Summer 8-15-2018

Methods to Record Transcription Factor Binding and Enhancer Activity throughout Cellular Differentiation

Zongtai Qi
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Molecular Genetics and Genomics

Dissertation Examination Committee:
Robi Mitra, Chair
Barak Cohen
Joseph Dougherty
Joshua Rubin
Gray Stormo
Andrew Yoo

Methods to Record Transcription Factor Binding and Enhancer Activity throughout Cellular Differentiation.

by
Zongtai Qi

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

August 2018
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

The pursuit of a PhD Degree is fruitless without people who provide support, mentorship, patience and guidance on the multiple facets of research and of life. First, I am incredibly indebted to my mentor, Rob Mitra. I am grateful to work under his mentorship. His trust in my abilities expands my potentials to do research and encourages me to ask tough questions in science. He taught me to think critically and creatively especially in the difficult times and instilled in me a sense of scientific rigor and curiosity. Rob sets up a good example of scientist whose creativity, energy and vision never cease to inspire me. I also recognize how fortunate I am to have a mentor who supports my life outside of lab. I am grateful that I could share with him my excitement of marriage and birth of my son.

I would like to sincerely thank my committee: Dr. Andrew Yoo, Dr. Joseph Dougherty, Dr. Barak Cohen, Dr. Gray Stormo and Dr. Joshua Rubin. Your feedbacks were crucial in the development of my dissertation. I appreciated your patience and support, especially through the difficult times. Your mentorship has taught me how to be successful in academia and I look forward to building upon the immense knowledge I have gained from you. I also would like to thank members of the Genetics Department especially Jim Skeath for their continued support throughout the entire graduate study.

I want to thank all the members of the Rob Mitra lab, past and present. I remember all their help such as brainstorming and troubleshooting. I am indebted to Xuhua Chen who helps a lot on my projects in cloning and cell culture experiments. Christian Shively, Matt Lalli and Denis Avey are great colleagues and are always willing to provide positive feedbacks. Especially, Matt always impressed and inspired me by his thoughts and ideas. I thank Justin Melendez, Jiayue Liu, Arnav Moudgil and Michael Wilkinson for providing their perspectives and help on my

research. Sumithra Sankararaman, Francesco Vallania, David Mayhew, Maximiliaan Schillebeeckx, Scott Hidgon, Tom Cohen, Gabe Bien-Willner, Yun Yue and Haoyi Wang are great old friends who shared with me so many memories of how Mitra lab has "evolved".

I want to specially thank Najla Kfoury from Josh Rubin's lab who has been a tremendous help and provided plenty of support on our collaborated project. I was lucky enough to work with her. She is a great friend in both work and life. I thank Nisha Iyer from Shelly Sakiyama-Elbert's lab who kindly engineered cell lines for my research, which was a pivotal step to start a project.

Family has been a constant source of support and inspiration. I would like to thank my parents, Guocai and Lanyin, for your continuous support and love. They have always allowed me to pursue my passions in life and have taught me that hard work is the key to success. I appreciate all the sacrifices each of them have made over the years. I also want to thank my wife Yinjiao and my son Jimmy. Yinjiao helped me through the ups and downs of research and of life. Her patience and understanding have brought me to where I am now. Jimmy is a source of pure happiness and "pushes" me to work even harder. They mean too much to my life and they are the reason that I am able to be stronger to overcome significant and difficult times.

Zongtai Qi

Washington University in St. Louis

August 2018

ABSTRACT OF THE DISSERTATION

Methods to Record Transcription Factor Binding and Enhancer Activity throughout Cellular Differentiation.

by

Zongtai Qi

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular Genetics and Genomics

Washington University in St. Louis, 2018

Professor Andrew Yoo, Chair

The ability to create distinct cell types is fundamental for the development of multicellular organisms. Since all cells in an organism contain the same genes, cellular diversity is achieved through the transcriptional network where transcription factors (TFs) interacts with cis-regulatory elements, leading to the selective transcription of different sets of genes. To better understand the functions of TFs and regulatory elements underlying cell fate decisions, we developed methods that are able to record their activities throughout cellular differentiation. In Chapter 2, we developed a degradation domain based induction system for "Calling Cards" method which maps the binding sites of TFs using *piggybac* transposons. The induction "Calling Cards" method offers an alternative to chromatin immunoprecipitation (ChIP) methods and furthermore has the ability to record TF binding at different time periods of the development. In Chapter 3, we applied the "Calling Card" method to study the role of master regulatory Brd4-bound enhancers for sex differences in glioblastoma. We revealed a set of sex-specific regulatory genes and networks, which are indicative of sex-specific transcriptional programs regulated by

Brd4-bound enhancers. Finally, to record the activity of regulatory elements or enhancers, in Chapter 4, we developed a CRE recombinase-mediated method for high-throughput functional identification of active enhancers at different time periods of the development, named as Developmental Enhancer Sequencing (DevEn-seq). We demonstrated that DevEn-seq is able to detect enhancers more efficiently than regular reporter methods and trace enhancer activities throughout cellular differentiation without being disturbed by the gene silencing effect caused by lentiviral sequences. With an *in vitro* neural differentiation protocol, we identified two neural progenitor-specific enhancers near HB9 and Olig2 genes respectively. In summary, this dissertation contributes to the field of developmental biology by providing useful methods for recording TF binding events and enhancer activities throughout development.

Chapter 1
Introduction

1.1 Development of an inducible transposon "Calling Card" method

1.1.1 The transposon "Calling Card" method can record transcription factor binding.

Vertebrate development is transcriptionally regulated, thus considerable effort has been expended on understanding the gene expression networks that control this process[1]. However, mapping these transcriptional networks has proven difficult. Methods like ChIP-Chip or ChIP-Seq can analyze transcription factor binding in homogenous populations of cells but are unable to record transcription factor binding events along different cellular lineages. This make it impossible to correlate transcription factor binding events in progenitor cells to the final fates of their progeny, complicating efforts to understand the mechanisms by which cell fates are determined.

We have developed transposon "Calling Cards", a method designed to record transcription factor binding during development[2]. The central idea behind the Calling Card method is to attach the transposase of a transposon to a transcription factor (TF), thereby bestowing on the TF the ability to direct insertion of the transposon into the genome near to where it binds. The transposon is a "Calling Card" that permanently marks the transcription factor's visit to a particular genomic location. By harvesting the transposon Calling Cards along with their flanking genomic DNA, it is possible to obtain a genome-wide map of transcription factor binding.

1.1.2 An "inducible" Calling Card method is useful in studying TF bindings underlying cell fate decisions.

In developmental biology, the decision of into what terminal cell fate a progenitor cell differentiates is a critical step in creating the architecture of multicellular organisms. These choices are often directed by some key TFs known as master regulators of gene regulation[3, 4]. These TFs may display disparate functions at different times during neural development[5, 6]. To use transposon "Calling Card" method to understand these key TFs' bindings underlying cell fate decisions in development, it is important to control the timing of TF-directed transposition events to those early in the differentiation process rather than the mature states. To achieve this goal, we need to create an inducible *piggybac* (PB) transposon system on which the CC method is built so that we are able to turn on and off the transposition events for TF recording. An inducible PB transposon system can be made at two molecular levels: transcriptional and post-translational levels. Because we want the transposon system to react promptly to an inducer, the molecular control at post-translational level is more favored.

1.1.3 Degradation domains allow a rapid post-translational induction system.

Molecular tools that regulate protein stability at a post-translational level in live cells include protein degron systems induced by auxin, light, or destabilizing domains (DDs)[7-10]. Here, we focus on the DD-based degron system, which involves genetically fusing the protein of interest to a small unstable protein domain. This DD fusion protein is recognized by the cellular protein quality control machinery, which then degrades the entire fusion protein. However, in the presence of a DD-specific small molecule, the DD assumes a folded state and becomes stable, allowing the target protein to perform its regular biochemical function[11].

Two versions of DD induction systems were developed that control target protein levels in a rapid, reversible and tunable fashion[9, 10]. The first engineered DD is based on a human FK506-rapamycin-binding protein (FKBP) with 107 amino acid residues. Point mutations in FKBP (F36V and L106P) confer instability to fusion partners, which can be rescued by the cell-permeable high-affinity small molecule, Shield-1[9]. The second DD is engineered from an *Escherichia coli* dihydrofolate reductase (DHFR) protein with 159 amino acid residues. Similarly, a few key mutations confer instability to DHFR, which can be rescued by a highly permeable small molecule, Trimethoprim (TMP)[10]. These two induction systems have been demonstrated to function in a variety of contexts, including mammalian cell cultures, live mice, viral infections, and in pathogens like Plasmodium and Toxoplasma[9, 10 12-15]. However, both FKBP and DHFR DDs display high levels of background in the absence of any molecule indcuer[15]. This basal level of expression is sometimes sufficient for the target fusion protein to perform its regular biochemical function and thereby precludes the observation of any loss-of-function phenotype[15]. The double architecture for DDs has recently been shown to minimize the background expression levels[15], but a systematic exploration should further expand the uses of these powerful induction systems.

1.1.4 Degradation domains are utilized to develop an inducible transposon "Calling Card" method.

We sought to develop a PB transposon induction system for "Calling Card" method that provides a tightly regulated transposase enzyme that displays a large difference in activity between the induced and un-induced state, that, when induced, deposits transposons with an efficiency equal to that of the native protein, that is highly active across a wide variety of cell

types, and that is induced with a small molecule that has minimal effects on general cellular functions. To do so, we tagged PB transposase with two different destabilized domain (DD) proteins, FK506- and rapamycin binding protein (FKBP)[9] and dihydrofolate reductase (DHFR)[10], and compared these to the ERT2 based induction system[16] in four different cell lines. To identify the optimal transposon induction system for "Calling Card" method, we systematically characterized, in four cell lines, 15 different PB transposase fusion proteins representing three different induction systems. We found that the FKBP-based system achieved the broadest dynamic range of induction across four cell lines. Remarkably, in the presence of chemical inducer, this system had transposition efficiencies that were almost as high (~95%) as the "wild-type" PB transposon. We also investigated the effects of the inducer molecules on differentiation and found that Shld1 does not affect ESC development. Taken together, the FKBP-based PB transposon induction system is selected for "Calling Cards" method. This work was described in chapter 2.

