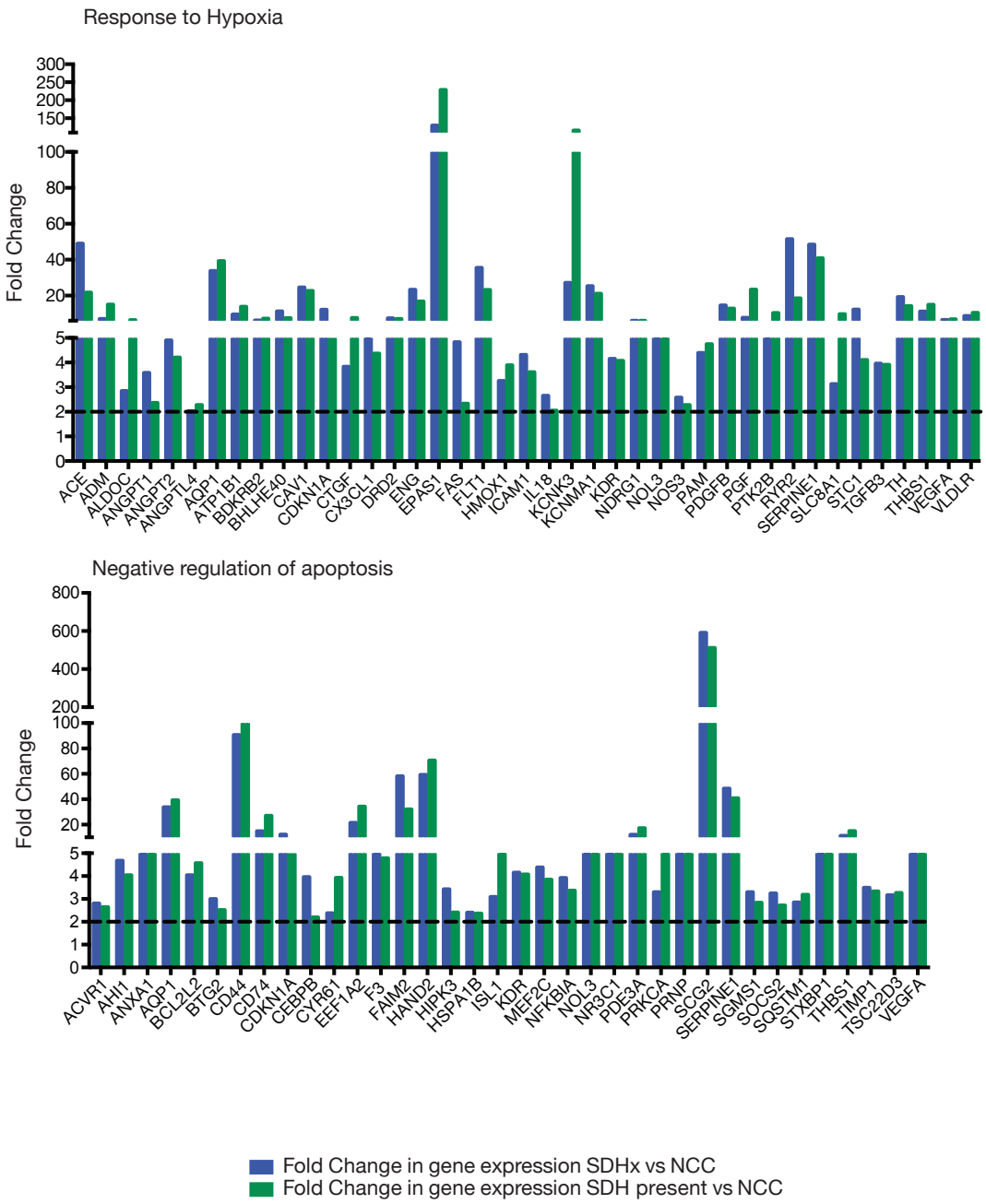
Supplementary Figure 4.19: Hypomethylated Distant Region (not associated with a CpG Island) body and 3'UTR regions overlap with several repeat elements.



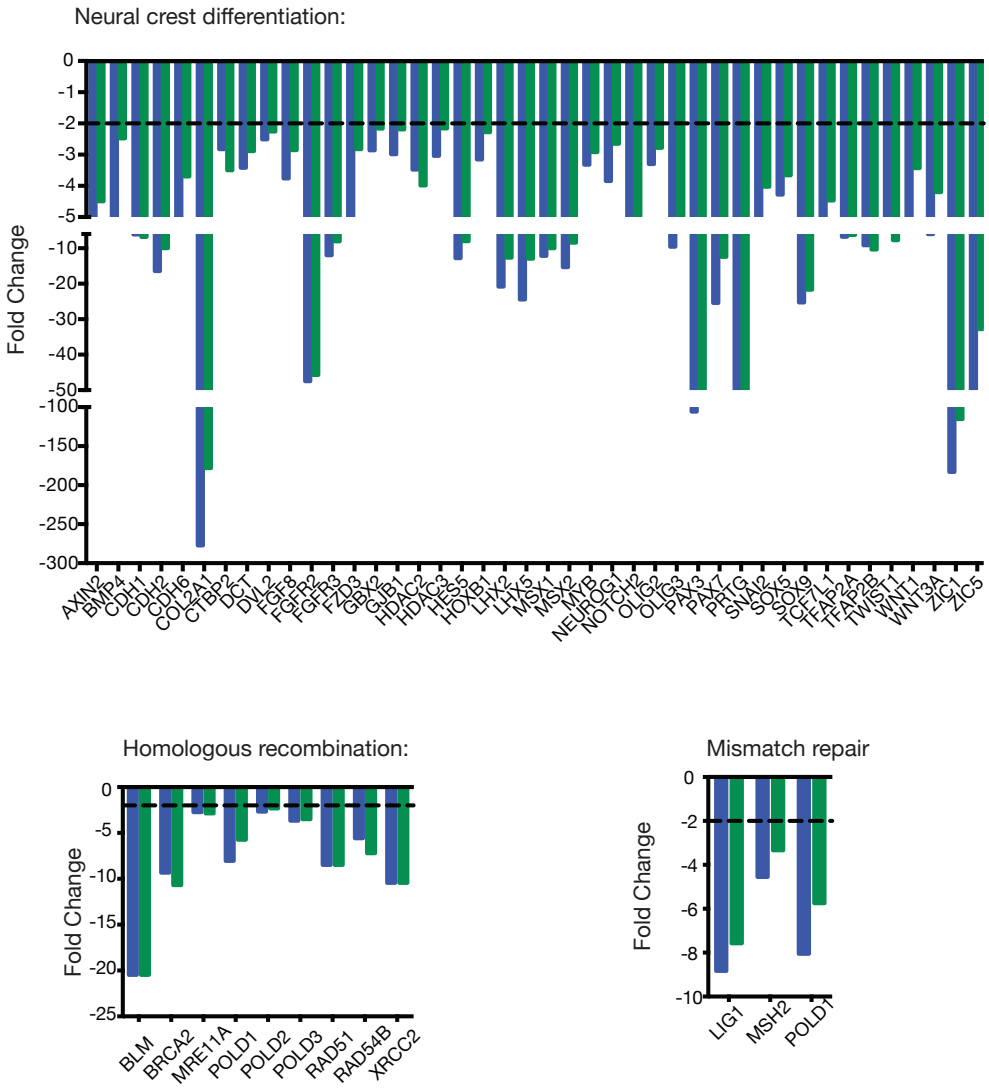
Supplementary Figure 4.20: Hypomethylated Distant Region (not associated with a CpG Island) intergenic regions overlap with several repeat elements.



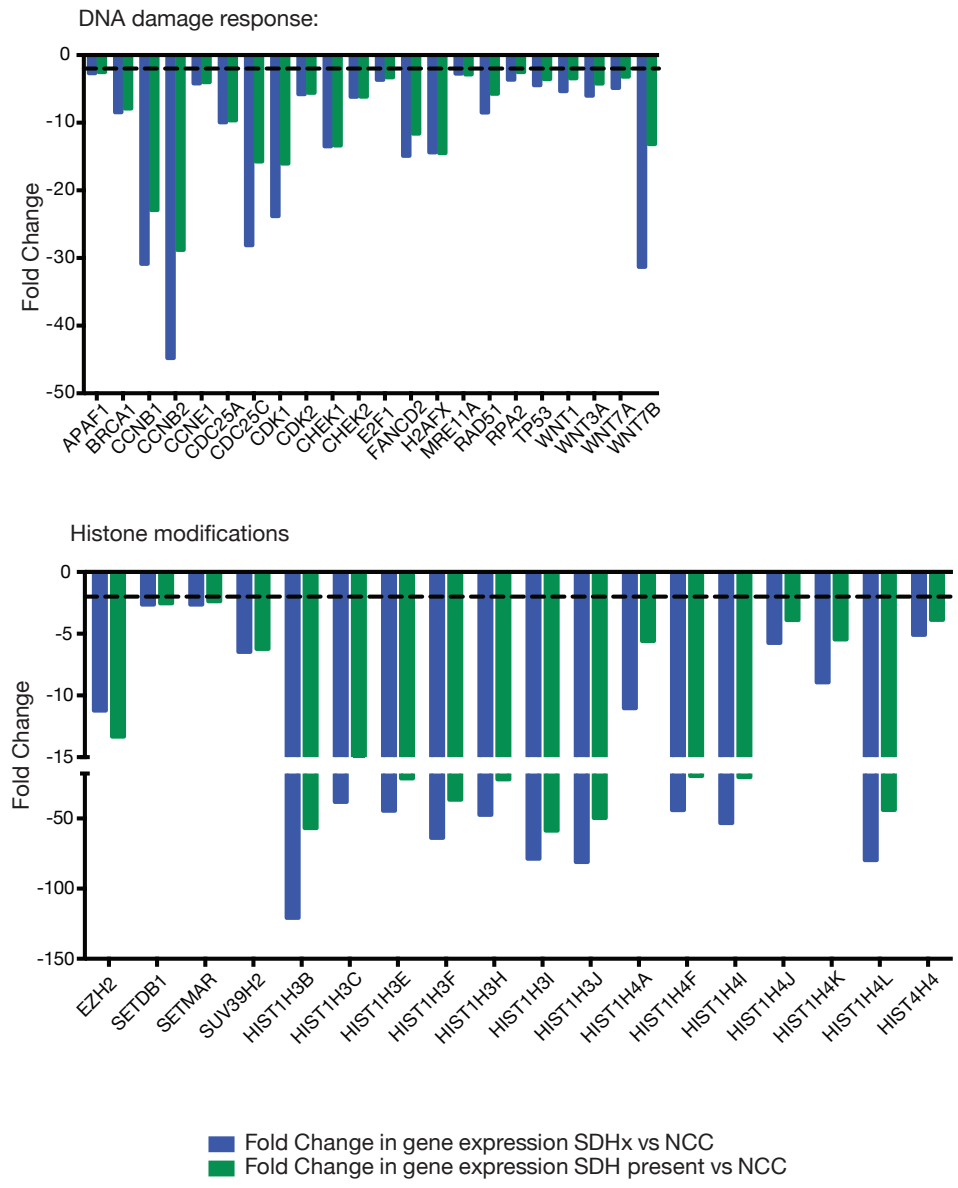
Supplementary Figure 4.21: Transcriptome analysis of SDH Deficient and SDH Present tumors as compared to neural crest cells. Upregulated gene candidates from enriched Go-terms (Figure 4.31) in SDH Deficient and SDH Present tumors over NCCs. (FDR < 0.001, fold change >= 2)



Supplementary Figure 4.21 continued.



Supplementary Figure 4.22: Transcriptome analysis of SDH Deficient and SDH Present tumors as compared to neural crest cells. Downregulated gene candidates from enriched Go-terms (Figure 6B) in SDH Deficient and SDH Present tumors over NCCs. (FDR < 0.001, fold change >= -2)



Supplementary Figure 4.8B continued.

4.7 References

1. Bardella C, Pollard PJ, Tomlinson I (2011) SDH mutations in cancer. Biochim Biophys Acta 1807: 1432-1443.

2. van Nederveen FH, Gaal J, Favier J, Korpershoek E, Oldenburg RA, et al. (2009) An immunohistochemical procedure to detect patients with paraganglioma and phaeochromocytoma with germline SDHB, SDHC, or SDHD gene mutations: a retrospective and prospective analysis. Lancet Oncology 10: 764-771.

3. Welander J, Soderkvist P, Gimm O (2011) Genetics and clinical characteristics of hereditary pheochromocytomas and paragangliomas. Endocrine Related Cancer 18: R253-R276.

4. Dahia PLM (2014) Pheochromocytoma and paraganglioma pathogenesis: learning from genetic heterogeneity. Nature Reviews Cancer 14: 108-119.

5. Hensen EF, Bayley JP (2011) Recent advances in the genetics of SDH-related paraganglioma and pheochromocytoma. Fam Cancer 10: 355-363.

6. Lendvai N, Pawlosky R, Bullova P, Eisenhofer G, Patocs A, et al. (2013) Succinate-to-fumarate ratio as a new metabolic marker to detect the presence of SDHB/D-related paraganglioma: initial experimental and ex vivo findings. Endocrinology 155: 27-32.