1.2 Application of the "Calling Card" method to study the role of Brd4-bound enhancers for sex differences in glioblastoma

1.2.1 Sex-dimorphism in glioblastoma is a cell-intrinsic property.

The sexual dimorphism observed in glioblastoma (GBM) is thought to be driven largely by cell-intrinsic factors. This conclusion is supported by two lines of evidence. First, pediatric brain cancers display the same sexual dimorphism in incidence as is observed in adult brain cancers. In these instances, the patients are pre-pubescent, so males and females have similar hormonal profiles[17, 18]. Second, the Rubin lab has recapitulated the observed sex-disparity in a cell culture model of GBM. In this system, male NF1 null astrocytes expressing a dominant negative form

of p53 (NF1-/-;DNp53) exhibit increased tumorigenesis and cell proliferation relative to female NF1-/-;DNp53 astrocytes, independent of hormonal exposure[19]. In other words, when the male transgenic astrocytes are implanted into female hosts, the increase in tumor size and growth is still greater than female transgenic astrocytes implanted into male hosts. These two lines of evidence support the hypothesis that the sexual dimorphism in GBM occurrence is a cell intrinsic property.

1.2.2 Epigenetic regulatory mechanism may play a key role in establishing sex differences in GBM.

Since sex differences in GBM are cell intrinsic, epigenetic regulatory mechanisms may be a key mediator of these differences[20, 21]. Evidence of sex specific epigenetic effects in the brain can be found in the bed nucleus of the stria terminalis (BNST), which is known to have sex-dependent differences in structure[22]. Histone deacetylase (HDAC) inhibitors applied to the BNST of male mice result in female-like volume and cell number, and suggests that histone marks play a role in sexually dimorphic cell growth and apoptosis[23]. Though epigenetic regulation of cancer states have been studied[24, 25], their role in GBM sex-disparity is unknown.

1.2.3 Epigenetic reader Brd4 establishes cell identity and regulates oncogenes.

One mediator of cell identity and disease state is the epigenetic reader BRD4, a protein from the BET family of chromatin regulators that bind acetylated histones via its bromodomains[26]. BRD4 activity is abolished when these domains are blocked by drugs such as JQ1[27]. BRD4 binding to acetylated histones mediates recruitment of transcriptional activators and promotes gene expression independent of common transcriptional recruiters such as

pTEFb[26], making it a key regular in gene transcription[28]. BRD4 is also known to be one of the few chromatin regulators that are present during cell division[29], which is consistent with its established role in promoting cellular proliferation. Interestingly, BRD4 has also been shown to acetylate H3K122, a mark which destabilizes the nucleosome and is correlated with active transcription of surrounding genes[30]. This correlates with BRD4's known role in maintaining cell proliferation in rapidly differentiating cells[31], and may be a key mechanism for maintaining the proliferative "stem-like" cancer cells.

BRD4 also has been shown to promote tumorigenesis in multiple myeloma (MM) cell line where BRD4 regulates the expression of specific oncogenes such as c-MYC. JQ1 treatment of MM cell lines results in a global decrease in BRD4 enhancer binding[32], and a subsequent decrease in gene expression. Global decrease in BRD4 binding results in a decrease in oncogene expression in human GBM cell lines as well, and demonstrates BRD4's role as a regulator of oncogenic potential[32].

1.2.4 The "Calling Card" method is used to decipher the Brd4-bound enhancers.

BRD4s regulation of oncogenic potential gives insight into the role of BRD4 in GBM. In addition, we have experimentally demonstrated that JQ1 has sex-specific effects on clonogenicity and tumorigenesis, in which males exhibit a decrease in these traits while females exhibit no effect or an increase in these traits. RVX208, a drug with a higher BRD2 affinity than a BRD4 affinity, does not exhibit such drastic effects. Therefore, we hypothesized that BRD4 is a mediator of the increased clonogenicity in males and used transposon "Calling Cards" method to identify the BRD4-bound enhancers in both male and female GMB cells.

Integration of Brd4 Calling Cards, H3K27ac ChIP-seq and RNA-seq data in GBM cells revealed a set of sex-specific regulatory genes and networks. Male-specific JQ1-sensitive targetable genes demonstrated functional enrichment for neoplasm metastasis, tumor angiogenesis, integrin signaling pathway, and DNA-repair-deficiency disorders. Female-specific JQ1-sensitive targetable genes showed an enrichment in pathways involved in semaphorin signaling, chromosome aberrations, glioblastoma, regulation of transcription, and glucose metabolism disorders. These results are indicative of sex-specific transcriptional programs regulated by Brd4-bound enhancers. This work is described in chapter 3.

1.3 Development of a high-throughput functional identification of enhancers using a CRE recombinase mediated reporter assay

1.3.1 Enhancer and super enhancer play a fundamental role in cellular differentiation.

Enhancers are a class of noncoding regulatory DNA sequences that can activate transcription independently of their location, distance or orientation with respect to the promoters of the genes they control[33, 34]. Their activity is frequently tissue-specific or cell type-specific[35-37]. Hundreds of thousands of candidate enhancers have been annotated in the human genome so far, compromising at least 12% of the total DNA sequences[38, 39]. A majority of these sequences display chromatin accessibility or characteristic patterns of histone modifications in a cell type-specific fashion, supporting their roles in mediating cell type-specific gene expression programs. Variations at enhancer sites, either single-nucleotide variants or copy-number variants, are known to be directly associated with a variety of human diseases[40, 41]. In mice, individual deletions of enhancers have been shown to considerably alter development[42-45]. Further, the

enhancer sequences are enriched in transcription factor-binding motifs, providing clues to the lineage-specific transcription factors that work at these sequences[46].

Among the large number of candidate enhancers predicted in the human and mouse genomes, a small fraction of them drew special attention. These sequences, referred to as super-enhancers, are dense clusters of enhancers and are found in virtually every mammalian cell and tissue type examined[47]. They are bound by an extraordinary amount of transcription factors and cofactors and are typically associated with high levels of transcription activities in specific cell types. Found particularly enriched at cell identity genes, super-enhancers are regarded as the key control elements of lineage specification.

1.3.2 Current methods have challenges for high-throughput identification of cell-specific enhancers in development.

DNase I hypersensitivity assay led to the discovery of numerous mammalian enhancers, such as the beta-globin locus control region (LCR)[48] critical for the developmental control of beta-globin genes in mammals[34, 49]. This approach has undergone several iterations from the laborious Southern blot analysis[49] to the high-throughput microarray [50] and the ultrahigh-throughput DNA-sequencing analysis that has been now broadly applied to hundreds of cell types[46, 51]. The latest iteration of this assay is the use of a genetically engineered TN5 transpose, instead of DNase I, in a highly streamlined experimental protocol called ATAC-seq (assay for transposase-accessible chromatin with high-throughput sequencing)[52]. ATAC-seq is especially amendable to very small cell numbers and even single cells[53, 54]. However, DNase I hypersensitivity is not unique to enhancers. Other cis regulatory elements such as promoters and

insulators also show this property[55]. Therefore, just the presence of DNase I hypersensitivity signal alone is not sufficient to distinguish these different classes of cis regulatory elements.

With the development of ChIP-seq to map transcription factor's binding sites and chromatin modification patterns in the genome[56, 57], it was possible to show that promoters, enhancers, and insulators are differentially occupied by transcription factors and are associated with distinct chromatin modification patterns[58, 59]. This allowed genome-wide determination of candidate enhancers. For instance, binding sites of p300, a coactivator protein and histone acetyltransferase that acts at enhancers, led to annotation of several thousand enhancers in embryonic brain and limbs in the mouse[60]. Mapping of DNA occupancy of transcription factors such as Sox2, Oct4, and Nanog also provided the first enhancer map in the human embryonic stem cells[61, 62].With the increased understanding of the role of histone modification processes in nucleosome dynamics and chromatin organization, it was natural to examine the chromatin modification patterns in the genome since different classes of cis regulatory elements may recruit different chromatin remodeling proteins that generate distinct patterns of histone modifications. This hypothesis was supported by the observation that enhancers are associated with high levels of H3K4me1 and low levels of H3K4me3, while the opposite is true for promoters[63]. The finding led to the development of chromatin signature-based strategy for mapping enhancers, and identification of such DNA in both mouse and human genomes[64, 65].

Despite these major advances in identifying enhancers, a number of important questions remain. What are the biological functions of the annotated enhancers? How do they contribute to cell type-specific gene expression programs? The lack of comprehensive, functionally validated enhancer data sets for most tissues and cell types has prohibited the systematic exploration of these questions.

1.3.3 The CRE-mediated reporter assay allows a high-throughput functional identification of cell specific enhancers in development.

To overcome the challenges mentioned above, we developed a CRE recombinase-mediated method for the functional identification of active enhancers at different time periods of the development, named as Developmental Enhancer Sequencing (DevEn-seq). DNA test fragments containing putative enhancers are linked to an inducible CRE recombinase and then are integrated into genome by lentiviruses. If the enhancer is active in a cell, CRE recombinase will be expressed and activate RFP reporter and permanently generate red cells in itself and its progeny. ES cell libraries can be differentiated *in vitro* before sorting and sequencing. We demonstrated that DevEn-seq is able to detect enhancers more efficiently than regular reporter methods and record enhancer activities throughout cellular differentiation without being disturbed by gene silencing effect caused by lentiviral sequences. Coupled with high-throughput sequencing, DevEn-seq enables a high throughput method for functional identification and tracking of enhancer activities. This work is described in chapter 4.

Chapter 2
An optimized, broadly applicable *piggyBac* transposon induction system

Published in Nucleic Acids Research. (2017)

2.1 Abstract

The piggyBac (PB) transposon has been used in a number of biological applications. The insertion of PB transposons into the genome can disrupt genes or regulatory regions, impacting cellular function, so for many experiments it is important that PB transposition is tightly controlled. Here, we systematically characterize three methods for the post-translational control of the PB transposon in four cell lines. We investigated fusions of the PB transposase with ERT2 and two degradation domains (FKBP-DD, DHFR-DD), in multiple orientations, and determined (i) the fold-induction achieved, (ii) the absolute transposition efficiency of the activated construct and (iii) the effects of two inducer molecules on cellular transcription and function. We found that the FKBP-DD confers the PB transposase with a higher transposition activity and better dynamic range than can be achieved with the other systems. In addition, we found that the FKBP-DD regulates transposon activity in a reversible and dose-dependent manner. Finally, we showed that Shld1, the chemical inducer of FKBP-DD, does not interfere with stem cell differentiation, whereas tamoxifen has significant effects. We believe the FKBP-based PB transposon induction will be useful for transposon-mediated genome engineering, insertional mutagenesis and the genome-wide mapping of transcription factor binding.

2.2 Introduction

Transposons are genetic elements that are able to mobilize throughout a host genome autonomously. Due to their mobility, DNA transposons and retrotransposons have been widely used as tools for generating mutation libraries and for delivering non-viral gene constructs into the genomes of a variety of organisms [66]. The use of DNA transposons in mammalian genomes, however, was initially impeded by a lack of active transposable elements. The synthetic resurrection of *Sleeping Beauty* (SB), a Tc1/mariner-like transposable element of the salmanoid fish genome, initiated the development of transposon technologies for use in mammalian cells [67]. Since then, several other transposon systems have been developed [68-70], including the *piggyBac* (PB) system, derived from the cabbage looper moth *Trichoplusia ni* [71]. This system is widely used because the PB transposon consistently exhibits high transposition efficiencies in different cell lines and organisms [72], can transpose cargos of up to 100 kb without a significant reduction in efficiency [73], mobilizes without leaving footprint mutations at the excision site, and is amenable to molecular engineering [16]. In addition, the PB system is highly efficient for germ line insertional mutagenesis in mice and lacks overproduction inhibition [74]. These unique characteristics make the *piggyBac* transposon an invaluable tool for a wide range of applications, including stable gene delivery [75], transgene excision [76], insertional mutagenesis [77], and the mapping of transcription factor (TF) binding in eukaryotic genomes [2, 78, 79]. For many of these applications, the PB transposon is engineered to act as a gene trap that will disrupt genes when inserted into the genome [77], so it is useful to be able to tightly control PB transposition so that insertion only occurs during the proper experimental window, and not, for example, during the propagation of cell lines or mice. For other applications, such as the mapping of TF binding to genomic DNA by PB transposition [2, 78, 79], the constitutive activity of the PBase prevents the

determination of the precise timing of the binding events, a feature that is particularly useful when studying developmental processes. Therefore, a method for induction that can tightly regulate the temporal activity of the PB transposon system is desirable. Furthermore, post-translational control is appealing because protein activity can be switched on or off considerably faster than is possible with transcriptional control schemes [9, 10].

Prior to the work presented here, the only existing inducible PB transposon system that operated post-translationally utilized the mutated ligand-binding domain of the estrogen receptor (ERT2) [16]. The PBase-ERT2 fusion is constitutively expressed but is sequestered outside of the nucleus by the cytoplasmic heat shock protein 90 (HSP90), which binds the ERT2 domain (Fig 2.1A). In the presence of a small molecule inducer (4-Hydroxy tamoxifen; 4OHT), HSP90 binding is abolished and the PBase-ERT2 fusion rapidly relocates to the nucleus where it directs transposition. The PBase-ERT2 fusion works well in several contexts [14], but our early experiences with this system revealed that dynamic range of induction varies widely between different cell types, possibly due to differences in HSP90 levels, which is not constant across cell types [80-83]. We further found that ERT2-PBase fusions are not highly active when induced (relative to the unfused transposase, see results). Finally, we found that the chemical inducer 4OHT inhibited the *in vitro* differentiation of mouse embryonic stem cells, an observation consistent with similar reports that 4OHT adversely affects neurogenesis [84], myelinogenesis [84], myometrial differentiation [85], and sexual maturation [86].

In light of these observations, we sought to develop a PB transposon induction system that provides a tightly regulated transposase enzyme that displays a large difference in activity between the induced and un-induced state, that, when induced, deposits transposons with an efficiency equal to that of the native protein, that is highly active across a wide variety of cell

types, and that is induced with a small molecule that has minimal effects on general cellular functions. To do so, we characterized and optimized fusions of the PB transposase with two different destabilized domain (DD) proteins, FK506- and rapamycin binding protein (FKBP) [9] and dihydrofolate reductase (DHFR) [10], and compared these to the ERT2 based induction system [16] in four different cell lines. Destabilized domains are small, inherently unstable proteins that bind small molecules and have been destabilized by mutation. When fused to PBase, the mutated domain unfolds and the fusion protein is degraded by the proteosome (Fig 2.1B). However, if the cognate small molecule ligand is provided, the DD is stabilized and the PBase fusion protein is rescued from degradation and transposase activity is restored in a rapid, dose-dependent, and reversible manner. Unlike the ERT2-based system where induction is mediated by a specific cytoplasmic protein HSP90, DD degradation is mediated by the ubiquitin-proteasome pathway, a common protein degradation mechanism that is active in all mammalian cells [11, 87].

To identify the optimal configuration for each of the three induction systems, we investigated 15 different fusions of the PB transposase with FKBP, DHFR or ERT2, and measured the fold-induction of activity between the uninduced and induced states, the maximum transposition rate of the activated constructs as compared to the unfused *piggyBac* transposase, and the effects of two of the inducer molecules on cellular transcription and differentiation.

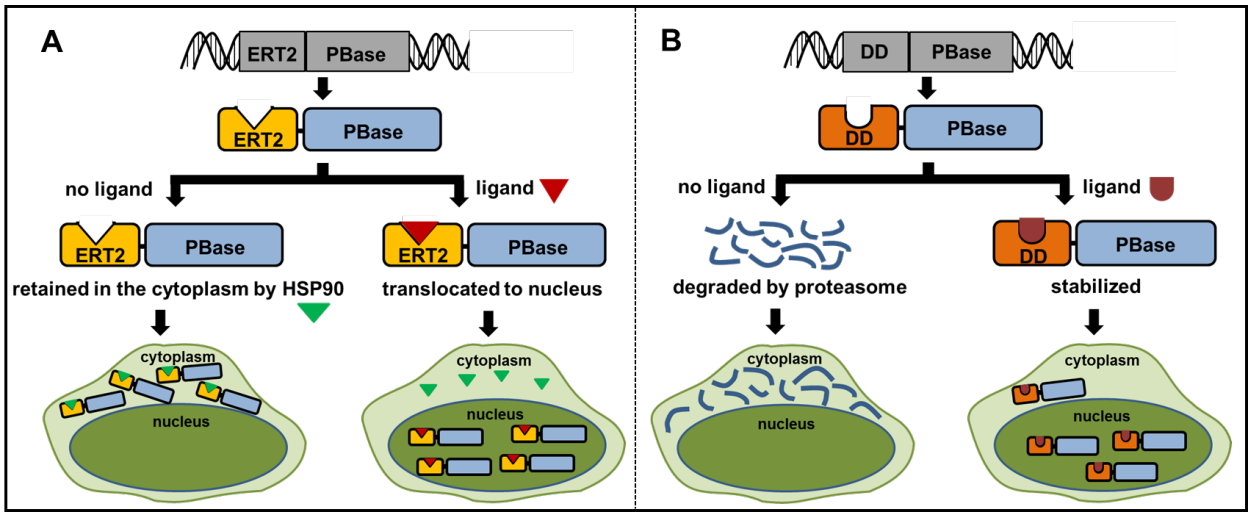


Figure 2.1 Schematic illustration of the ERT2 and DD based PB transposon induction systems. (A) For the ERT2-based PB transposon induction system, the PBase-ERT2 fusion is constitutively expressed but sequestered outside of the nucleus by HSP90, which binds the ERT2 domain. In the presence of ER antagonist 4-OHT, HSP90 dissociates and the PBase-ERT2 fusion rapidly relocates to the nucleus where it directs transposition. (B) For the DD-based PB transposon induction system, a DD (either FKBP or DHFR) was fused to the PBase. The DD confers the instability to the fused protein such that the PBase fusion protein was constitutively degraded. However, binding of a small molecule ligand (Shld1 for FKBP; TMP for DHFR) to the DD prevents PBase from degradation and stabilizes it.

2.3 Results

2.3.1 Overview

We evaluated and optimized three induction systems for their abilities to provide tight post-translation control over the PB transposon. We used one degradation domain (DD) derived from human FKBP12 (FKBP) [9] and another from *Escherichia coli* dihydrofolate reductase

(DHFR) [10] and compared these against the ERT2-based system in four different cell lines: human embryonic kidney cells (HEK293 and HEK293T), human colorectal cancer cells (HCT116), and RW4 mouse embryonic stem cells (ESCs). To obtain the most efficient induction possible, a number of different fusions constructs were investigated for each system (Fig 2.2). We fused DD or ERT2 induction domains in-frame at the N terminus, C terminus, or at both termini of the PBase. For N terminal fusion proteins, an 18 amino acid linker sequence was included between the PBase and the induction domain [88]. For the C terminal fusion protein, a 24 amino acid linker was added between the PBase and induction domain [16].

In order to evaluate these different fusion proteins, we needed a simple, robust way to measure PB transposition. We adopted the commonly used method of co-transfecting cells with two plasmids: a 'helper' plasmid expressing the PBase and a 'donor' plasmid carrying the PB transposon with a drug-resistance marker [2]. Once transfected into the host cells, the PBase mobilizes the transposon from the donor plasmid and integrates it into the host genome, conferring the cell with drug resistance. The number of drug-resistant colonies is closely correlated with the number of transposon integration events and is a key indicator of the efficiency of the PB transposon system [2, 16, 89]. We used this system to measure, for each post-translational control system, the absolute transposition efficiency of the activated construct relative to the unfused *piggyBac* transposase, and the fold-induction achieved by each method in four different cell lines. We then went on to optimize the most promising system, characterize the dynamics of the temporal control achieved by the optimal system and evaluate the effects of two of the inducer molecules on cellular transcription and function.