7. Smith EH, Janknecht R, Maher LJ (2007) Succinate inhibition of alpha-ketoglutarate-dependent enzymes in a yeast model of paraganglioma. Human Molecular Genetics 16: 3136-3148.

8. Selak MA, Armour SM, MacKenzie ED, Boulahbel H, Watson DG, et al. (2005) Succinate links TCA cycle dysfunction to oncogenesis by inhibiting HIF-alpha prolyl hydroxylase. Cancer Cell 7: 77-85.

9. Xiao M, Yang H, Xu W, Ma S, Lin H, et al. (2012) Inhibition of alpha-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. Genes Dev 26: 1326-1338.

10. Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. Cell 128: 707-719.

11. Sharma S, Kelly TK, Jones PA (2010) Epigenetics in cancer. Carcinogenesis 31: 27-36.

12. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics 13: 484-492.

13. Jones PA, Baylin SB (2007) The epigenomics of cancer. Cell 128: 683-692.

14. Baylin SB (2005) DNA methylation and gene silencing in cancer. Nature Clinical Practice Oncology 2: S4-S11.

15. Akiyama Y, Watkins N, Suzuki H, Jair KW, van Engeland M, et al. (2003) GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer. Molecular and Cellular Biology 23: 8429-8439.

16. Cervera AM, Bayley J-P, Devilee P, McCreath KJ (2009) Inhibition of succinate dehydrogenase dysregulates histone modification in mammalian cells. Molecular Cancer 8: 89.

17. Xu W, Yang H, Liu Y, Yang Y, Wang P, et al. (2011) Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. Cancer Cell 19: 17-30.

18. Ratcliffe PJ (2007) HIF-1 and HIF-2: working alone or together in hypoxia? Journal of Clinical Investigation 117: 862-865.

19. Vander Heiden MG, Cantley LC, Thompson CB (2009) Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science 324: 1029-1033.

20. Gilkes DM, Xiang L, Lee SJ, Chaturvedi P, Hubbi ME, et al. (2014) Hypoxia-inducible factors mediate coordinated RhoA-ROCK1 expression and signaling in breast cancer cells. Proceedings of the National Academy of Sciences of the United States of America 111: E384-393.

21. Killian JK, Kim SY, Miettinen M, Smith C, Merino M, et al. (2013) Succinate dehydrogenase mutation underlies global epigenomic divergence in gastrointestinal stromal tumor. Cancer Discov 3: 648-657.

22. Letouze E, Martinelli C, Loriot C, Burnichon N, Abermil N, et al. (2013) SDH mutations establish a hypermethylator phenotype in paraganglioma. Cancer Cell 23: 739-752.

23. Ilona Linnoila R, Keiser HR, Steinberg SM, Lack EE (1990) Histopathology of benign versus malignant sympathoadrenal paragangliomas: clinicopathologic study of 120 cases including unusual histologic features. Human Pathology 21: 1168-1180.

24. Lee G, Chambers SM, Tomishima MJ, Studer L (2010) Derivation of neural crest cells from human pluripotent stem cells. Nature Protocols 5: 688-701.

25. Knecht AK, Bronner-Fraser M (2002) Induction of the neural crest: a multigene process. Nature Reviews Genetics 3: 453-461.

26. Burnichon N, Vescovo L, Amar L, Libe R, de Reynies A, et al. (2011) Integrative genomic analysis reveals somatic mutations in pheochromocytoma and paraganglioma. Human Molecular Genetics 20: 3974-3985.

27. Bajpai R, Chen DA, Rada-Iglesias A, Zhang J, Xiong Y, et al. (2010) CHD7 cooperates with PBAF to control multipotent neural crest formation. Nature 463: 958-962.

28. Potok ME, Nix DA, Parnell TJ, Cairns BR (2013) Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. Cell 153: 759-772.

29. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454: 766-770.

30. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, et al. (2012) IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. Nature 483: 479-483.

31. Unland R, Schlosser S, Farwick N, Plagemann T, Richter G, et al. (2012) Epigenetic silencing of the dopamine receptor D4 in medulloblastoma. Klinische Pädiatrie 224: A4.

32. Ju JH, Yang W, Lee KM, Oh S, Nam K, et al. (2013) Regulation of cell proliferation and migration by keratin19-induced nuclear import of early growth response-1 in breast cancer cells. Clin Cancer Res 19: 4335-4346.

33. Fukui T, Kondo M, Ito G, Maeda O, Sato N, et al. (2005) Transcriptional silencing of secreted frizzled related protein 1 (SFRP 1) by promoter hypermethylation in non-small-cell lung cancer. Oncogene 24: 6323-6327.

34. Takada T, Yagi Y, Maekita T, Imura M, Nakagawa S, et al. (2004) Methylation-associated silencing of the Wnt antagonist SFRP1 gene in human ovarian cancers. Cancer Science 95: 741-744.

35. Suzuki H, Watkins DN, Jair K-W, Schuebel KE, Markowitz SD, et al. (2004) Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. Nature Genetics 36: 417-422.

36. Shih YL, Hsieh CB, Lai HC, Yan MD, Hsieh TY, et al. (2007) SFRP1 suppressed hepatoma cells growth through Wnt canonical signaling pathway. Int J Cancer 121: 1028-1035.

37. Lo PK, Mehrotra J, D'Costa A, Fackler MJ, Garrett-Mayer E, et al. (2006) Epigenetic suppression of secreted frizzled related protein 1 (SFRP1) expression in human breast cancer. Cancer Biol Ther 5: 281-286.

38. Harris T, Marquez B, Suarez S, Schimenti J (2007) Sperm motility defects and infertility in male mice with a mutation in Nsun7, a member of the Sun domain-containing family of putative RNA methyltransferases. Biology of Reproduction 77: 376-382.

39. Gao Q, Steine EJ, Barrasa MI, Hockemeyer D, Pawlak M, et al. (2011) Deletion of the de novo DNA methyltransferase Dnmt3a promotes lung tumor progression. Proceedings of the National Academy of Sciences 108: 18061-18066.

40. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, et al. (2014) Mutational landscape and significance across 12 major cancer types. Nature 502: 333-339.

41. Werner S, Frey S, Riethdorf S, Schulze C, Alawi M, et al. (2013) Dual roles of the transcription factor grainyhead-like 2 (GRHL2) in breast cancer. Journal of Biological Chemistry 288: 22993-23008.

42. Wang F, Travins J, DeLaBarre B, Penard-Lacronique V, Schalm S, et al. (2013) Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. Science 340: 622-626.

43. Berdasco M, Ropero S, Setien F, Fraga MF, Lapunzina P, et al. (2009) Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. Proceedings of the National Academy of Sciences 106: 21830-21835.

44. Morris MR, Ricketts CJ, Gentle D, McRonald F, Carli N, et al. (2011) Genome-wide methylation analysis identifies epigenetically inactivated candidate tumour suppressor genes in renal cell carcinoma. Oncogene 30: 1390-1401.

45. Tessema M, Yingling CM, Grimes MJ, Thomas CL, Liu Y, et al. (2012) Differential epigenetic regulation of TOX subfamily high mobility group box genes in lung and breast cancers. PLoS One 7: e34850.

46. Smith KT, Sardiu ME, Martin-Brown SA, Seidel C, Mushegian A, et al. (2012) Functional characterization of the human FAM60A protein: a new subunit of the Sin3 deacetylase complex. Molecular and Cellular Proteomics 11: 1815-1828.

47. Park S, Kim J-H, Jang J-H (2007) Aberrant hypermethylation of the FGFR2 gene in human gastric cancer cell lines. Biochemical and Biophysical Research Communications 357: 1011-1015.

48. Fortier AM, Asselin E, Cadrin M (2013) Keratin 8 and 18 loss in epithelial cancer cells increases collective cell migration and cisplatin sensitivity through claudin1 up-regulation. Journal of Biological Chemistry 288: 11555-11571.

49. Zhou Y-H, Wu X, Tan F, Shi Y-X, Glass T, et al. (2005) PAX6 suppresses growth of human glioblastoma cells. Journal of Neuro-Oncology 71: 223-229.

50. Shyr CR, Tsai MY, Yeh S, Kang HY, Chang YC, et al. (2010) Tumor suppressor PAX6 functions as androgen receptor co-repressor to inhibit prostate cancer growth. Prostate 70: 190-199.

51. Mayes DA, Hu Y, Teng Y, Siegel E, Wu X, et al. (2006) PAX6 suppresses the invasiveness of glioblastoma cells and the expression of the matrix metalloproteinase-2 gene. Cancer Res 66: 9809-9817.

52. Kilinc D, Ozdemir O, Ozdemir S, Korgali E, Koksal B, et al. (2012) Alterations in promoter methylation status of tumor suppressor HIC1, SFRP2, and DAPK1 genes in prostate carcinomas. DNA and Cell Biology 31: 826-832.

53. Veeck J, Noetzel E, Bektas N, Jost E, Hartmann A, et al. (2008) Promoter hypermethylation of the SFRP2 gene is a high-frequent alteration and tumor-specific epigenetic marker in human breast cancer. Molecular Cancer 7: 83.

54. Aleman A, Adrien L, Lopez-Serra L, Cordon-Cardo C, Esteller M, et al. (2007) Identification of DNA hypermethylation of SOX9 in association with bladder cancer progression using CpG microarrays. British Journal of Cancer 98: 466-473.

55. Guryanova OA, Sablina AA, Chumakov PM, Frolova EI (2005) Downregulation of TRIP6 gene expression induces actin cytoskeleton rearrangements in human carcinoma cell lines. Molecular Biology 39: 792-795.

56. Huynh HT, Larsson C, Narod S, Pollak M (1995) Tumor suppressor activity of the gene encoding mammary-derived growth inhibitor. Cancer Research 55: 2225-2231.

57. Song MA, Tiirikainen M, Kwee S, Okimoto G, Yu H, et al. (2013) Elucidating the landscape of aberrant DNA methylation in hepatocellular carcinoma. PLoS One 8: e55761.

58. Tchougounova E, Jiang Y, Bråsäter D, Lindberg N, Kastemar M, et al. (2009) Sox5 can suppress platelet-derived growth factor B-induced glioma development in Ink4a-deficient mice through induction of acute cellular senescence. Oncogene 28: 1537-1548.

59. Qin YR, Tang H, Xie F, Liu H, Zhu Y, et al. (2011) Characterization of tumor-suppressive function of SOX6 in human esophageal squamous cell carcinoma. Clin Cancer Res 17: 46-55.

60. Wolf SF, Jolly DJ, Lunnen KD, Friedmann T, Migeon BR (1984) Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. Proceedings of the National Academy of Sciences 81: 2806-2810.

61. Hellman A, Chess A (2007) Gene body-specific methylation on the active X

chromosome. Science 315: 1141-1143.

62. Kondo Y, Issa J-PJ (2010) DNA methylation profiling in cancer. Expert reviews in molecular medicine 12: e23.

63. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16: 6-21.

64. Scott KL, Nogueira C, Heffernan TP, van Doorn R, Dhakal S, et al. (2011) Proinvasion metastasis drivers in early-stage melanoma are oncogenes. Cancer Cell 20: 92-103.

65. Senchenko VN, Krasnov GS, Dmitriev AA, Kudryavtseva AV, Anedchenko EA, et al. (2011) Differential expression of CHL1 gene during development of major human cancers. PLoS ONE 6: e15612.

66. Kong HK, Yoon S, Park JH (2012) The regulatory mechanism of the LY6K gene expression in human breast cancer cells. Journal of Biological Chemistry 287: 38889-38900.

67. Oshima H, Ishikawa T, Yoshida GJ, Naoi K, Maeda Y, et al. (2013) TNF-alpha/TNFR1 signaling promotes gastric tumorigenesis through induction of Noxo1 and Gna14 in tumor cells. Oncogene 356: 1-10.

68. Liu R, Zhou Z, Huang J, Chen C (2011) PMEPA1 promotes androgen receptor-negative prostate cell proliferation through suppressing the Smad3/4-c-Myc-p21 Cip1 signaling pathway. J Pathol 223: 683-694.

69. Trerotola M, Cantanelli P, Guerra E, Tripaldi R, Aloisi AL, et al. (2013) Upregulation of Trop-2 quantitatively stimulates human cancer growth. Oncogene 32: 222-233.

70. Fulda S (2013) Targeting c-FLICE-like inhibitory protein (CFLAR) in cancer. Expert Opin Ther Targets 17: 195-201.

71. Baldanzi G, Mitola S, Cutrupi S, Filigheddu N, van Blitterswijk WJ, et al. (2004) Activation of diacylglycerol kinase alpha is required for VEGF-induced angiogenic signaling in vitro. Oncogene 23: 4828-4838.

72. Yanagisawa K, Yasuda S, Kai M, Imai S-i, Yamada K, et al. (2007) Diacylglycerol kinase alpha suppresses tumor necrosis factor-alpha-induced apoptosis of human melanoma cells through NF-kappaB activation. Biochimica et biophysica acta 1771: 462-474.

73. Bacchiocchi R, Baldanzi G, Carbonari D, Capomagi C, Colombo E, et al. (2005) Activation of alpha-diacylglycerol kinase is critical for the mitogenic properties of anaplastic lymphoma kinase. Blood 106: 2175-2182.