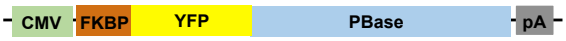
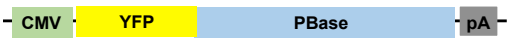
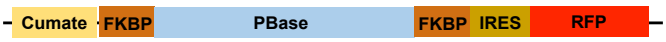
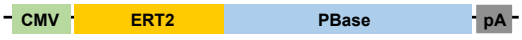
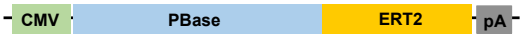
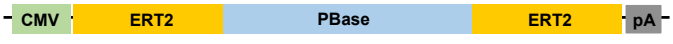
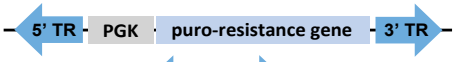
Category	Construct ID and name	Construct structure
FKBP(DD)-based helper constructs	1. FKBP-PBase	
	2. PBase-FKBP	
	3. FKBP-PBase-FKBP	
	4. NLS-FKBP-PBase	
	5. FKBP(2)-PBase	
	6. FKBP(2)-PBase-FKBP	
	7. FKBP-YFP-PBase	
	8. YFP-PBase	
	9. Cumate>iPBase	
DHFR(DD)-based helper constructs	10. DHFR-PBase	
	11. PBase-DHFR	
	12. DHFR-PBase-DHFR	
ERT2-based helper constructs	13. ERT2-PBase	
	14. PBase-ERT2	
	15. ERT2-PBase-ERT2	
Donor constructs	16. 5'LTR-puro-3'LTR	
	17. PBSplitGFP	

Figure 2.2 Schematic illustration of the constructs used in this study. For the helper constructs: PBase, PiggyBac transposase; CMV, cytomegalovirus promoter; pA, bovine growth hormone polyA signal. FKBP, a destabilizing domain derived from FK506 and rapamycin binding protein. DHFR, a destabilizing domain derived from Escherichia coli dihydrofolate reductase; YFP, yellow fluorescent protein; ERT2, mutated ligand-binding domain of the estrogen receptor; Cumate, cumate operator (mutated cmv promoter). For the donor construct: 5'TR-puro-3'TR contains the minimal PB terminal repeats (5'LR and 3'TR) flanking a PGK promoter driven puromycin resistance cassette; EF1, elongation factor 1 promoter; GFP, green fluorescent protein.

2.3.2 Evaluating the transposition efficiencies of inducible PBase

We first measured the transposition activity of the induced PBase fusion proteins relative to that of unfused PBase. The transposition activity for each variant of helper constructs was obtained by transfecting the above-mentioned four cell lines with an equimolar ratio of donor and helper plasmids in the presence of chemical inducers (Shld1 for FKBP fusions; TMP for DHFR fusions; 4OHT for ERT2 fusions). Typical images of stained colonies from negative control, inducer-treated and non-treated co-transfected experiments, and positive control with unfused PBase are shown in Fig 2.3A. In our assays, the DD-based induction systems had significantly higher activities than the ERT2-based induction system in all cell lines. The highest PBase activity was observed for the constructs of FKBP-PBase and DHFR-PBase, both of which had activities that were ~95% of "wild-type" levels. In contrast, all of the ERT2 fusion proteins displayed activities less than 25% that of the unfused PBase. Thus, we conclude that, under inducing conditions, the FKBP and DHRF DDs have little effect on PBase transposition activity while ERT2 appears to significantly interfere with the PBase activity. For the two DD-based induction systems, the transposition activities between N and C terminally tagged PBase fusions are not significantly different ($p>0.05$). However, we observed a significant decrease of PBase activity when both termini were attached to either DD, suggesting that at least one terminus should be exposed for near-optimal PBase transposition activity.

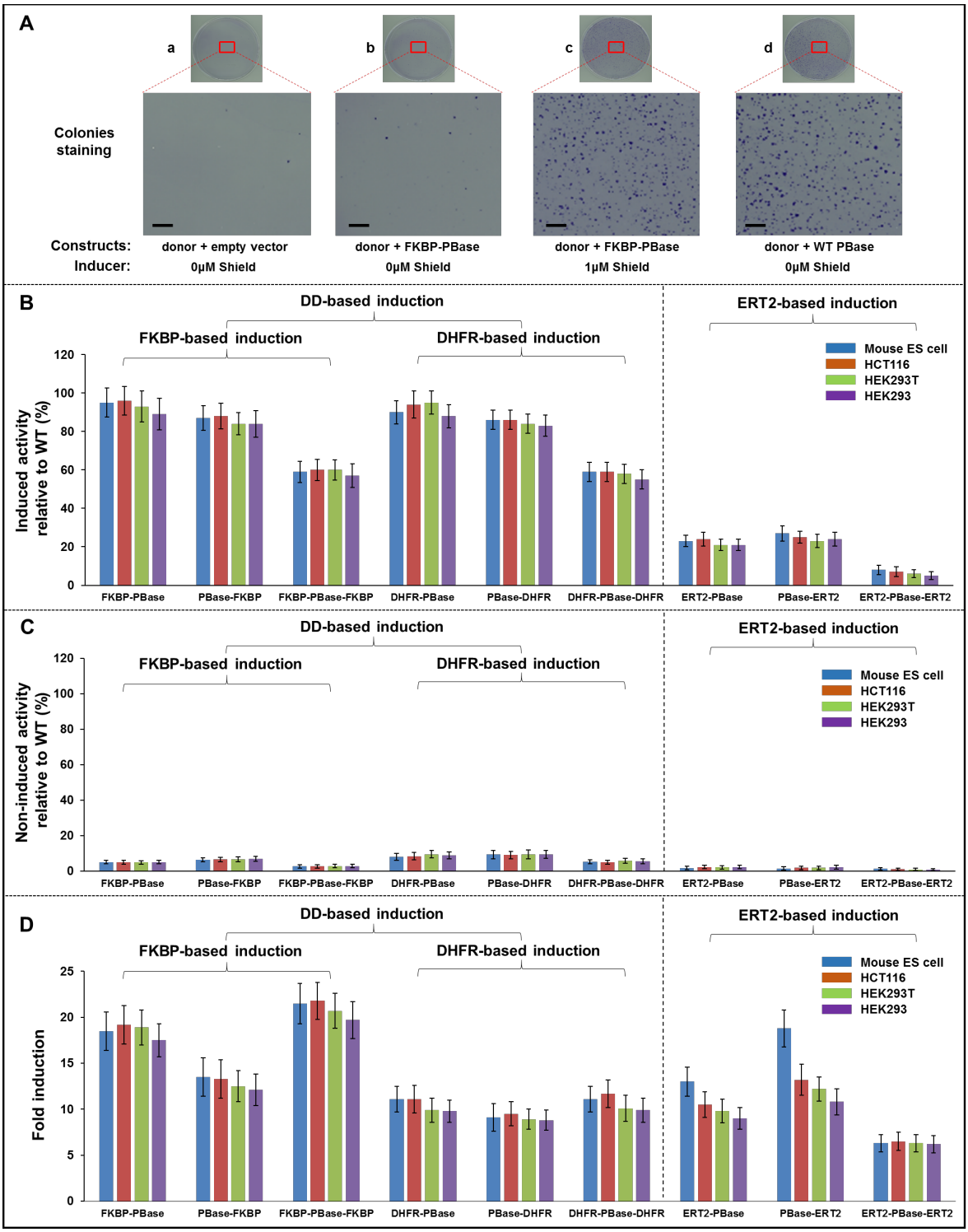


Figure 2.3 The performance of different helper constructs in DD and ERT2 based PB transposon induction systems. (A) Typical images of colony forming and staining assays to evaluate the transposition efficiency. The scale bar in the enlarged image equals 0.1mM. The

construct used is FKBP-PBase. a) Cells transfected with donor and helper backbone plasmids (yeast shuttle vector pRS314) were used to estimate the background or random insertions. b) and c) Cells transfected with FKBP-PBase and donor plasmids either in the absence or presence of Shld1 were used to evaluate the transposition efficiency. d) Cells transfected with unfused PBase and donor plasmid were used to estimate the maximum transposition efficiency. (B) Induced PBase activity of different PBase fusions relative to wild-type PBase across four cell lines. The number of puromycin-resistant colonies formed from the cells transfected with both donor and helper plasmids was normalized to that with donor and helper backbone plasmids prior to any further calculations (background deduction). The induced transposition activity of an inducible domain (i.e. FKBP, DHFR or ERT2) fused PBase was calculated as the normalized number of colonies from the PBase fusion divided by that from "wild type" unfused PBase. Experiments were done in triplicates. (C) Non-induced PBase activity of different PBase fusions across four cell lines. Experimental conditions were the same as in B except that no drug was added for non-induced samples. (D) Fold induction of different PBase fusions across four cell lines. The induction fold was calculated as the normalized number of colonies from chemical inducer treated samples in B divided by that from untreated samples in C.

2.3.3 The FKBP DD system achieved the highest fold-change between induced and uninduced states

We next sought to quantify how tightly the different induction systems regulated the PBase protein. To do so, we again transfected the four cell lines with donor and helper plasmids; however, in this assay, we omitted the small-molecule inducers. We found that the DHFR-based PB induction system had a much higher background than did the FKBP and ERT2-based ones

(Fig 2.3C). To determine which PBase fusion protein provided the largest dynamic range, we used the corresponding induced and non-induced transposition rates to calculate the fold-induction for each protein (Fig 2.3D). This analysis revealed that the FKBP-based induction system displayed a significantly higher fold-induction than both the DHFR and ERT2-based systems in all cell lines ($p<0.05$). The FKBP-PBase-FKBP fusion protein showed the highest fold-induction while the ERT2-PBase-ERT2 fusion protein showed the lowest. Taking fold-induction and the absolute activity of induced PBase into consideration, the FKBP-based inducible PB transposon system outperformed its DHFR and ERT2-based counterparts.