74. Carter H, Samayoa J, Hruban RH, Karchin R (2010) Prioritization of driver mutations in pancreatic cancer using cancer-specific high-throughput annotation of somatic mutations (CHASM). Cancer Biol Ther 10: 582-587.

75. Hayes GM, Carrigan PE, Miller LJ (2007) Serine-arginine protein kinase 1 overexpression is associated with tumorigenic imbalance in mitogen-activated protein kinase pathways in breast, colonic, and pancreatic carcinomas. Cancer Research 67: 2072-2080.

76. Hu-Lowe DD, Chen E, Zhang L, Watson KD, Mancuso P, et al. (2011) Targeting activin receptor-like kinase 1 inhibits angiogenesis and tumorigenesis through a mechanism of action complementary to anti-VEGF therapies. Cancer Research 71: 1362-1373.

77. Krneta J, Kroll J, Alves F, Prahst C, Sananbenesi F, et al. (2006) Dissociation of angiogenesis and tumorigenesis in follistatin-and activin-expressing tumors. Cancer Research 66: 5686-5695.

78. Subimerb C, Pinlaor S, Lulitanond V, Khuntikeo N, Okada S, et al. (2010) Circulating CD14(+) CD16(+) monocyte levels predict tissue invasive character of cholangiocarcinoma. Clinical and Experimental Immunology 161: 471-479.

79. Lindsey SF, Byrnes DM, Eller MS, Rosa AM, Dabas N, et al. (2013) Potential role of meiosis proteins in melanoma chromosomal instability. J Skin Cancer 2013: 190109.

80. Span PN, Rao JU, Oude Ophuis SBJ, Lenders JWM, Sweep FCGJ, et al. (2011) Overexpression of the natural antisense hypoxia-inducible factor-1alpha transcript is associated with malignant pheochromocytoma/paraganglioma. Endocrine Related Cancer 18: 323-331.

81. Lopez-Jimenez E, Gomez-Lopez G, Leandro-Garcia LJ, Munoz I, Schiavi F, et al. (2010) Research resource: transcriptional profiling reveals different pseudohypoxic signatures in SDHB and VHL-Related pheochromocytomas. Molecular Endocrinology 24: 2382-2391.

82. De Santa F, Totaro MG, Prosperini E, Notarbartolo S, Testa G, et al. (2007) The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. Cell 130: 1083-1094.

83. Margueron R, Reinberg D (2010) Chromatin structure and the inheritance of epigenetic information. Nature Reviews Genetics 11: 285-296.

84. Laget S, Joulie M, Le Masson F, Sasai N, Christians E, et al. (2010) The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. PLoS ONE 5: e11982.

85. Du Y, Liu B, Guo F, Xu G, Ding Y, et al. (2012) The essential role of Mbd5 in the regulation of somatic growth and glucose homeostasis in mice. PLoS ONE 7: e47358.

86. Cox BJ, Vollmer M, Tamplin O, Lu M, Biechele S, et al. (2010) Phenotypic annotation of the mouse X chromosome. Genome Res 20: 1154-1164.

87. Kooistra SM, Helin K (2012) Molecular mechanisms and potential functions of histone demethylases. Nature Reviews Molecular Cell Biology 13: 297-311.

CHAPTER 5

DISCUSSION

5.1 Human Sperm is Poised for Embryo Development by the Presence of Histone Modifications and DNA Hypomethylation

During spermiogenesis, which is the last stage of spermatogenesis, chromatin packaging in sperm undergoes significant compaction by the exchange of majority of its histone for protamine. Protamines are basic, small, toroid shaped proteins that wrap DNA tightly and hence are imperative towards compaction of chromatin in mature sperm[1,2]. This compaction contributes towards sperm maturation, fertility and is crucial for genome transport[3,4]. In human sperm, only ~5% of histone is retained in the genome and the rest is replaced by protamine[5]. We asked the intriguing question whether the remaining histones marked promoter regions of genes involved in early embryo development or if they were marked promoters of genes involved in spermatogenesis and hence were simply remnant of that developmental program.

We found that canonical histones (H2A, H2B, H3 and H4) and testis-specific histone variant (tH2B) as well as H2Az comprised the majority of retained histone in the mature human sperm. We performed ChIP (chromatin immunoprecipitation) in order to investigate if histone modifications marked promoters of genes involved in early embryo development. We challenged the notion that the retained histones were simply remnants of a spermatogenesis program or left behind as a consequence of incomplete histone to protamine exchange. Strikingly, we observed that histones were indeed retained at promoters of several genes including those involved in early development, signaling pathways, miRNAs and spermiogenesis. However, in order to have any potential paternal contribution towards embryonic development, the developmental genes and signaling pathways needed have secondary modifications or variants in order to distinguish them from the majority of the genome that would acquire acetylated histone modifications following protamine replacement after fertilization[5].

Since a large percentage of the histone retained was testis-specific H2B, we investigated its distribution in mature sperm via ChIP. Analysis showed that it was enriched at genes involved in ion channels and spermiogenesis. We further investigated the localization of another histone variant, H2Az, which has been implicated in embryonic stem cells to mark promoters of genes that are targets of the polycomb complex. Also, the presence of H2Az is anticorrelated with the presence of DNA methylation and hence genes that need to be poised for activation during early embryonic development may enrich for this mark. However, consistent with previously published immunohistochemistry staining, H2Az enriched at pericentric heterochromatin in mature sperm. Hence, we speculated that alternate histone modifications, such as H3K4me3, H3K4me2 and H3K27me3 may enrich at the promoters of developmental genes and indeed our data confirmed these expectations. Interestingly, H3K4me3 mostly marked genes involved in spermatogenesis where as genes necessary for embryonic development were bivalently marked with H3K4me2 and H3K27me3, a hallmark also observed in ES cells. These bivalently marked genes were also DNA hypomethylated. We speculated that the presence of H3K27me3 (a repressive mark) and H3K4me2 (an active mark) allowed these genes to be transcriptionally repressed in the sperm but poised for activation in the early embryo. These regions are protected from DNA methylation, since the presence of this epigenetic mark is viewed as “locking in” the decision to be transcriptionally silenced long term. Notably, some of these developmental promoters are downstream targets of transcription factors involved in self-renewal, such as OCT4, SOX2, NANOG, KLF4 and FOXD3. While these self-renewal factors themselves acquire DNA methylation (DNAme) at their promoters during spermatogenesis, their downstream target genes remain hypomethylated, consistent with findings in mice. These developmental promoters will selectively gain DNA methylation during development when the cells will commit

to a differentiation lineage. Overall, our data shows that histone retention and DNA hypomethylation contribute to a poised state in the mature human sperm that may contribute towards transcriptional activation of early developmental genes in the embryo[5].

Furthermore, these findings were also confirmed in mouse sperm and zebrafish sperm. The genome in mouse sperm is also packaged by protamine and histone. However, in contrast to humans, 99% of the genome in mouse sperm is packaged by protamine and only 1% is packaged by histone [6]. The study confirmed that genes important for spermatogenesis and spermiogenesis were marked with H3K4me2. Genes involved in early development were either mostly marked by H3K27me3 or were bivalently marked by H3K27me3 and H3K4me2 and both classes of these genes were generally DNA hypomethylated[7]. In zebrafish sperm, the overarching theme was also preserved where developmentally important genes were enriched with multivalent chromatin marks. In contrast to human sperm, the zebrafish sperm genome is packaged exclusively in histone. Hence, if these marks are instructive for development of the early embryo, the presence of histone variants and histone modifications at promoters of genes was even more significant in this system to maintain robustness in the embryo and distinguish them from canonical histones. These genes were marked by multivalent marks including the repressive H3K27me3 and active H3K4me2/me3 and H2Az variant [8].

5.2 Aberrations in DNA Methylation at Distinct Imprinted Loci in Sperm of Infertile Patients with Abnormal Protamine Ratios

A small percentage of offspring conceived through ART develop imprinting disorders. This could be either due to the ART procedure or due to aberant DNA methylation patterns at imprinted genes in the gametes of the parents. One goal

to understanding chromatin aberrations in infertility was to gain further insight into how this might impact imprinting abnormalities associated with offspring conceived through ART (assisted reproductive technologies) and if this could potentially serve as a diagnostic tool. While other possible contributing factors such as *in vitro* manipulations of embryos or gametes and ovulation induction medication cannot be excluded, our previous work on and interest in the role of chromatin packaging in male gametes guided us to focus our efforts on investigating if aberrations in DNA methylation at imprinted regions in infertile males was misregulated and hence a contributing factor to increased imprinting disorders in IVF offspring[9-12]. As previously mentioned, during spermiogenesis, histones are replaced for protamines and this is crucial for the tight compaction of the sperm genome. Only about 5% histone is retained while the remainder of the genome is packaged in protamine. One important aspect of the chromatin packaging of the sperm genome is the proportion of protamine 1 vs protamine 2. In fertile men, the ratio of P1/P2 is strictly regulated and alterations in the ratio have been observed in infertile men[13]. These altered P1/P2 ratios in infertile men are not only associated with altered sperm quality but also decreased embryo quality and IVF (*in vitro* fertilization) outcome in comparison to infertile men with normal P1/P2 ratios[14,15]. We speculated that the root of misregulation in infertile men with altered P1/P2 ratio might lie in chromatin packaging. Hence, we were interested in understanding if histone retention and localization was also affected. If histone modifications retained at the promoters of developmental genes are instructive for embryo development, it could potentially explain decreased embryo quality and IVF outcome experienced by this group of patients. In addition, we were interested in understanding if DNA methylation was also aberrantly affected. Previously published studies had reported changes in DNA methylation at imprinted genes such as H19 and Mest[9,10,16]. Hence, we focused our efforts on understanding whether imprinted genes suffered from

aberrant DNA methylation in infertile patients with altered P1/P2 ratio and if this misregulation could correlate to poor IVF outcome and developmental disorders in offspring produced by ART[17].

Along these lines, we chose to study the DNA methylation patterns at seven imprinted genes, Peg3, Mest (involved in Silver Russell syndrome), SNRPN (involved in Angelman and Prader-Wili syndrome), LIT1 (involved in Beckwith-Weidemann syndrome), IGF2 and H19 (involved in Wilms tumors, Silver Russell syndrome and Beckwith-Weidemann syndrome) and Zac (involved in transient neonatal diabetes mellitus)[10,16,18-20]. We studied two groups of infertility where patients suffering from oligozoospermia (low sperm count) and patients with abnormal P1/P2 ratio in their sperm were included. Our results confirmed the previously published finding that MEST was hypermethylated in oligozoospermic patients. However, patients with abnormal protamine ratios had significant changes in DNA methylation at LIT1 and SNRPN. Interestingly, patients with aberrant DNA methylation at LIT1 parsed into three categories: those that were unaffected / hypomethylated, completely methylated and partially methylated. We predicted that infertile patients that had complete hypermethylation at CpGs in the DMR of LIT1 in their sperm would increase the risk of their offspring developing Beckwidth-Weidermann syndrome. Also of note, CpGs in the DMR of SNRPN were not completely methylated. We observed partial but significant DNA methylation at this loci and hence again would predict that the offspring of these infertile men might have a higher risk of developing Prader-Wili syndrome. Notably, not all patients or alleles were affected to the same extent. Also, there was no significant co-variance of methylation defects at LIT1 and the other loci. Hence, the risk of transmitting epigenetic alterations may vary with a subset of genes that are misregulated, and with degree of change in DNA methylation at these imprinted regions. This difference raises an important question regarding the variable risk associated with different CpGs and if there is a certain

threshold level for conferring disease risk. So, while our study did not provide a causal link for the trans-generational inheritance of DNA methylation defects and imprinting diseases, it showed a strong correlation between infertility in males and aberrations of DNA methylation at select imprinted loci and serves as a diagnostic tool to help inform infertile men of their possible risks of having a child with an imprinting disorder. How and to what degree these epigenetic aberrations impact the risks associated with developmental disorders in IVF offspring will be a major focus of long-term outcome studies pursued in the field in the future.