2.3.4 HSP90 levels correlate with the dynamic range of the ERT2 system

The dynamic ranges of the ERT2 controlled PBase fusion proteins were highly variable across the different cell lines (Fig 2.3D). For example, the PBase-ERT2 fusion displayed a 12-fold change in transposase activity between induced and uninduced conditions in mouse RW4 embryonic stem cell, but a significantly lower fold-induction in the other cell lines ($p<0.05$). One possible explanation for this observation is that the HSP90 protein is typically highly expressed in mouse ES cells. HSP90 is the molecule that sequesters the fusion protein in the absence of inducer, and it has been previously reported the HSP90 RNA [90] and protein [91] levels are significantly higher in mouse ES cells than in differentiated cell lines or embryoid bodies. To confirm HSP90 is indeed highly expressed in our mESCs, we performed RNA sequencing on the RW4 and HCT116 lines. We found that the beta-actin normalized HSP90 expression level in the RW4 mouse ES cell line was 4-fold higher than that in HCT116 cell line, which is consistent with the theory that HSP90 levels explain the differences observed between RW4 mouse ESCs and HCT116 cells.

2.3.5 There is a trade-off between maximal PBase activity and fold-induction

In the FKBP-based induction system, the PBase fusion protein shuttles between the cytoplasm where FKBP-directed degradation occurs and nucleus where PB transposition occurs. Therefore, we hypothesized that this nucleo-cytoplasmic transport could be fine-tuned to create an induction system with both efficient transposition and rapid degradation. To try to further improve transposition efficiency, we fused the nuclear localization sequence (NLS) from the SV40 large T-antigen upstream of, and in-frame with, the FKBP-PBase fusion, which had the highest transposition activity in our previous tests (Fig 2.3B). We reasoned that the NLS should direct more fusion protein to the nucleus, increasing the transposition rate upon administration of Shld1. We used transient transfection to introduce an equimolar ratio of NLS-FKBP-PBase fusion-carrying plasmid and donor plasmid across four different cell lines, and compared the results to that of FKBP-PBase plasmid without the NLS. After induction, the engineered fusion protein (NLS-FKBP-PBase) displayed an increase in transposition activity relative to the FKBP-PBase fusion protein (Fig 2.4A, $p<0.05$). However, we also observed higher background transposition in the uninduced sample for this construct. As a result, the fold-induction for the NLS-FKBP-PBase fusion was lower than that of FKBP-PBase (Fig 2.4B, $p<0.05$). This suggests that the NLS sequence sequesters some protein in the nucleus and prevents efficient degradation in the absence of Shld1. Based on this observation, we concluded that adding NLS to PBase fusion protein is not optimal for an inducible PB transposon system.

To test the possibility of further reducing the background transposition that occurs in the absence of chemical inducer, we made two more fusion constructs. One contains two tandem FKBP domains at the N terminus of PBase (Fig 2.2, construct 5) since our previous results showed that an N-terminally tagged PBase was more tightly regulated than the corresponding C-

terminally tagged PBase. The second construct utilized three FKBP domains: two in tandem on the N-terminus and one on the C-terminus (Fig 2.2, construct 6). We tested the transposition efficiency and dynamic range of these constructs as before, and found that the triple FKBPs fusion protein (FKBP-FKBP-PBase-FKBP) showed the highest fold-induction in all cell lines (Fig 2.4A, right panel), followed by the double FKBPs fusion proteins (FKBP-PBase-FKBP and FKBP-FKBP-PBase). All three constructs achieved a significantly higher fold induction than did FKBP-PBase ($p < 0.05$).

We next examined the maximal induced transposition efficiency of these constructs relative to the unfused PBase (Fig 2.4B, right panel). We found that the transposition efficiencies of FKBP-FKBP-PBase-FKBP, FKBP-FKBP-PBase, and FKBP-PBase-FKBP were 60%, 79% and 80% of the unfused PBase, which is significantly lower than the 95% relative activity we observed for FKBP-PBase ($p<0.05$), indicating that transposition activity gradually decreased when more FKBP domains were added to PBase. Taken together, these results suggest that there is a trade-off between the efficient degradation of the PBase fusions under non-inducing conditions and the enzymatic activity of these fusions under fully inducing conditions (Fig 2.4C). Specifically, if high activity upon induction is the objective, then the FKBP-PBase fusion is optimal, while if a high fold-change in induction is required, the FKBP-FKBP-PBase-FKBP fusion is the best choice.

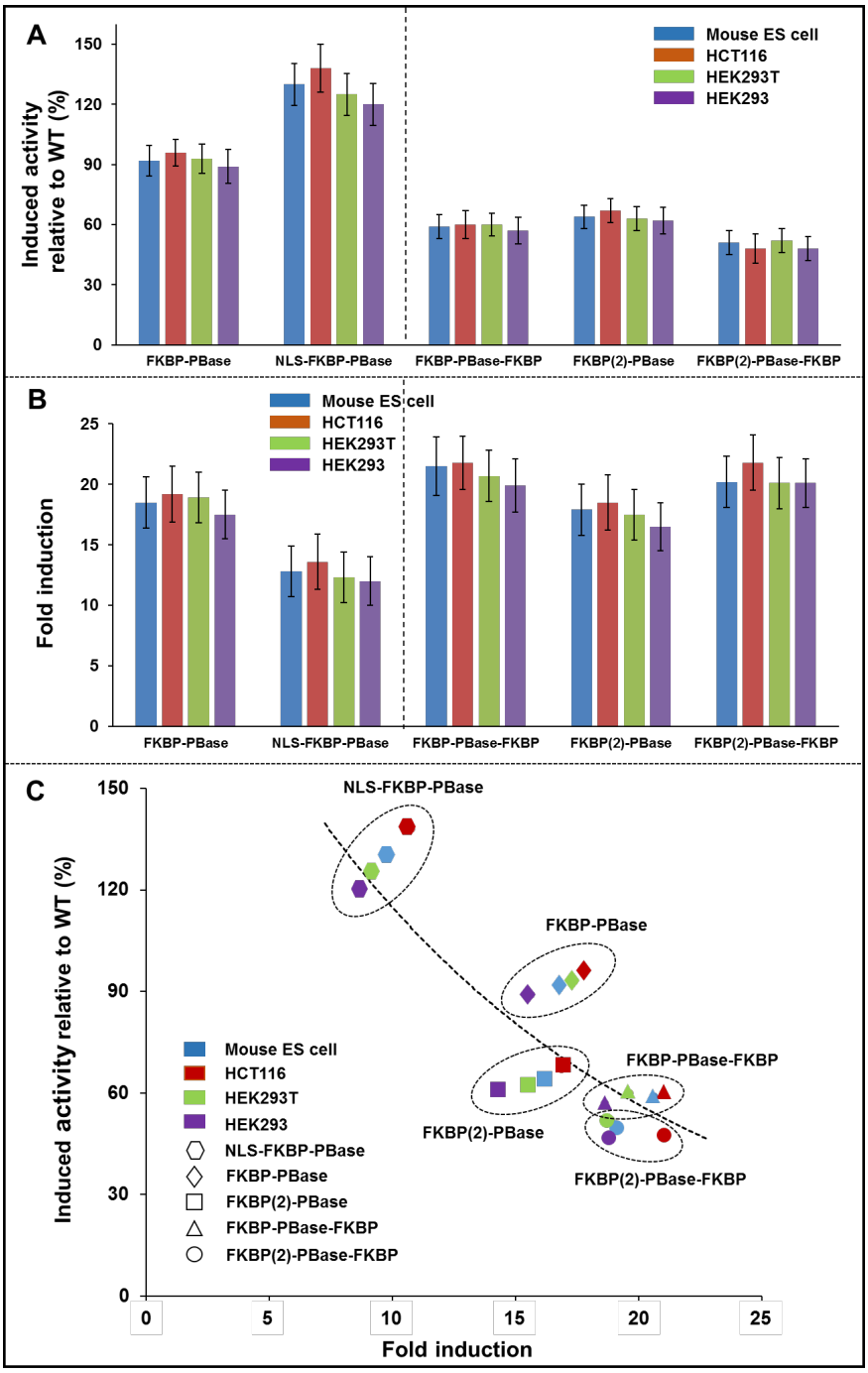


Figure 2.4 Optimization of the induced transposition activity and fold induction for FKBP-based PB transposon induction system. (A) Induced transposition activity of different PBase fusion proteins relative to wild-type PBase across four cell lines. Experiments were done in triplicates. The induced transposition activity of a PBase fusion protein was calculated as the

normalized number of colonies from the PBase fusion in the presence of corresponding chemical inducer divided by that from "wild type" unfused PBase. (B) Fold induction of different PBase fusion proteins across four cell lines. The induction fold was calculated as the normalized number of colonies from chemical inducer treated samples divided by that from untreated samples. (C) Comparison of the induced transposition activity with the fold induction for different PBase fusion proteins across four cell lines.

2.3.6 Fusion protein levels can be tuned to achieve low background transposition and high inducibility

For some application, it is not important to obtain the maximum possible PBase activity, but it is instead preferable to have an assay with low background levels and to achieve a large fold-change in activity upon induction. We reasoned that the high background levels observed in Figs. 2.3 and 2.4 might be due to the fact that the PBase-DD fusion proteins were all highly expressed and could be overloading the proteasome. To test this hypothesis, we created lentivirus containing an FKBP-PBase-FKBP-IRES-RFP fusion gene under the control of a cumate inducible promoter [92] (Fig 2.2, construct 9), which allows gene expression to be tuned by varying the cumate concentration in the culture media. Next, we infected cymR expressing HEK 293 cells, and isolated transduced cells by performing FACS to purify RFP positive cells. We transfected these cells with donor plasmid (Fig 2.2, construct 16), titrated FKBP-PBase-FKBP protein levels by culturing the cell with varying amounts of cumate, in the presence or absence of Shld1, and determined PBase activity by counting puromycin resistant colonies as previously described. The results are shown in Fig 2.5. As lower concentrations of cumate were added to the media, much lower background and larger fold-changes in induction were achieved. For

example, when 10 ug/ml cumate was added to the media, no background transposition was observed; therefore, we observed a more than 100-fold change in PBase activity is achieved after induction with Shld1. These results demonstrate that, when the FKBP-PBase-FKBP protein is expressed at moderate levels, this system has essentially no background transposition yet is still robustly inducible. This level of control comes at some cost, however, as the maximum level of PBase activity is roughly one-fourth the maximum rate we observed (i.e. compare the 10 ug/ml cumate and Shld1 culture condition to the PBase activity of cells grown with 70 ug/ml cumate and Shld1). For experiments that do not require rapid, post-transcriptional induction, it is possible to have the best of both worlds, namely zero background and high absolute activity, by growing cells in the absence of cumate and Shld1 and then adding both chemicals to induce transposition.