5.3 Similar DNA Methylation and Transcription Profiles in SDH Deficient and Present PGLs may be Explained by Misregulation of Epigenetic Enzymes Targeting a Similar Pathway

One of the themes of our research was to understand how germ cell DNA was packaged and if misregulation of this packaging can be seen in infertility and cancer. While our previous studies focussed on chromatin packaging in mature sperm from normal donors and infertile patients, our most recent study involved looking at paragangliomas (PGLs). We investigated PGLs as we were interested in understanding how defects in metabolic enzymes such as SDH, FH and IDH can have a direct impact on the epigenome of tumors and thus the transcriptome of cells. Xiao *et al*., demonstrated in *in vitro* and cell line assays that succinate, and its structurally similar metabolite, fumarate, competitively inhibit the activity of α-KG-dependent dioxygenases such as JHDM2A (a human histone H3K36 demethylase), CeKDM7A (a *Caenorhabditis elegans* dual-specificity demethylase that recognizes methylated H3K9), HIF (hypoxia inducible factor involved in oxygen sensing), as well as TET1 and TET2 (involved in DNA methyl-cytosine hydroxylation). Further, upon ectopic expression of tumor-derived SDH and FH mutations, they showed an accumulation of succinate or fumarate, respectively,

which led to alterations in genome-wide histone and DNA methylation levels. These changes were speculated to contribute towards tumorigenesis[21,22]. Another group[23] was first to perform a restricted methylome analysis on a large cohort of paragangliomas harboring mutations in RET, NF1, VHL and SDHx genes. They found that SDH Deficient tumors gained DNA methylation at genes in promoters of CpG Islands, however, very few of these changes correlated with downregulation of gene expression. Namely, genes involved in neuroendocrine differentiation and catecholamine metabolism were mainly affected. They also found a single gene involved in EMT (epithelial-to-mesenchymal transition) and a single tumor suppressor to be affected. While their findings revealed interesting differences between SDH Deficient and Present tumors, their analysis was limited by a lack of comparison to a progenitor cell. Consequently, the highlighted changes explain differences in PGLs with different genetic backgrounds but failed to determine gene candidates that may be involved in tumor initiation and progression. Further, their DMR analysis was performed by a single CpG analysis where the majority of DNA methylation changes observed were less than 15%. Finally, they limited their analysis to changes of DNA methylation in CpG Islands, where as several studies have shown that changes in DNA methylation at CpG Shores (approximately 2kb upstream or downstream from a CpG Island) may also play a significant role in modulating gene expression which in turn may drive tumorigenesis[24-27].

Similar to SDH Deficient PGLs, gliomas have mutations in another TCA cycle enzyme, isocitrate dehydrogenase (IDH). IDH catalyzes the oxidative decarboxylation of isocitrate, producing alpha-ketoglutarate. However, mutant IDH loses the ability to sufficiently generate the physiologically normal product α-KG and instead gains the function of mainly producing the onco-metabolite 2-hydroxyglutarate (2-HG). 2-HG is shown to act as a competitive inhibitor of the a-KG-dependent dioxygenases[28-31]. A recent study profiled changes in DNA

methylation in IDH mutant and wildtype gliomas, correlated these changes to gene expression and then further demonstrated that IDH mutations were sufficient to establish both DNA methylation changes and transcriptome changes in immortalized primary human astrocytes[32-35]. Additionally Lu *et al.* reported that IDH mutations impair histone demethylation, blocking the differentiation of lineage-specific progenitor cells into terminally differentiated cells. They demonstrated bulk gain of several histone modifications in their adipocyte cells expressing IDH-mutants and showed enrichment of these modifications at promoters of genes involved in adipocyte differentiation, resulting in transcriptional repression[36-42]

Extending the link between the Krebs cycle, epigenetic changes and cancer, another study highlighted divergent global changes in DNA methylation in gastrointestinal stromal tumors (GISTs) that harbored mutations in SDH genes or in other non-SDH related susceptibility genes[21,43,44]. They validated the link between SDH mutations and methyl-divergence in tumorigenesis by comparing SDH Deficient, hereditary paragangliomas to adrenal medulla as reference tissue and IDH mutant gliomas to normal glial tissue. Globally, by principal component analysis, they found that the Krebs-cycle mutant tumors were more closely related to each other than nonmutant tumors. They also found similar targets that were hypermethylated and hypomethylated in the Krebs-cycle mutant tumors. While this study identifies similarities in DNA methylation patterns from tumors of different developmental lineages sharing mutations in related Krebs-cycle enzymes, it notably did not establish a link between DNA methylation changes and gene expression. From previously published reports[23] and our study, only a small fraction (~10-20%) of regions that have differential DNA methylation actually correlate with changes in transcription. Hence, in the Killian *et al.* study the interpretation of the biological contribution of the reported DNA methylation changes in the tumorigenesis of Kreb-cycle mutant tumors is limited.

In our study, we demonstrated that SDH Deficient and SDH Present PGLs share very similar DNA methylation changes compared to a progenitor cell type, neural crest cells (NCCs). The two PGL subclasses also share very similar transcriptome profiles compared to NCCs. This is in stark contrast to previously published reports that have focused on changes between tumor subtypes – those having mutations in SDH and those having mutations in other susceptibility genes such as RET, NF1, TMEM127 and MAX.

The Illumina 450K array focuses on CpG rich regions, thus harboring an inherent bias towards promoters. Hence, this analysis is by no means as comprehensive as whole genome bisulfite sequencing (WGBS) which can be extremely cost prohibitive for getting enough coverage over all CpGs in regulatory regions for a large sample set. However, our analysis reveals genes hypermethylated at CpG Island promoter regions and transcriptionally repressed such as DRD4, KRT19, and FRZB[45-52]. These genes have been reported to be epigenetically silenced in a range of different cancers and may serve as potential tumor suppressors. Interestingly, these were the only genes we parsed in our entire study that were exclusively affected in SDH Deficient paragangliomas. All other gene candidates were equally affected in both tumor subclasses. DNMT3A, GRHL2, KAZALD1, NSD1, ATP5G2 and TOX3 were hypermethylated in their CpG Islands promoters in both tumor subclasses and were either transcriptionally repressed or mutated in other cancers[53-59]. In promoters located in CpG Shores, genes involved in cell morphology changes, cell migration (FAM60A, KRT8, TRIP6), tumor suppressors (FGFR2, PAX6) and genes that are epigenetically silenced in other tumors such as SFRP2 and SOX9 were also found hypermethylated and downregulated in PGLs[60-69]. Intriguingly, JARID2, TET1, SALL1 and SALL4 are all genes involved in epigenetic regulation and were also found to be hypermethylated at their CpG Shores and transcriptionally downregulated in both PGL subclasses.

At promoters which fell into regions distant from CpG Islands three potential tumor suppressors (FABP3, SOX5 and SOX6) are also downregulated and DNA hypermethylated. They were also reported previously to be affected in a variety of other cancers[70-73].

We compared our gene candidates to previously published tumors with mutations in TCA cycle genes[21]. Four genes, BMP4, FABP3, FRZB and TRIP6 intersected with high statistical significance ($p < 0.001$) with hypermethylated genes in IDH1 gliomas, SDH Deficient GISTs, and PGLs from Killian *et al*. Two other genes, KRT19 and RPP25, overlapped with less statistical significance ($p < 0.0127$) with hypermethylated genes in SDH Deficient PGLs and SDHB knock out mouse chromaffin cells from Letouze *et al*. Finally, we observed 6 genes TGIF1, RPP25, TMEM159, DNMT3A, TRIP6, GAP43, in common with IDH1 gliomas from Turcan *et al.* ($p <0.0085$). Notably, these genes were found hypermethylated (in most cases) and repressed in both SDH Deficient and Present PGLs used in our study. Hence, while mutations in Krebs-cycle genes may be strongly associated with misregulation of these genes, we speculate that both tumors may adopt different mechanisms to achieve misregulation of similar gene targets. While, in SDH Deficient tumors, misregulation is driven by the accumulation of a co-factor required for the activity of epigenetic enzymes, SDH Present tumors may harbor mutations in epigenetic enzymes that could phenocopy the misregulation observed in SDH Deficient tumors.

Another interesting observation from our data is that a majority of the differentially methylated regions (DMRs) were actually losing DNA methylation. From our working model, succinate accumulation would inhibit TET enzymes from hydroxylating 5mC to 5hmC and hence we would expect a gain in DNA methylation[22]. However, on a global level we observed that SDH Deficient and SDH Present PGLs, compared to ESCs and NCCs, were hypomethylated. While these

findings may seem paradoxical, the presence of hypomethylated regions in PGLs is not completely surprising since global hypomethylation has been observed in many cancers and is usually associated with repetitive regions that are normally DNA methylated. Hypomethylation of repeat elements is correlated to a decrease in genome stability, which is mediated by recombination between nonallelic repeats, causing an increase in chromosome rearrangements or translocations. Also hypomethylated, and thus more active retrotransposons, can integrate into genes and disrupt them[24,26]. While this level of hypomethylation in SDH disrupted tumors would not be predicted from our working model, it highlights the fact that epigenetic misregulation in PGLs is a dynamic process involving a combination of gain and loss of DNA methylation acting in concert with mislocalization of histone modifications to potentially promote tumorigenesis. It is important to note, however, that PGLs deficient in SDH globally lose less DNA methylation than SDH Present PGLs when compared to progenitor cells, arguing strongly for SDH's role in inhibiting DNA demethylation.

Among genes that were hypomethylated and gained gene expression, we found several interesting candidates involved in promoting invasion and metastasis[28,30,31] (ACP5, CHL1, CPEB4, DOCK2, LY6K), cell proliferation and tumor growth [32,34,35] (GNA14, PMEPA1, TACSTD2), inhibition of cell death[36,38-42] (BCL2L1, CFLAR, DGKA, SRPK3), promoting angiogenesis[43,44] (ALK1 and FSTL3), chronic inflammation[74] (CD14) and genomic instability (REC8)[75]. None of these genes were previously reported to change in PGLs, since previously published studies focused mainly on regions that were gaining DNA methylation. Our data demonstrate that it is important to look at all changes regardless of their location (CpG Island, Shore or Distant Regions) and direction of change (both gain and loss of DNAme), to get a complete picture of all the aberrant changes potentially contributing to tumorigenesis.

Finally, we looked at the transcriptional changes in SDH Deficient and SDH Present PGLs and compared them to NCCs. Our data agreed with previously published reports that showed upregulation of genes involved in hypoxia, angiogenesis, blood vessel development, inflammatory response, defense response, myc transcription, negative regulation of apoptosis and cap-dependent translation. Further, our data also agree with studies showing downregulation of genes involved in DNA damage response, homologous recombination, base excision repair, nucleotide excision repair, neural crest differentiation and histone modifications[23,76-78]. Strikingly, both SDH Deficient and SDH Present PGLs had very similar transcriptome profiles compared to NCCs. One reason for this could be that we are the first to compare tumor transcriptome profiles to a progenitor cell type, instead of comparing them to each other. Another reason for this could be that the SDH Present tumors of our study have been categorized as wildtype/ benign tumors by other studies. Previous reports have demonstrated that wildtype PGLs can be transcriptionally similar to either SDH Deficient and VHL mutated tumors or to RET, NF1 and TMEM127 tumors[77]. The reason why wildtype tumors may be related to either class remains unclear.

To identify a link possibly explaining the similar epigenetic and transcription profiles of SDH Deficient and Present PGLs, we performed whole-exome sequencing, where preliminary data point towards mutations in epigenetic genes in SDH Present tumors. These genes may be part of the same epigenetic axis affected in SDH Deficient, succinate-accumulating PGLs. For example, mutations in KDM6B and MLL4, key epigenetic enzymes, are very intriguing candidates. KDM6B is a H3K27me2 and me3 demethylase and is found in a complex with MLL4, an H3K4 methyltransferase[79]. MLL4 and KDM6B are perhaps recruited to promoters of genes by transcription factors to methylate H3K4 and demethylate H3K27, allowing for transcriptional activation of the gene[79]. Mutations in MLL4 and

KDM6B were not found in their catalytic domains, however, it is entirely possible that the mutations in these genes may disrupt interactions with transcription factors and/or interactions with each other or modulate their enzymatic activity. This would cause the affected gene to remain aberrantly silenced. Mutations in other H3K4 methyltransferases, SETD1A and SETD1B have also been observed that might deter them from being recruited to their appropriate target regions leading to inactivation of the gene. JARID2, a member of the Jumonji family of proteins lacking demethylase activity, is known to bind GC-rich DNA and recruit the Ezh2/ PRC2 complex to its target sites where Ezh2 methylates H3K27 to transcriptionally silence the genes[80]. We hypothesize that mutations in JARID2 may disrupt its ability to interact or recruit PRC2 to its target sites, leading to aberrant gene activation. MBD5 contains a methyl-binding domain but does not bind to methylated DNA. It is known to associate with heterochromatin and hence may contribute to its formation[81]. Its deficiency is linked to developmental disorders, as knockout mice show growth retardation and preweaning lethality[82]. While its exact role may not be fully understood, mutations in this gene might affect its interaction with heterochromatin and this would be the first report implicating a mutation in MBD5 in cancer. Knock-out mouse models of KDM5C, which encodes an H3K4me2 and me3 demethylase, have neurulation and cardiac looping defects[83]. Mutations in KDM5C are associated with mental retardation, autism and renal carcinoma [84]. Taken together, while these genes do not represent an exhaustive list of all possible genes that are either mutated or transcriptionally misregulated in SDH Deficient and SDH Present tumors, these candidates may provide a novel and fascinating link, explaining the strikingly similar epigenetic and transcriptional profiles observed in our SDH Deficient and SDH Present PGLs.

Finally, while mutations in different susceptibility genes (SDHx, VHL, EPAS1, RET, NF1, TMEM127, MAX) have been reported in paragangliomas,

misregulation in each tumor subclass can mechanistically converge onto shared pathways towards proliferation. For example, gain-of-function mutations in RET, and loss-of-function mutations in NF1 and TMEM127 can activate the PI3K pathway which will consequently activate mTOR. Myc, no longer bound by MAX (due to loss of function mutations), cooperates with mTOR and thus activates it as well. mTOR activation can regulate cell growth through increased synthesis of nucleic acids, lipids, fatty acid, proteins and most importantly can activate HIF[85]. Accordingly, mutations in the SDH complex cause accumulation of succinate, which competitively inhibits PHDs, leading to stabilization of HIF. Activated HIF regulates transcription of its downstream targets involved in increased glucose uptake, glycolysis, angiogenesis and metastasis. Succinate accumulation can also inhibit the activity of α-KG dependent epigenetic enzymes, which in turn can lead to epigenetic misregulation of downstream targets that may contribute further to PGL oncogenesis as previously discussed. We speculate that in PGLs lacking SDH mutations, epigenetic enzymes may harbor mutations that could phenocopy the misregulation in SDH Deficient tumors.

5.4 Perspective and Future Directions

In conclusion, from our sequencing data, we have identified mutations in epigenetic enzymes in our SDH Present PGLs . These enzymes lie downstream to the SDH complex and hence SDH Present PGLs are able to phenocopy epigenetic misregulation observed in the SDH Deficient PGLs. However, as a near term goal, we need to also perform whole exome sequencing on the germline from SDH Present patients. This would help us determine if the identified mutations are germline or tumor associated. In addition, sequencing more additional PGLs along with their germline is required to confirm if the common theme of mutations in epigenetic enzymes in maintained in the majority of PGLs, especially those that

do not harbor mutations in the other susceptibility genes (RET, NF1, MAX, and TMEM127). Further, while we have identified gene candidates involved in PGL oncogenesis that are aberrantly methylated and transcriptionally affected, it is imperative that we continue to probe the chromatin landscape and identify regions of histone modification accumulation and mislocalization. It has been established in previous studies that succinate accumulation can cause bulk accumulation of histone modifications such as H3K4me1, me3 and H3K27me2 in HEK293T and HeLa cells. However, their genomic mislocalization has not been determined in the context of neural-crest derived tumors. This is most likely due to the lack of enough material from patient-derived paragangliomas, pointing towards the need to develop tumor derived cell lines with the ability to accumulate succinate. A PGL cell line system would allow us to identify the genomic locations of histone modification alterations, such as H3K27me3 and H3K4me3. We would focus efforts on understanding mislocalization of these histone modifications since many of the PGL tumors sequenced by us harbor nonsynomous mutations in KDM6B, an H3K27 demethylase, and MLL, SETD1A, SETD1B and PRDM2, H3K4 methyltransferases. In addition, we would like to identify direct downstream targets of transcription factors such as HIF and Myc that may be involved in both driving and progressing PGL oncogenesis. Once PGL cell lines have been established, we can screen for candidates that directly impact growth, proliferation and drug resistance using the CRISPR-Cas9 system. Recent studies used this powerful tool for systematic genetic analysis where they developed over 65,000 single-guide RNAs (sgRNAs) to target every protein-coding gene in the genome of mammalian cells, to screen for genes resistant to selective pressures such as drug treatment targeting proliferation and DNA repair machinery[86,87]. Both studies were able to successfully identify gene candidates that conferred a selective growth advantage to cells that developed resistance to the therapeutic drugs. Hence, we could apply

this system targeting specific genes and tease out the drivers contributing towards malignancy, growth and proliferation of PGLs.

Little is known about the repertoire of noncoding RNAs that are either enriched or lost in PGLs. Recently, a published study profiled the presence and loss of miRNAs in mutant GISTs compared to pediatric/wildtype GISTs. In adult mutant patients, germline mutations in KIT, PDGFRA and BRAF were observed, whereas in the wildtype/pediatric cases no mutations in known susceptibility genes were reported. The pediatric cases were deficient for the SDH complex by immunohistochemistry, whereas the adult mutant cases had no loss of the SDH complex. The authors observed striking differential expression patterns in clusters of miRNAs present on chromosome 14 between the SDH Deficient pediatric tumors and adult mutant GISTs. In order to understand the mechanism for this observed pattern, they speculated that there is a loss of chromosome 14 in the pediatric GISTs, as seen in adult GISTs with germline mutations in susceptibility genes. Chromosome 14 contains two imprinted genes, MEG3 (maternally expressed) and DLK1 (paternally expressed). Hence, loss of the maternal allele on chromosome 14 correlates with transcriptional silencing of the miRNA cluster in adult mutants. However, they did not observe a deletion in chromosome 14 in pediatric cases. So, another possibility was that there may be an aberrant gain in DNA methylation at the regulatory region controlled the imprinted genes, MEG3 (maternally expressed) and DLK1 (paternally expressed), and presumably, the miRNA cluster. However, they did not observe any changes in DNA methylation by their methylation specific PCR assay[88]. It is entirely possible though that the region they surveyed through their PCR assay is not the regulatory region controlling the transcription of the miRNA cluster. In order to truly eliminate gain of DNA methylation as an epigenetic mechanism controlling expression of these miRNAs, one would have to survey a larger region on chromosome 14 using a technique that allows for deep coverage

of the methylation status of every CpG in the region of interest (such as MiSeq technology). It is also possible that other epigenetic marks, such as repressive histone modifications, mark the regulatory region hence repressing the cluster of miRNAs on chromosome 14. It is unclear if the miRNAs found downregulated in pediatric GISTs, which were unable to assemble the SDH complex, are also found depleted in our PGLs since our transcription data only sequenced RNA species larger than 75 nt. We did not have enough RNA to perform additional sequencing of smaller noncoding RNA species. Hence, more resources are required to profile the repertoire of miRNAs and other noncoding RNA species present in PGLs, followed up by comparative analysis with previously published data in related tumors. Further, it will be important to parse the function of these small noncoding RNAs in providing cells with a selective growth advantage, leading to progression and metastasis. This again can be potentially achieved by targeting candidate miRNAs with the CRISPR-Cas9 system and studying the effects on proliferation, apoptosis and senescence.

Finally, the ultimate goal of understanding the drivers involved in PGL oncogenesis is to develop suitable therapies for these patients and to reduce the risk of metastasis. A powerful solution would be to model patient-derived tumors in animal models, catering to the idea of personalized cancer therapeutics[89]. The rationale for this is based on the Center for Personalized Therapeutics led by Dr. Ross Cagan, where one can sequence patient derived tumors and determine mutations in upto 15 potential tumor drivers. As a starting point, these mutations can be combined in a fly model and targeted to specific tissues, in this case tissues derived from neural crest cell lineage, where cells may mimic patient tumors by exhibiting overgrowth, inhibited cell death and senescence and metastatic-like cell behavior. These patient-matched flies can then be screened in a high-throughput format with drugs/drug combinations for those that suppress tumor growth. Finally,

the patient tumors can be modeled in mammalian systems, where the previously identified drug cocktails can be tested for efficacy and toxicity. Hence, in order to develop and optimize therapeutic targets, we must first identify genetic lesions and epigenetic changes in patient derived PGLs, and we hope that our work provides a good starting point in understanding the complexity of this cancer.

5.5 References

1. Ward WS, Coffey DS (1991) DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells. Biology of Reproduction 44: 569-574.

2. Balhorn R, Reed S, Tanphaichitr N (1988) Aberrant protamine 1/protamine 2 ratios in sperm of infertile human males. Cellular and Molecular Life Sciences 44: 52-55.

3. Balhorn R, Brewer L, Corzett M (2000) DNA condensation by protamine and arginine-rich peptides: analysis of toroid stability using single DNA molecules. Molecular Reproduction and Development 56: 230-234.

4. Gatewood J, Cook G, Balhorn R, Bradbury E, Schmid C (1987) Sequence-specific packaging of DNA in human sperm chromatin. Science 236: 962-964.

5. Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, et al. (2009) Distinctive chromatin in human sperm packages genes for embryo development. Nature 460: 473-478.

6. Brunner AM, Nanni P, Mansuy IM (2014) Epigenetic marking of sperm by post-translational modification of histones and protamines. Epigenetics Chromatin 7: 2.

7. Brykczynska U, Hisano M, Erkek S, Ramos L, Oakeley EJ, et al. (2010) Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nat Struct Mol Biol 17: 679-687.

8. Wu SF, Zhang H, Cairns BR (2011) Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. Genome Research 21: 578-589.

9. Marques CJ, Costa P, Vaz B, Carvalho F, Fernandes S, et al. (2007) Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. Molecular Human Reproduction 14: 67-74.

10. Kobayashi H, Sato A, Otsu E, Hiura H, Tomatsu C, et al. (2007) Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. Human Molecular Genetics 16: 2542-2551.

11. Cummins JM, Jequier AM (1994) Treating male infertility needs more clinical andrology, not less. Human Reproduction 9: 1214-1219.

12. Bowdin S, Allen C, Kirby G, Brueton L, Afnan M, et al. (2007) A survey of assisted reproductive technology births and imprinting disorders. Hum Reprod 22: 3237-3240.

13. Corzett M, Mazrimas J, Balhorn R (2002) Protamine 1: Protamine 2 stoichiometry in the sperm of eutherian mammals. Molecular Reproduction and Development 61: 519-527.

14. Oliva R (2006) Protamines and male infertility. Human Reproduction Update 12: 417-435.

15. Carrell DT, Emery BR, Hammoud S (2007) Altered protamine expression and diminished spermatogenesis: what is the link? Human Reproduction Update 13: 313-327.

16. Marques CJ, Francisco T, Sousa S, Carvalho F, Barros A, et al. (2010) Methylation defects of imprinted genes in human testicular spermatozoa. Fertility and Sterility 94: 585-594.

17. Hammoud SS, Purwar J, Pflueger C, Cairns BR, Carrell DT (2010) Alterations in sperm DNA methylation patterns at imprinted loci in two classes of infertility. Fertility and Sterility 94: 1728-1733.

18. Niemitz EL, DeBaun MR, Fallon J, Murakami K, Kugoh H, et al. (2004) Microdeletion of LIT1 in familial Beckwith-Wiedemann syndrome. American Journal of Human Genetics 75: 844-849.

19. DeBaun MR, Niemitz EL, Feinberg AP (2003) Association of in vitro fertilization with Beckwith-Wiedemann syndrome and epigenetic alterations of LIT1 and H19. American Journal of Human Genetics 72: 156-160.

20. Gosden R, Trasler J, Lucifero D, Faddy M (2003) Rare congenital disorders, imprinted genes, and assisted reproductive technology. Lancet 361: 1975-1977.

21. Killian JK, Kim SY, Miettinen M, Smith C, Merino M, et al. (2013) Succinate dehydrogenase mutation underlies global epigenomic divergence in gastrointestinal stromal tumor. Cancer Discov 3: 648-657.

22. Xiao M, Yang H, Xu W, Ma S, Lin H, et al. (2012) Inhibition of alpha-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. Genes Dev 26: 1326-1338.

23. Letouze E, Martinelli C, Loriot C, Burnichon N, Abermil N, et al. (2013) SDH mutations establish a hypermethylator phenotype in paraganglioma. Cancer Cell 23: 739-752.

24. Kondo Y, Issa J-PJ (2010) DNA methylation profiling in cancer. Expert Reviews in Molecular Medicine 12: e23.

25. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, et al. (2009) The human

colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nature Genetics 41: 178-186.

26. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16: 6-21.

27. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics 13: 484-492.

28. Scott KL, Nogueira C, Heffernan TP, van Doorn R, Dhakal S, et al. (2011) Proinvasion metastasis drivers in early-stage melanoma are oncogenes. Cancer Cell 20: 92-103.

29. Xu W, Yang H, Liu Y, Yang Y, Wang P, et al. (2011) Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. Cancer Cell 19: 17-30.

30. Senchenko VN, Krasnov GS, Dmitriev AA, Kudryavtseva AV, Anedchenko EA, et al. (2011) Differential expression of CHL1 gene during development of major human cancers. PLoS ONE 6: e15612.

31. Kong HK, Yoon S, Park JH (2012) The regulatory mechanism of the LY6K gene expression in human breast cancer cells. Journal of Biological Chemistry 287: 38889-38900.

32. Oshima H, Ishikawa T, Yoshida GJ, Naoi K, Maeda Y, et al. (2013) TNF-alpha/TNFR1 signaling promotes gastric tumorigenesis through induction of Noxo1 and Gna14 in tumor cells. Oncogene 356: 1-10.

33. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, et al. (2012) IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. Nature 483: 479-483.

34. Liu R, Zhou Z, Huang J, Chen C (2011) PMEPA1 promotes androgen receptor-negative prostate cell proliferation through suppressing the Smad3/4-c-Myc-p21 Cip1 signaling pathway. J Pathol 223: 683-694.

35. Trerotola M, Cantanelli P, Guerra E, Tripaldi R, Aloisi AL, et al. (2013) Upregulation of Trop-2 quantitatively stimulates human cancer growth. Oncogene 32: 222-233.

36. Fulda S (2013) Targeting c-FLICE-like inhibitory protein (CFLAR) in cancer. Expert Opin Ther Targets 17: 195-201.

37. Lu C, Ward PS, Kapoor GS, Rohle D, Turcan S, et al. (2012) IDH mutation impairs histone demethylation and results in a block to cell differentiation. Nature 483: 1-7.

38. Baldanzi G, Mitola S, Cutrupi S, Filigheddu N, van Blitterswijk WJ, et al. (2004) Activation of diacylglycerol kinase alpha is required for VEGF-induced angiogenic signaling in vitro. Oncogene 23: 4828-4838.