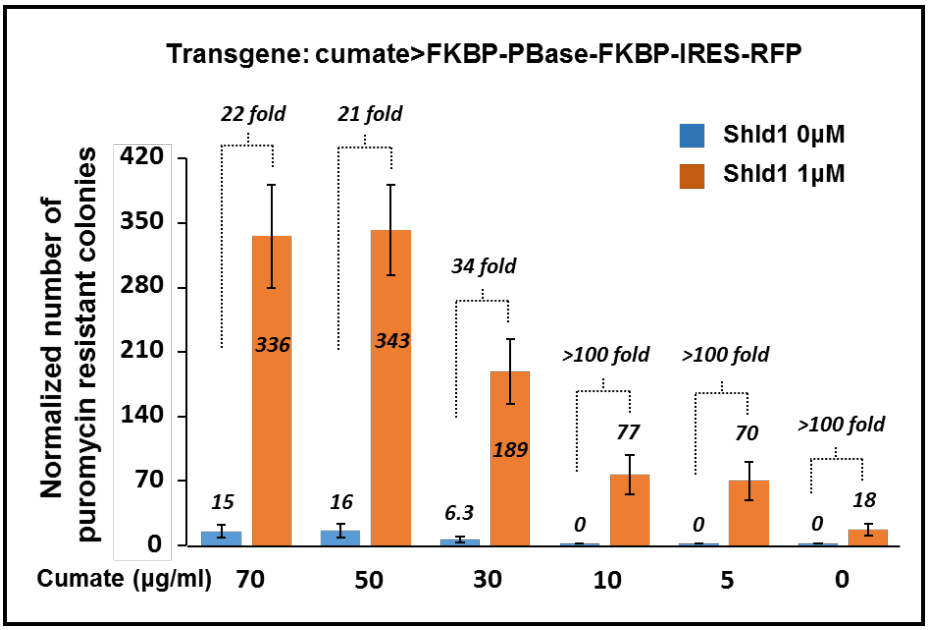


Figure 2.5 The performance of cumate-regulated FKBP-based PB transposon induction system. The cumate chemical ranging from 70 μg/ml to 0 μg/ml was added respectively to regulate the strength of cumate promoter. For each cumate concentration, Shld1 or vehicle (70%

ethanol) was added to the medium to activate the PB transposon induction system. The normalized number of puromycin-resistant colonies was calculated by deducting the background colonies from the transfection with donor and empty vector. The induction fold was calculated as the normalized number of colonies from inducer treated samples divided by that from untreated samples.

2.3.7 The FKBP-PBase-FKBP protein levels can be tuned post-transcriptionally

We next sought to determine if PB activity could be post-transcriptionally tuned by adjusting Shld1 concentration. To do so, we transfected four cell lines with donor and helper plasmids (Fig2.2, construct 3 and 16) and subjected these transfected cells to various concentrations of Shld1: 0nM, 8nM, 40nM, 200nM and 1μM. We calculated the normalized number of puromycin resistant cell colonies for each Shld1 concentration. We found that the colony number increased almost linearly with the increase of the concentration across all cell lines (r=0.81, Fig 2.6A), indicating the FKBP-based PB transposon induction is tunable and dose-dependent.

Counting puromycin-resistant colonies reveals the number of cells with at least one integrated transposon; however, it does not provide information about the average number of transposition events that have occurred per cell. To address this, we also measured PBase activity at various concentrations of Shld1 using an alternative readout for transposition. We created a donor vector (Fig 2.2, construct 17) in which a GFP gene is split by a PB transposon, rendering it inactive (Supplementary Fig. 2.S1A). Nuclear PBase will excise the transposon creating a functional GFP gene that is then expressed, and so the average fluorescence of the cell population is proportional to the number of transposon excision events. The results of these

experiments are shown in Supplementary Figs. 2.S1B and 2.S1C. The mean cellular fluorescence increases monotonically with Shld1 in a near-linear fashion, validating the results obtained with the puromycin donor, and supporting the thesis that PB induction is tunable and dose-dependent.

2.3.8 The FKBP based PBase system is reversible

To determine if the degradation of FKBP-PBase is reversible, we fused a yellow fluorescent protein (YFP) in-frame between FKBP and PBase by an 18 amino acid linker at N terminus of PBase [88] (Fig 2.2, construct 7), allowing the visualization and semi-quantitative analysis of the expression of the PBase fusion by monitoring fluorescence intensity. The four cell lines mentioned above were transfected with the helper plasmid (FKBP-YFP-PBase) and separately, a control plasmid (YFP-PBase). Shld1 was added to the culture medium immediately upon transfection and incubated for one day. Next, the cells were passaged and cultured in drug-free medium for three days. Finally, Shld1 was added back again to re-stabilize the FKBP-YFP-PBase fusion for another three days. The typical images from fluorescence microscopy at three turning points were shown in Fig 2.6B: 1) one day after treatment with Shld1, 2) three days after removing Shld1, and 3) three days after retreatment with Shld1. We also measured the percentage of YFP positive cells every day by flow cytometry. To control for plasmid dilution during the experiment, the percentages of YFP positive cells from the FKBP-YFP-PBase fusions were normalized by the percentage of YFP positive cells in the corresponding control YFP-PBase samples. The results are plotted in Fig 2.6C. Taking Fig 2.6B and 2.6C together, we observed a sharp drop in the YFP positive cell population at the second measurement time point, indicating that without the inducer Shld1, the fusion proteins were quickly degraded. The normalized percentage of YFP positive cells was nearly completely restored to the original value

three days after Shld1 was reapplied, indicating the system is reversible. By fitting the curves in Fig 2.6C to an exponential decay functions, we estimated the half-life of the FKBP-PB fusion protein to be 30h. Our results suggested FKBP-based inducible PB transposon system is tunable and reversible. These levels of regulation could prove invaluable when a certain number of transposition events is required in the induction system.

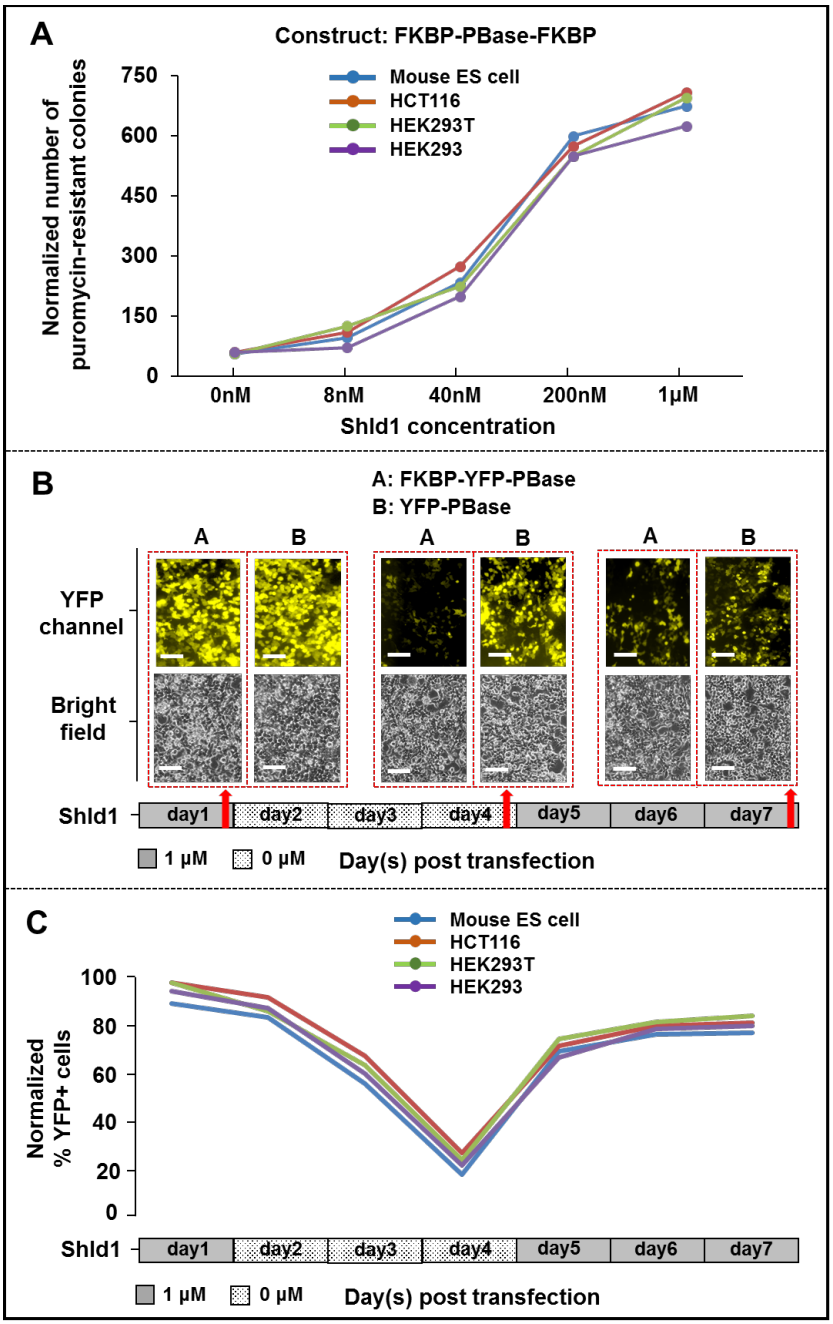


Figure 2.6 Tunability and inducibility of the FKBP-based PB transposon induction system.

(A) Normalized puromycin-resistant colonies under different concentrations of Shld1 across four cell lines. Cells were transfected with donor and helper plasmids (FKBP-PBase-FKBP) and subjected to various concentrations of Shld1: 0nM, 8nM, 40nM, 200nM and 1μM. For each condition, the number of puromycin resistant cell colonies was normalized to that from cells

transfected with the donor plasmid and helper vector (pRS314). (B) Fluorescent images taken prior to FACS analysis at day1, day4 and day 7 of Shld1 treatment for transfected HEK293T cells. The white scale bar equals 50μm. The red arrows indicate the time points of measurement. Bright field and fluorescent images were shown at top and bottom panel respectively. (C) Normalized percentage of YFP positive cells along Shld1 treatment across four cell lines. A minimum of 10000 YFP positive cells were analyzed from each FACS and post-sort analysis was performed with FloJo software to obtain the percentage of YFP positive cells.

2.3.9 Unlike 4OHT, Shld1 does not interfere with general cellular functions

It has previously been reported that 4OHT adversely affects developmental processes such as neurogenesis [84], myelinogenesis [84], myometrial differentiation [85], and sexual maturity [86]. Therefore, we next sought to test if cells treated with Shld1 or 4OHT display any developmental phenotypes. We added these chemicals to embryoid bodies (EBs) generated from mouse ESCs, and differentiated them into ventral spinal neural cells with retinoic acid and smoothened agonist [93]. We found that in the presence of 2μM 4OHT, EB differentiation was inhibited and neuron-like cells were not generated. In contrast, EBs that were mock-treated or treated with 2μM Shld1 differentiated normally (Fig 2.7A). These results suggest that 4OHT inhibits EB differentiation, while Shld1 does not.

To further explore the effects of the inducer molecules on differentiation, we sought to quantify the extent to which these Shld1 and 4OHT perturb the transcriptional network of EBs. EBs were treated with Shld1 and 4OHT respectively for 2 days, induced with retinoic acid and smoothened agonist for 2 more days, and then subjected to gene expression profiling by RNA-Seq. Cells mock-treated with 95% ethanol as vehicle were used as controls. For each condition,

two biological replicates were performed and correlation co-efficiencies between them are above 0.96, demonstrating reproducibility. The mean of the normalized counts from drug-treated samples was plotted against log2 fold change of drug-treated samples to mock-treated ones under the category of Shld1 vs Mock and 4OHT vs Mock (Fig 2.7B and 2.7C). A significant change of gene expression profile was observed for 4OHT-treated samples but not in Shld1-treated ones. After Benjamini-Hochberg correction for multiple hypotheses (see methods), we identified 260 differentially expressed genes in the 4OHT-treated samples. We performed gene ontology (GO) analysis on these genes using the DAVID Bioinformatics package [94, 95] and functional annotation clustering revealed that "heat shock" and "stress response" were the most over-represented terms (adjusted $p < 0.01$), a result that supports our experimental observation that 4OHT is toxic to EBs. In contrast, only one gene (FOS, NM_005252) was differentially expressed in the Shld1 treated samples, suggesting that this chemical has little to no effect on the transcriptional network that controls development of EBs.

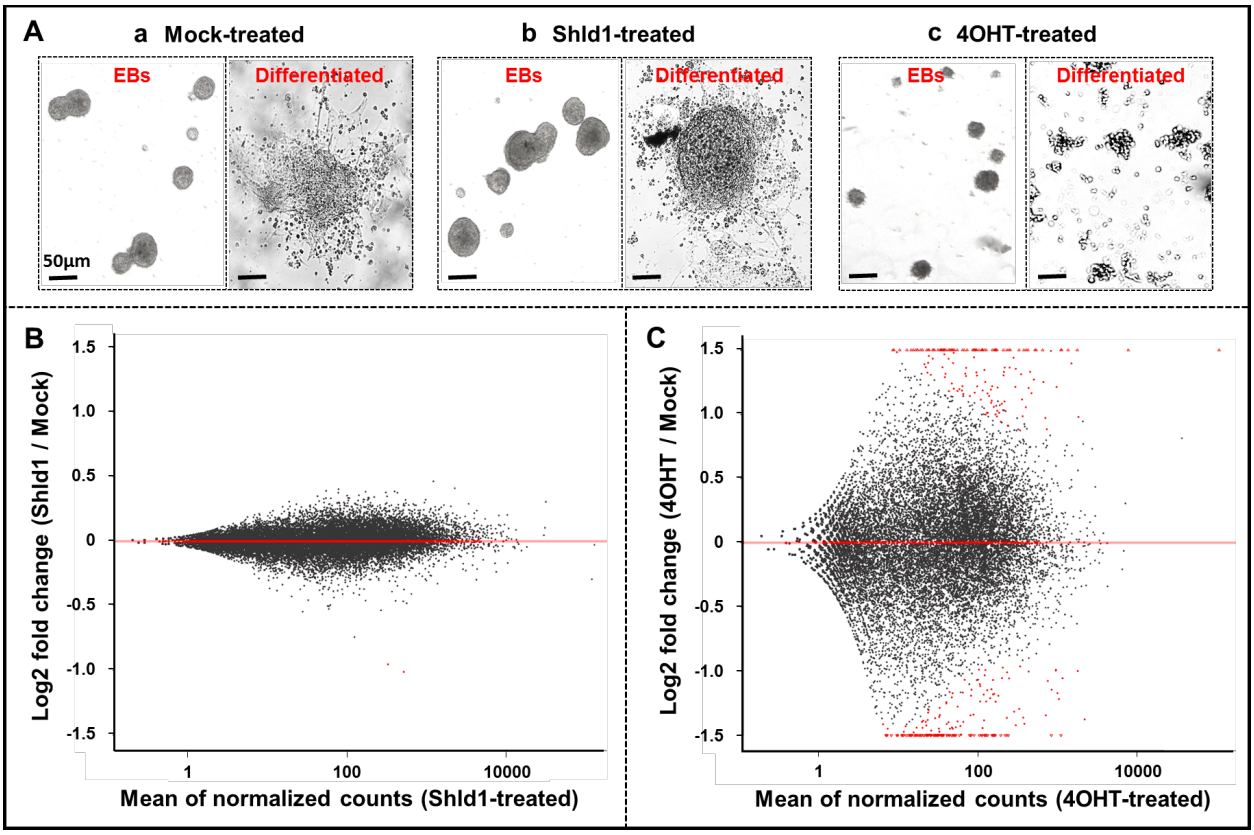


Figure 2.7 The effects of chemical inducer on RW4 mouse ES cells. (A) Images of embryoid bodies (EBs) and differentiated EBs treated with (a) 95% ethanol as vehicle, (b) 2μM Shld1 and (c) 2μM 4-hydroxytamoxifen during EB formation and neural differentiation from mouse ESCs to neural lineages. (B and C) The comparison of gene expression profile of drug-treated samples with that of mock-treated ones. Experiments were done in duplicate. Mean of the normalized counts from drug-treated samples was plotted against log2 fold change of drug-treated samples to mock-treated ones under the category of (B) Shld1 vs Mock and (C) 4OHT vs Mock. Red dots indicate the differentially expressed genes.

2.4 Discussion

We systematically characterized, in four cell lines, 15 different PB transposase fusion proteins representing 3 different induction systems. We found that the FKBP-based system achieved the broadest dynamic range of induction across four cell lines. Remarkably, in the presence of chemical inducer, this system had transposition efficiencies that were almost as high (~95%) as the "wild-type" PB transposon. These results, coupled with the fact that Shld1 does not affect ESC development, suggest that this will be the preferred induction system for many types of experiments, especially those involving cellular differentiation or organismal development.

In our plasmid-based experiments (Fig 2.3), we observed a lower dynamic range for the ERT2-based PB transposon induction system that was previously observed by Cadinanos et al. (~20 fold vs ~270 fold) [16]. This discrepancy is explained by the fact that, in our donor plasmid, the puromycin resistance gene is driven by a PGK promoter whereas in Cadinanos et al., the puromycin resistance gene is promoterless and downstream of a splicing acceptor site. In that system, only insertions in active genes will produced colonies, but in our system, all insertions will produce resistant clones. This would explain why we observed more background puromycin-resistant clones when the transposase was not induced since both studies normalized to background and small absolute changes in the denominator can lead to large changes in the calculated fold change.

The moderate background transposition of the DD induction system observed in our plasmid experiments (Fig 2.3) was largely the result of the over-expressed fusion protein overloading the cells' proteasome system, since when the FKBP-fused PBase was placed under the control of a cumate-titratable promoter and delivered at lower copy number, we observed no

background transposition while still achieving robust transposition after induction. Since no background hops were observed at several cumate concentrations, the fold-induction of the system is not defined; we conservatively estimate it as > 100-fold.

An interesting difference between DD-based systems and the ERT2 based system is that the performance of the DD systems was more uniform across the difference cell lines. We hypothesize that this difference is due to the fact that the DD system uses the ubiquitin-proteasome system (UPS) to constantly degrade FKBP-fused PBase. The UPS is a general degradation machinery and uniformly expressed and works efficiently in all mammalian cell types. In contrast, the ERT2-based induction system is dependent on a specific cytoplasmic protein HSP90, a gene whose expression varies widely between cells. Our RNA sequencing data showed that the HSP90 expression level in the RW4 mouse ES cell line is about 4-fold higher than in the HCT116 cell line, suggesting that a low amount of HSP90 levels may lead to either a high basal activity without induction or a low induced activation.