39. Yanagisawa K, Yasuda S, Kai M, Imai S-i, Yamada K, et al. (2007) Diacylglycerol kinase alpha suppresses tumor necrosis factor-alpha-induced apoptosis of human melanoma cells through NF-kappaB activation. Biochimica et biophysica acta 1771: 462-474.

40. Bacchiocchi R, Baldanzi G, Carbonari D, Capomagi C, Colombo E, et al. (2005) Activation of alpha-diacylglycerol kinase is critical for the mitogenic properties of anaplastic lymphoma kinase. Blood 106: 2175-2182.

41. Carter H, Samayoa J, Hruban RH, Karchin R (2010) Prioritization of driver mutations in pancreatic cancer using cancer-specific high-throughput annotation of somatic mutations (CHASM). Cancer Biol Ther 10: 582-587.

42. Hayes GM, Carrigan PE, Miller LJ (2007) Serine-arginine protein kinase 1 overexpression is associated with tumorigenic imbalance in mitogen-activated protein kinase pathways in breast, colonic, and pancreatic carcinomas. Cancer Research 67: 2072-2080.

43. Hu-Lowe DD, Chen E, Zhang L, Watson KD, Mancuso P, et al. (2011) Targeting activin receptor-like kinase 1 inhibits angiogenesis and tumorigenesis through a mechanism of action complementary to anti-VEGF therapies. Cancer Research 71: 1362-1373.

44. Krneta J, Kroll J, Alves F, Prahst C, Sananbenesi F, et al. (2006) Dissociation of angiogenesis and tumorigenesis in follistatin-and activin-expressing tumors. Cancer Research 66: 5686-5695.

45. Unland R, Schlosser S, Farwick N, Plagemann T, Richter G, et al. (2012) Epigenetic silencing of the dopamine receptor D4 in medulloblastoma. Klinische Pädiatrie 224: A4.

46. Ju JH, Yang W, Lee KM, Oh S, Nam K, et al. (2013) Regulation of cell proliferation and migration by keratin19-induced nuclear import of early growth response-1 in breast cancer cells. Clin Cancer Res 19: 4335-4346.

47. Fukui T, Kondo M, Ito G, Maeda O, Sato N, et al. (2005) Transcriptional silencing of secreted frizzled related protein 1 (SFRP 1) by promoter hypermethylation in non-small-cell lung cancer. Oncogene 24: 6323-6327.

48. Takada T, Yagi Y, Maekita T, Imura M, Nakagawa S, et al. (2004) Methylation-associated silencing of the Wnt antagonist SFRP1 gene in human ovarian cancers. Cancer Science 95: 741-744.

49. Suzuki H, Watkins DN, Jair K-W, Schuebel KE, Markowitz SD, et al. (2004) Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. Nature Genetics 36: 417-422.

50. Shih YL, Hsieh CB, Lai HC, Yan MD, Hsieh TY, et al. (2007) SFRP1 suppressed hepatoma cells growth through Wnt canonical signaling pathway. Int J Cancer 121: 1028-1035.

51. Lo PK, Mehrotra J, D'Costa A, Fackler MJ, Garrett-Mayer E, et al. (2006) Epigenetic suppression of secreted frizzled related protein 1 (SFRP1) expression in human breast cancer. Cancer Biol Ther 5: 281-286.

52. Harris T, Marquez B, Suarez S, Schimenti J (2007) Sperm motility defects and infertility in male mice with a mutation in Nsun7, a member of the Sun domain-containing family of putative RNA methyltransferases. Biology of Reproduction 77: 376-382.

53. Gao Q, Steine EJ, Barrasa MI, Hockemeyer D, Pawlak M, et al. (2011) Deletion of the de novo DNA methyltransferase Dnmt3a promotes lung tumor progression. Proceedings of the National Academy of Sciences 108: 18061-18066.

54. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, et al. (2014) Mutational landscape and significance across 12 major cancer types. Nature 502: 333-339.

55. Werner S, Frey S, Riethdorf S, Schulze C, Alawi M, et al. (2013) Dual roles of the transcription factor grainyhead-like 2 (GRHL2) in breast cancer. Journal of Biological Chemistry 288: 22993-23008.

56. Wang F, Travins J, DeLaBarre B, Penard-Lacronique V, Schalm S, et al. (2013) Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. Science 340: 622-626.

57. Berdasco M, Ropero S, Setien F, Fraga MF, Lapunzina P, et al. (2009) Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. Proceedings of the National Academy of Sciences 106: 21830-21835.

58. Morris MR, Ricketts CJ, Gentle D, McRonald F, Carli N, et al. (2011) Genome-wide methylation analysis identifies epigenetically inactivated candidate tumour suppressor genes in renal cell carcinoma. Oncogene 30: 1390-1401.

59. Tessema M, Yingling CM, Grimes MJ, Thomas CL, Liu Y, et al. (2012) Differential epigenetic regulation of TOX subfamily high mobility group box genes in lung and breast cancers. PLoS One 7: e34850.

60. Smith KT, Sardiu ME, Martin-Brown SA, Seidel C, Mushegian A, et al. (2012) Functional characterization of the human FAM60A protein: a new subunit of the

Sin3 deacetylase complex. Molecular and Cellular Proteomics 11: 1815-1828.

61. Park S, Kim J-H, Jang J-H (2007) Aberrant hypermethylation of the FGFR2 gene in human gastric cancer cell lines. Biochemical and Biophysical Research Communications 357: 1011-1015.

62. Fortier AM, Asselin E, Cadrin M (2013) Keratin 8 and 18 loss in epithelial cancer cells increases collective cell migration and cisplatin sensitivity through claudin1 up-regulation. Journal of Biological Chemistry 288: 11555-11571.

63. Zhou Y-H, Wu X, Tan F, Shi Y-X, Glass T, et al. (2005) PAX6 suppresses growth of human glioblastoma cells. Journal of Neuro-Oncology 71: 223-229.

64. Shyr CR, Tsai MY, Yeh S, Kang HY, Chang YC, et al. (2010) Tumor suppressor PAX6 functions as androgen receptor co-repressor to inhibit prostate cancer growth. Prostate 70: 190-199.

65. Mayes DA, Hu Y, Teng Y, Siegel E, Wu X, et al. (2006) PAX6 suppresses the invasiveness of glioblastoma cells and the expression of the matrix metalloproteinase-2 gene. Cancer Res 66: 9809-9817.

66. Kilinc D, Ozdemir O, Ozdemir S, Korgali E, Koksal B, et al. (2012) Alterations in promoter methylation status of tumor suppressor HIC1, SFRP2, and DAPK1 genes in prostate carcinomas. DNA and Cell Biology 31: 826-832.

67. Veeck J, Noetzel E, Bektas N, Jost E, Hartmann A, et al. (2008) Promoter hypermethylation of the SFRP2 gene is a high-frequent alteration and tumor-specific epigenetic marker in human breast cancer. Molecular Cancer 7: 83.

68. Aleman A, Adrien L, Lopez-Serra L, Cordon-Cardo C, Esteller M, et al. (2007) Identification of DNA hypermethylation of SOX9 in association with bladder cancer progression using CpG microarrays. British Journal of Cancer 98: 466-473.

69. Guryanova OA, Sablina AA, Chumakov PM, Frolova EI (2005) Downregulation of TRIP6 gene expression induces actin cytoskeleton rearrangements in human carcinoma cell lines. Molecular Biology 39: 792-795.

70. Huynh HT, Larsson C, Narod S, Pollak M (1995) Tumor suppressor activity of the gene encoding mammary-derived growth inhibitor. Cancer Research 55: 2225-2231.

71. Song MA, Tiirikainen M, Kwee S, Okimoto G, Yu H, et al. (2013) Elucidating the landscape of aberrant DNA methylation in hepatocellular carcinoma. PLoS One 8: e55761.

72. Tchougounova E, Jiang Y, Bråsäter D, Lindberg N, Kastemar M, et al. (2009) Sox5 can suppress platelet-derived growth factor B-induced glioma development

in Ink4a-deficient mice through induction of acute cellular senescence. Oncogene 28: 1537-1548.

73. Qin YR, Tang H, Xie F, Liu H, Zhu Y, et al. (2011) Characterization of tumor-suppressive function of SOX6 in human esophageal squamous cell carcinoma. Clin Cancer Res 17: 46-55.

74. Subimerb C, Pinlaor S, Lulitanond V, Khuntikeo N, Okada S, et al. (2010) Circulating CD14(+) CD16(+) monocyte levels predict tissue invasive character of cholangiocarcinoma. Clinical and Experimental Immunology 161: 471-479.

75. Lindsey SF, Byrnes DM, Eller MS, Rosa AM, Dabas N, et al. (2013) Potential role of meiosis proteins in melanoma chromosomal instability. J Skin Cancer 2013: 190109.

76. Span PN, Rao JU, Oude Ophuis SBJ, Lenders JWM, Sweep FCGJ, et al. (2011) Overexpression of the natural antisense hypoxia-inducible factor-1alpha transcript is associated with malignant pheochromocytoma/paraganglioma. Endocrine Related Cancer 18: 323-331.

77. Burnichon N, Vescovo L, Amar L, Libe R, de Reynies A, et al. (2011) Integrative genomic analysis reveals somatic mutations in pheochromocytoma and paraganglioma. Human Molecular Genetics 20: 3974-3985.

78. Lopez-Jimenez E, Gomez-Lopez G, Leandro-Garcia LJ, Munoz I, Schiavi F, et al. (2010) Research resource: transcriptional profiling reveals different pseudohypoxic signatures in SDHB and VHL-Related pheochromocytomas. Molecular Endocrinology 24: 2382-2391.

79. De Santa F, Totaro MG, Prosperini E, Notarbartolo S, Testa G, et al. (2007) The histone H3 lysine-27 demethylase JMJD3 links inflammation to inhibition of polycomb-mediated gene silencing. Cell 130: 1083-1094.

80. Margueron R, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. Nature 469: 343-349.

81. Laget S, Joulie M, Le Masson F, Sasai N, Christians E, et al. (2010) The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. PLoS ONE 5: e11982.

82. Du Y, Liu B, Guo F, Xu G, Ding Y, et al. (2012) The essential role of Mbd5 in the regulation of somatic growth and glucose homeostasis in mice. PLoS ONE 7: e47358.

83. Cox BJ, Vollmer M, Tamplin O, Lu M, Biechele S, et al. (2010) Phenotypic annotation of the mouse X chromosome. Genome Res 20: 1154-1164.

84. Kooistra SM, Helin K (2012) Molecular mechanisms and potential functions of histone demethylases. Nature Reviews Molecular Cell Biology 13: 297-311.

85. Dahia PLM (2014) Pheochromocytoma and paraganglioma pathogenesis: learning from genetic heterogeneity. Nature Reviews Cancer 14: 108-119.

86. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, et al. (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343: 84-87.

87. Wang T, Wei JJ, Sabatini DM, Lander ES (2014) Genetic screens in human cells using the CRISPR-Cas9 system. Science 343: 80-84.

88. Kelly L, Bryan K, Kim SY, Janeway KA, Killian JK, et al. (2013) Post-transcriptional dysregulation by miRNAs is implicated in the pathogenesis of gastrointestinal stromal tumor [GIST]. PLoS ONE 8: e64102.

89. Dar AC, Das TK, Shokat KM, Cagan RL (2012) Chemical genetic discovery of targets and anti-targets for cancer polypharmacology. Nature 486: 80-84.

APPENDIX A

TO TEST IF SUCCINATE ACCUMULATION IS NECCESSARY AND SUFFICIENT FOR CHANGES IN EPIGENOME OF SDH DEFICIENT PGLS

A.1 Experimental Design

1) To test for sufficiency of succinate accumulation:

We treated Rencells and NCCs (controls) wit SDH inhibitor, TTFA.

2) To test for necessity of succinate accumulation: We would overexpress the SDHB gene in a PGL cell.

For both scenarios, we would:

1) Test for accumulation of histone modifications: H3K4me3, H3K27me3, H3K9me3

2) Validate loss/accumulation of succinate levels.

A.2 Results and Discussion

We treated Rencells with 50 uM TTFA (Thenoyltrifluoroacetone). TTFA is a noncompetitive inhibitor that inhibits the ubiquinol binding site in SDHD. Electron influx (produced by the oxidation of succinate to fumarate) is inhibited at SDHD[1]. Hence, it is entirely possible that succinate is still oxidized to fumarate in these cells, albeit at a slower rate. Previous studies studied the effect of TTFA on Hep3B, HT1080 and PC12 cells and saw an increase in bulk histone modifications, H3K9me3, H3K27me3 and H3K36me3 [2]. Hence, we were interested in pursuing the effect of TTFA in neural cell types, Rencells and NCCs (neural crest cells).

Although RenCells are not neural crest cell derived, they are a neural progenitor cell line with the ability to differentiate into neurons and glial cells. These were isolated from the cortical region of the human fetal brain and immortalized by retroviral transduction with the c-myc gene. Also, they were commerially available and could be maintained in culture. This was in contrast to NCCs which can only be maintained in culture for 3 days and each time must be freshly differentiated from ES cells [3].

So, we treated RenCells with 50 uM TTFA and NCCs with 25 uM TTFA and

probed for bulk changes in histone modifications. We observed bulk accumulation of H3K4me3, H3K9me3 and H3K27me3 in TTFA treated RenCells (Figure A.1). Due to limited material from NCCs, we only probed for H3K27me3 and overall did not see an accumulation in bulk levels (Figure A.2). The difference in results could possibly be explained by the fact that histone accumulation by TTFA can be variable according to cell type. We then performed a DNA methylation analysis on the TTFA treated RenCells using the 450K Illumina methylation array. Unfortunately, we saw no difference in DNA methylation upon comparing untreated RenCells to TTFA treated RenCells (data not shown). Changes in DNA methylation are usually observed over several passages [4] depending on cell type and hence we speculate that perhaps the TTFA treatment to the cells was not long enough to elicit a change in DNA methylation response. It is also possible that in this cell type, the epigenome is impacted mostly on a histone level – where modifications accumulate and mislocalize in the genome and illicit aberrant changes in transcription.

While the results from the treated RenCells appear to be partially promising, we were concerned with the fact that the DNA methylome of untreated RenCells was significantly different than that of NCCs. This could be due to the fact that they both are different progenitor cells where RenCells are a more differentiated cell type in comparison to NCCs. Also, RenCells have been immortalized and this can have an impact on their methylome [5]. Finally, we argued that since PGLs are derived from neural crest cell lineage, NCCs would represent a better progenitor cell for comparisons of DNA methylation changes and transcription changes to PGL tumors. Hence, we did not perform follow up experiments on TTFA treated RenCells and eliminated the use of untreated cells for DNA methylation comparisons.

Currently, there are no published PGL cell lines available to the scientific community. PGLs are slow growing tumors and hence establishing tumor derived cell lines proves to be difficult since they may not have the right growth conditions

available in culture. We were able to establish two cell lines from patient derived tumors. However, like most primary cell lines that are not immortalized, our cells senesced after remaining in culture for seven passages. This made it difficult to overexpress the SDHB gene in them to test for necessity of succinate accumulation for impacts on the epigenome. Nonetheless, we characterized these cell lines by staining for a marker for glomus cells, synaptophysin. We also stained these cells for the presence of the SDH complex. Both cell lines stained positive for synaptophysin and positive for the SDH complex. Hence, the SDH complex was intact in both cell lines. To test for SDH activity, we performed metabolomics analysis using GC-MS and probed for levels of succinate and fumarate. As a positive control, we included HEK293T cells that were stably transduced with a lentiviral construct expressing either an shRNA against luciferase or an shRNA against SDHB. Figure A.3 shows >80% loss of the SDHB protein in SDHB knockdown HEK293T cells compared to those transduced with the shRNA against luciferase. Another positive control we included in our analysis was a fumarate hydratase deficient (FH^{-}/FH^{-}) cell line derived from a hereditary leiomyomatosis renal cell carcinoma [6]. We found 10-fold succinate accumulation in SDH Deficient HEK293T cells and 30-fold fumarate accumulation in the FH Deficient cells. However, we did not see any accumulation of succinate in either of the PGL cell lines, regardless of normoxic or hypoxic growth conditions (Figure A.4). Finally, we also surveyed bulk levels of histone modifications for the PGL cell lines compared to RenCells, HEK293T cells and FH Deficient cells. We were unable to see bulk accumulation of histone modifications for our western blots (Figure A.5 and A.6).

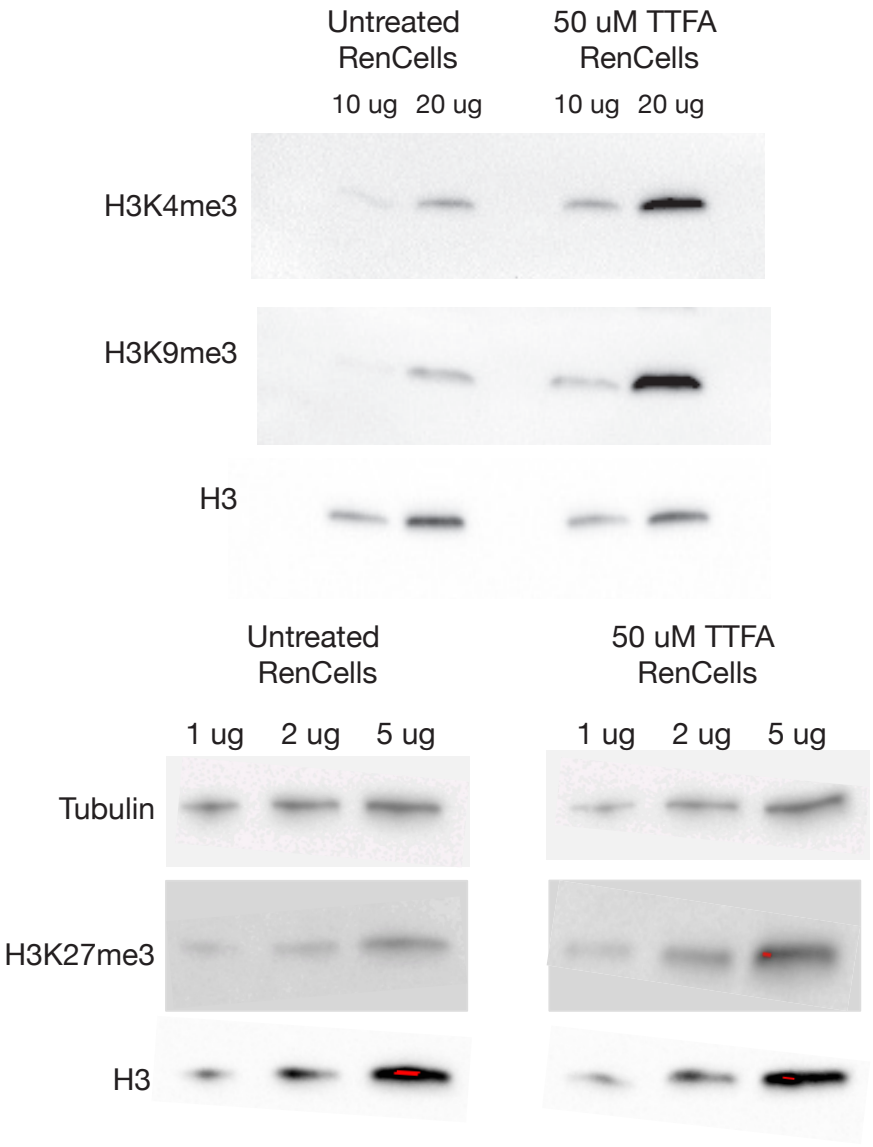


Figure A.1: Western blots probing for bulk levels of H3K4me3, H3K9me3 and H3K27me3 for untreated and treated RenCells (with 50 uM TTFA). Protein levels indicated is total protein levels.

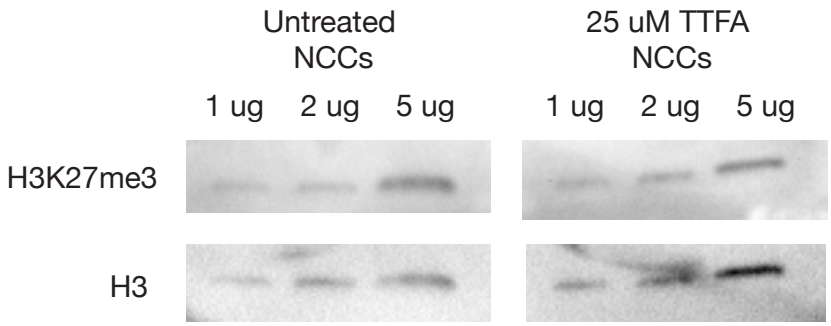


Figure A.2: Western blot probing for bulk levels of H3K27me3 for untreated and treated NCCs (with 25 uM TTFA). Protein levels indicated is total protein levels.

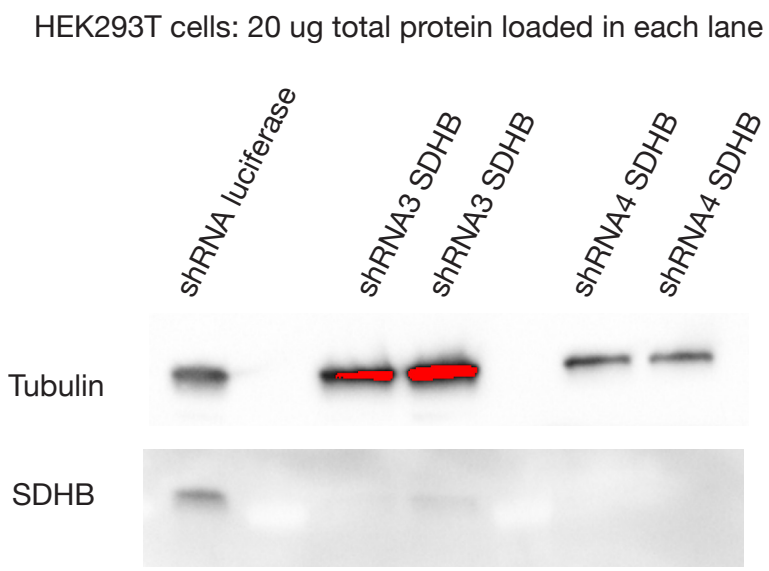


Figure A.3: Western blots showing knockdown of SDHB gene in HEK293T cells using two different shRNAs.

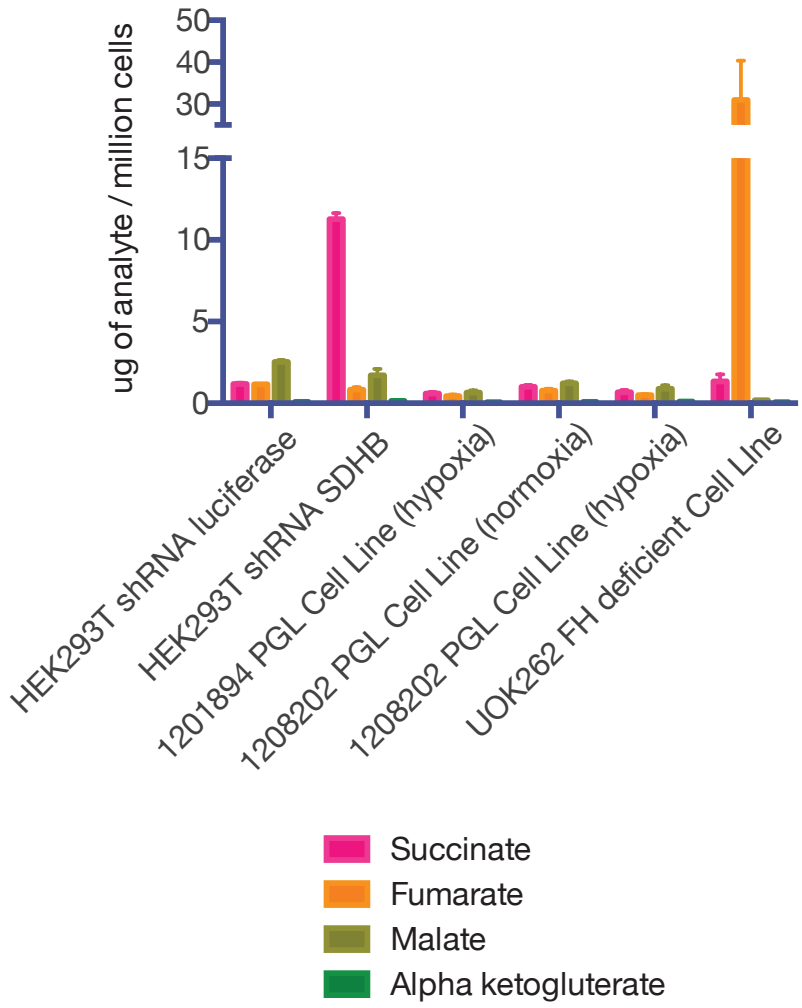


Figure A.4: Metabolomic analysis for succinate, fumarate, malate and alpha-ketoglutarate in control cell lines (HEKs transduced with either shRNA against luciferase or SDHB), PGL cell lines 1201894 and 1208202 (grown in both normoxc and hypoxic conditions) and FH Deficient renal leiyomatosis derived cell line.

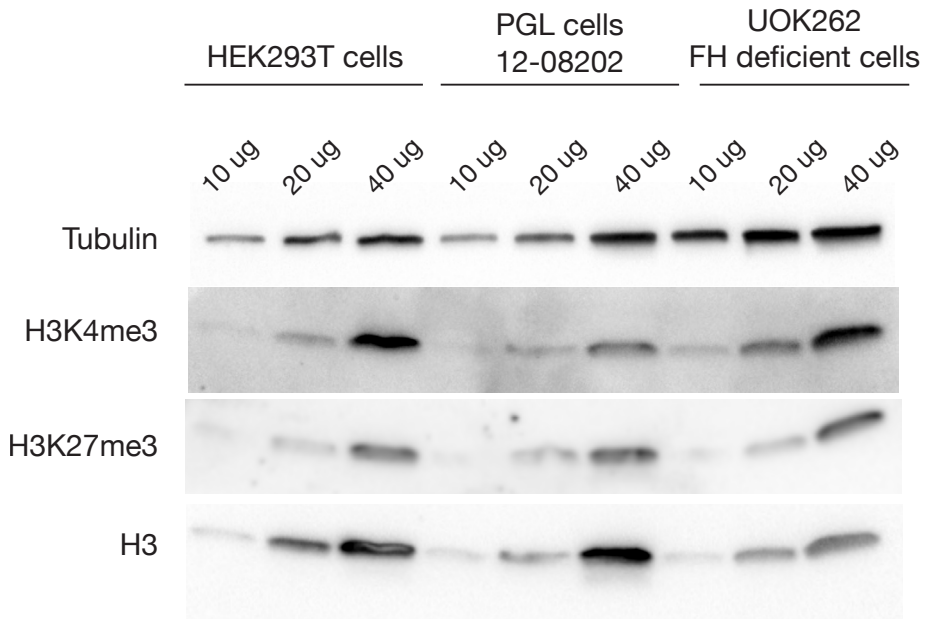


Figure A.5: Western blot showing bulk levels of H3K4me3, H3K27me3, histone H3 and tubulin in HEK293T cells (control), PGL cells 12-08202 and FH Deficient UOK262 cells. Protein levels indicated are for total protein.

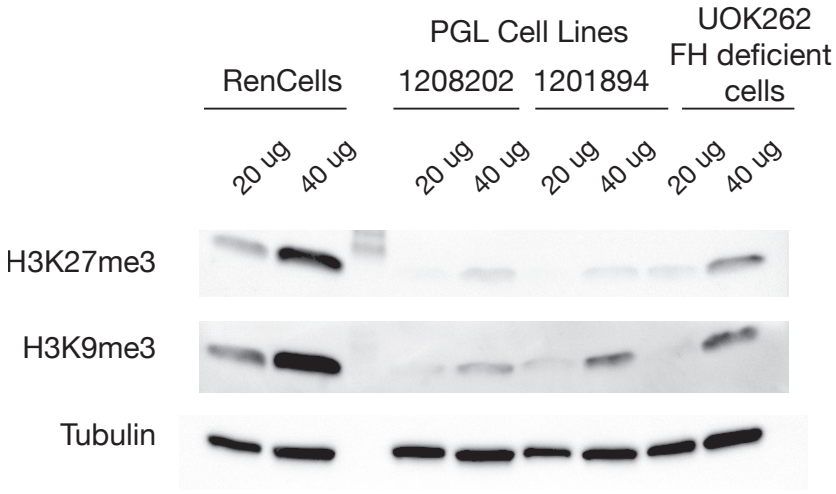


Figure A.6: Western blot showing bulk levels of H3K9me3, H3K27me3, and tubulin in RenCells (control), PGL cells 12-01894 and 12-08202 and FH Deficient UOK262 cells. Protein levels indicated are for total protein.

A.3 References

1. Guzy RD, Sharma B, Bell E, Chandel NS, Schumacker PT (2008) Loss of the SdhB, but not the SdhA, subunit of complex II triggers reactive oxygen species-dependent hypoxia-inducible factor activation and tumorigenesis. Molecular and Cellular Biology 28: 718-731.

2. Cervera AM, Bayley J-P, Devilee P, McCreath KJ (2009) Inhibition of succinate dehydrogenase dysregulates histone modification in mammalian cells. Molecular Cancer 8: 89.

3. Bajpai R, Chen DA, Rada-Iglesias A, Zhang J, Xiong Y, et al. (2010) CHD7 cooperates with PBAF to control multipotent neural crest formation. Nature 463: 958-962.

4. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, et al. (2012) IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. Nature 483: 479-483.

5. Grafodatskaya D, Choufani S, Ferreira JC, Butcher DT, Lou Y, et al. (2010) EBV transformation and cell culturing destabilizes DNA methylation in human lymphoblastoid cell lines. Genomics 95: 73-83.

6. Yang Y, Valera VA, Padilla-Nash HM, Sourbier C, Vocke CD, et al. (2010) UOK 262 cell line, fumarate hydratase deficient (FH-/FH-) hereditary leiomyomatosis renal cell carcinoma: in vitro and in vivo model of an aberrant energy metabolic pathway in human cancer. Cancer Genetics and Cytogenetics 196: 45-55.

APPENDIX B

METABOLOMIC ANALYSIS OF ALPHA-KETOGLUTARATE, SUCCINATE, FUMARATE AND MALATE USING GC-MS IN PARAGANGLIOMAS

B.1 Experimental procedure

James Cox developed this method in the Metabolomics Core, University of Utah.

B.1.1 Samples Included in the Analysis

We performed the analysis on a total of eight paragangliomas. In the first run, we submitted five tumors out of which two tumors stained deficient for the SDH complex and three tumors stained present for the SDH complex by immunohistochemistry. In the second run, we submitted three tumors out of which one tumor stained deficient for the SDH complex, one stained present and one tumor whose complex assembly status has not been determined by staining yet.

B.1.2 Sample Extraction from Tissue

Tissues are extracted by first placing weighed, snap-frozen tissues into precooled 2 mL homogenization tubes containing ceramic beads (1.4 mm). Next, a volume of ice-cold MeOH (100%) containing 1 µg of the internal standard D_4-succinate is added to give a final concentration of 80% MeOH (assuming tissue density is 1 g/mL) and the tissue is homogenized with the Omni Bead Ruptor 24 in one, 30 second cycle. The samples are put in the freezer for 1 hour, centrifuged at 14,000 g for 10 minutes at 4 °C, and then three aliquots of 100 µL of the upper phase are collected, transferred to an Eppendorf tube and dried under reduced pressure. Once dry, 40 µL of methoxyamine hydrochloride (40 mg/mL in pyridine) is added and heated at 40 °C for 1 hour followed by the addition of 40 µL of MSTFA and then heated at 40 °C for 1 hour. The derivatized sample is then transferred into a GC/MS vial with insert for analysis.

B.1.3 Calibration Curve

Analytical standards are prepared by creating a stock solution of 20 µg/mL (10 µg/tube) followed by a serial dilution (1:1) to generate 10 samples. The internal standard D_4-succinate is added to each dilution (1 uL of 1 mg/mL). The dilution series is then dried under reduced pressure and derivatized as described above. Each standard was run in triplicate and the resulting area under the curve was averaged.

B.1.4 GC-MS Analysis

GC/MS analyses are conducted using an HP6890 instrument interfaced with an MSD-HP5973 detector and equipped with a Zebron ZB-5MSi Guardian (30 m x 0.25 mm ID, 0.25 µm film thickness; Phenomenex) column and an HP7682 injector. Helium is used as a carrier gas at a flow rate of 13.8 mL/min with a 10:1 split ratio at an injection volume of 1 µL. The injector temperature is 250 °C. The oven temperature gradient was programmed as follows: 95 °C held for 1.5 minute increased at a rate of 40 °C/minute to 118 °C, held for 1 minute, increased at a rate of 5 °C/minute to 250 °C, increased at a rate of 25 °C/minute to 330 °C and held for 12.3 minute. MS spectra are obtained in EI mode from a range of *m/z* 244 – 305. The MS quad temperature is 150 °C, MS source temperature is 230 °C, solvent cut time of 4 minute, and scanned at 16 scans/second.

B.2 Data Analysis

Data are first collected on an Agilent MSD Chemstation, translated using an Agilent GC MSD translator, then analyzed using Agilent MassHunter Quant. The resulting area under the curve is processed using Microsoft Excel.

The raw area for each analyte (succinate, fumarate, malate, and

α-ketoglutarate) was averaged then normalized based on the response factor of the internal standard. Masses are the averaged results from three aliquots taken from the extractions. The measurement of succinate from the tissue samples extends just beyond the largest dilution standard. Measurements of α-ketoglutarate had the most variability.

B.3 Results and Discussion

From our first run, succinate did not accumulate in the three tumors that stained positive for the SDH complex via IHC. Interestingly, we observed succinate accumulation in only one of the two tumors that had no SDH complex assembly (Figure B1). We speculate that this is due to the fact that one of the tumors had not been collected in the correct manner. This includes that the tumor be snap-frozen immediately after being surgically removed from the patient. It is possible that the tumor that stained negative for the SDH complex was either not snap-frozen fast enough and hence metabolite levels were affected in that tumor. It is also possible that the SDH complex in that tumor does indeed assemble and the staining result is negative. Again, this could be due to poor handling of the tumor causing degradation of proteins.

From our second run, we observed succinate accumulation in the tumor that lacked SDH complex assembly as well as the tumor with the unknown SDH assembly status. As expected, we did not see succinate accumulation in the tumor that stained positive for the SDH complex (Figure B.2).

Together, from our results, we conclude that staining for the SDH complex via IHC agrees well with the metabolite data for succinate accumulation, granted that the tumor is snap-frozen immediately after being surgically removed from the patient.

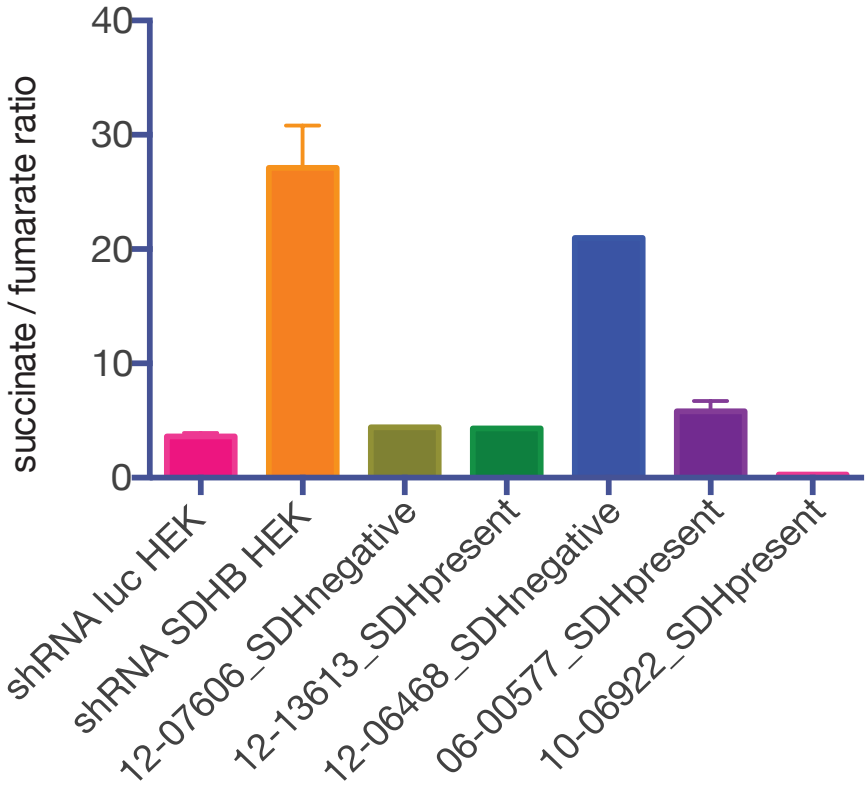


Figure B.1:Succinate to fumarate ratio for HEK cell lines transduced with either a shRNA against luciferase or SDHB (controls), SDH negative PGLs (stained deficient / absent for the SDH complex by IHC) and SDH Present PGLs (stained present / positive for the SDH complex by IHC).

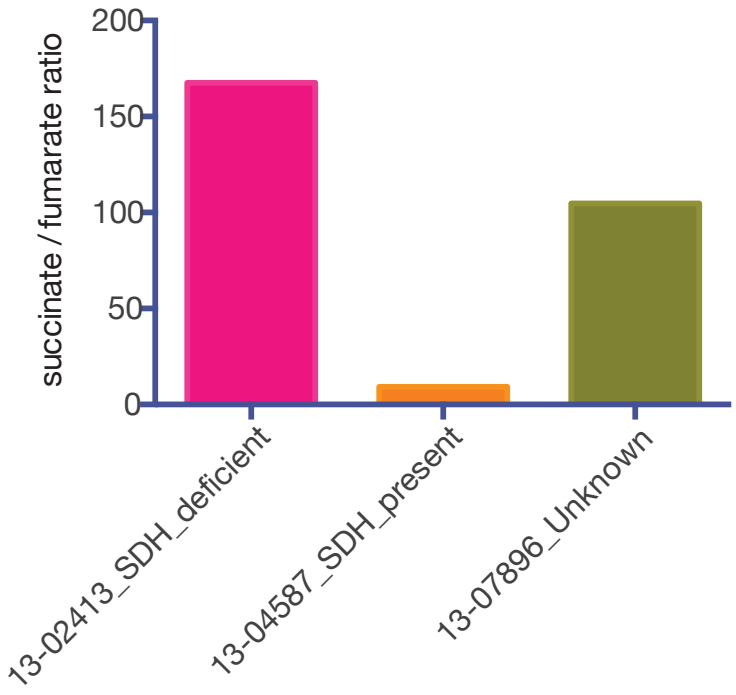


Figure B.2:Succinate to fumarate ratio additional PGL tumors that were either SDH Deficient PGLs (stained deficient / absent for the SDH complex by IHC), SDH Present PGLs (stained present / positive for the SDH complex by IHC) or staining status of PGL was unknown.