In the FKBP-based PB transposon induction system, we observed that N-terminally tagged PBase (FKBP-PBase) has a higher fold induction than the C-terminally tagged counterpart (PBase-FKBP), which may suggest that FKBP fused at the N terminus is more likely to be exposed and thereby recognized by the proteasome than when fused to the C terminus. This terminal specific higher activity may relate to the distribution of functional components of PBase. It was reported that the PBase has a functional nuclear targeting signal in the 94 C-terminal residues [96]. C-terminal fused PBase, therefore, is more likely to negatively affect the nuclear translocation of the PBase protein, which in turn increases the likelihood of being recognized and degraded by the proteasome in the cytoplasm. We also observed that fusing additional DDs to the PB transposase increases the fold-change in activation between the

induced and uninduced conditions. However, the addition of multiple DDs causes the absolute activity of the induced protein to be significantly reduced, suggesting a trade off between how tightly a protein can be regulated, and its maximum activity in the induced state. Our results suggest that the FKBP-DD is optimal when a high level of PB transposition is preferred, while FKBP-FKBP-PBase-FKBP is the choice when tight regulation is the priority.

The FKBP system will be likely prove considerably more useful than the ERT2 system for the study of developmental processes, since we found that the ERT2 inducer tamoxifen is a strong inhibitor of embryoid body differentiation. Our results are consistent with previous reports that tamoxifen has deleterious effects on a number of developmental systems [84-86]. In contrast, we found that Shld1, the FKBP inducer, has no such phenotype, and furthermore, few transcriptional changes were observed in our RNA-Seq analysis of embryoid bodies treated with Shld1. Together, these results suggest that the FKBP DD system is preferable to the ERT2 system for experiments involving cellular differentiation.

2.5 Method

2.5.1 Plasmid construction

The plasmids that contain mutated destabilized domains, FKBP and DHFR, were purchased from Addgene with ID 31763 and 29326 respectively. The PBase-ERT2 plasmid (mPBase-L3-ERT2) was a kind gift from Dr. Bradley [16]. The coding sequence of mPBase-L3-ERT2 was sub-cloned into a yeast shuttle vector pRS314 containing CEN6, ARS, and TRP1 to use “gap repair” cloning technique in yeast cells (Strathern and Higgins 1991). This engineered yeast shuttle vector was used as a “parental” plasmid (pRM1056) to derive other variants of PBase constructs by gap repair method [2]. Briefly, PCR-generated sequences were cloned into

linearized vectors by recognizing a 40 bp overlap at their ends. This 40 bp overlap can be engineered by designing primers for amplification of the desired sequences. For example, to replace the ERT2 sequence with the FKBP sequence, the pRM1056 plasmid was linearized by restriction digestion that cut the plasmid within the ERT2 sequence. The FKBP sequence was amplified with a pair of primers that have a 40bp sequence that is homologous to pRM1056. The FKBP PCR products and linearized pBM1056 were co-transformed into yeast cells and the yeast cells were selected for Trp+ colonies. DNA extracted from Trp+ yeast colonies was introduced into *E. coli*. Finally, the plasmid was isolated by QIAprep Spin Miniprep Kit (QIAGEN) and was confirmed by Sanger sequencing. The engineered constructs used in this study are depicted in Fig 2.

2.5.2 Cell culture and neural differentiation

Human embryonic kidneys cell lines (HEK293 and HEK293T) and human colon adenocarcinoma cell line HCT116 were maintained in Dubecco's Modified Eagle Media (DMEM; Gibco) supplemented with 10% FBS. RW4 mouse embryonic stem cells (ESCs) were cultured in complete media consisting of DMEM supplemented with 10% new born calf serum, 10% fetal bovine serum (FBS; Gibco), and 0.3 mM of each of the following nucleosides: adenosine, guanosine, cytosine, thymidine, and uridine (Sigma-Aldrich). To maintain their undifferentiated state, cells were also cultured in the presence of 1000 U/mL leukemia inhibitory factor (LIF; Chemicon) and 20 mM β-mercaptoethanol (BME; Invitrogen) in flasks coated with a 0.1% gelatin solution (Sigma-Aldrich). Mouse ESCs were differentiated into ventral spinal neural cells using a retinoic acid (RA) and smoothened agonist (SAG) induction protocol as described [93].

2.5.3 Cell transduction and transgenic cell line generation

All lentiviruses used for cell transduction were produced by the Hope Center Viral Vectors Core at Washington University School of Medicine. A CymR-expressing HEK293 cell line was made by transducing the cell with lentivirus containing the cumate operator repressor driven by an EF1a promoter at a multiplicity of infectivity (MOI) of 20 to ensure that CymR is overexpressed. Two days after transduction, neomycin was added to the culture medium (500ng/ml) and maintained for 7 days. Then, the CymR-expressing HEK293 cell line was further transduced with lentivirus containing an FKBP-PBase-FKBP-IRES-RFP fusion gene driven by a cumate inducible promoter at MOI of 0.5 to favor single-copy integration. To obtain a pure population, transduced cells were then sorted for red fluorescent protein (RFP) positive cells.

2.5.4 Cell transfection and drug administration

All plasmids used for the transfection of cells were prepared using EndoFree Plasmid Maxi Kit (Qiagen) following the manufacturer's protocol. About 10^5 cells were electroporated with a total of 0.6 μg of DNA (0.1 μg helper plasmid and 0.5 μg donor plasmid) by the Neon transfection system (Invitrogen) and plated to one well of a 6-well plate. Immediately upon transfection, cells were treated with 1 μM 4-hydroxy tamoxifen (4OHT, Sigma) for the ERT2 based induction system, 1 μM Shld1 (Clontech) for FKBP based induction system, 10 μM trimethoprim (TMP, Sigma) for the DHFR based induction system. Negative controls were mock-treated with 95% ethanol as vehicle. For the transgenic cell line expressing inducible PBase driven by a cumate promoter, only donor plasmid was used in transfection. Cumate inducer (System Biosciences) was added at concentrations ranging from 70 μg/ml to 0 μg/ml to regulate the strength of cumate promoter. For the purpose of reproducibility, experiments were

done in triplicates. To select cells in which the PB transposon transposed, cells were trypsinized 2 days after transfection and cultured on 10cm dish (Corning) with 10 ml fresh media containing puromycin (1 μg/ml). The puromycin selection normally takes 7 days before visible cell colonies are formed.

For experiments testing the reversibility of the FKBP-based PB transposon system, Shld1 (1 μM) was added to the culture medium immediately after transfection. One day after transfection, a small aliquot of cells were used for imaging and flow cytometry analyses and the rest of the cells were passaged to a 12-well plate and provided fresh Shld1-free medium. The passaged cells were grown to about 100% confluence before another passage to a 12-well plate. One well of cells were used for imaging and flow cytometry analyses every day. After 3 days of destabilization, the cells were treated with fresh medium containing 1 μM Shld1 to re-stabilize the FKBP-YFP-PBase fusion protein. The re-stabilization lasted for another 3 days during which the cells were analyzed by flow cytometry every day.

2.5.5 Cell colony staining and counting

The visible drug-resistant cell colonies were fixed with PBS containing 4% paraformaldehyde for 1 hour and then stained with 1% methylene blue in 70% EtOH for 30 min, washed in distilled water and air-dried overnight. Colonies with diameters more than 0.3 mm were counted by ImageJ software (National Institutes of Health). The number of puromycin-resistant colonies formed from the cells transfected with both donor and helper plasmids was normalized to that with donor and helper backbone plasmids (yeast shuttle vector pRS314) prior to any further calculations. The transposition activity of a PBase fusion protein was calculated as the normalized number of colonies from the inducible domain (i.e. FKBP, DHFR or ERT2)

fused PBase divided by that from “wild type” unfused PBase. For experiments that used the transgenic cell line expressing FKBP-PBase-FKBP driven by a cumate promoter, we used normalized number of puromycin-resistant colonies to estimate the transposition activity due to the unavailability of the cell line expressing “wild-type” unfused PBase. Fold-induction was calculated as the normalized number of colonies from the induced samples divided by that from untreated samples.

2.5.6 Imaging and flow cytometry

Fluorescent images were taken on Zeiss Axioskop fluorescence microscope equipped with a QICAM FAST 1394 digital CCD camera. Cells were grown to 90% confluence, trypsinized from the plate, and suspended in phosphate-buffered saline (PBS), washed once with PBS, and resuspended in Hank’s balanced salt solution supplemented with 2 mM EDTA. Cellular fluorescence was analyzed on an iCyt Reflection HAPS2 cell sorter at the Washington University Siteman Flow Cytometry Core. The gate was set relative to the cells transfected with non-fluorescent control plasmids to eliminate background. Cells transfected with a positive control fluorescent reporter plasmid were also used to eliminate false positive singles. About 10000 cells were analyzed from each FACS and post-sort analysis was performed with FloJo software to obtain the percentage of fluorescent positive cells.

2.5.7 RNA extraction and sequencing

Total RNA was isolated from the RW4 mouse ESCs using the PureLink RNA Mini kit (Ambion) according to the manufacturer's instructions. The quantity of RNA was measured using a spectrophotometer (NanoDrop 2000c; Thermo Scientific). Samples with a RNA

concentration (A260/A280 ≥1.8 ng/µl) and purity (A230/A260 ≥2.0 ng/µl) were selected. The Agilent 2100 Bioanalyser was used to determine the RNA integrity number (RIN). The degradation level was identified using the RNA 6000 Nano LabChip kit (Agilent). Samples with RIN >9.8 were further processed using TruSeq mRNA Library Preparation Kit (Illumina) and then sequenced by Illumina MiSeq platform at the Genome Technology Access Center (GTAC) at Washington University in St. Louis. The gene expression data generated for this study can be found under the NCBI Gene Expression Omnibus (GEO) accession number GSE78857. The expression data is also publically available at the Center for Genome Sciences by request.

2.5.8 Reads mapping and statistical analysis

Trimmomatic (v0.32) [97] was employed on RNA-Seq FASTQ files to clip the illumina adaptors and remove the reads of low quality. The cleaned reads were mapped back to mm10 genome reference from UCSC database using STAR (v2) [98]. We used the HTSeq package[99] to estimate the count of uniquely mapped reads for each of the annotated genes in the mm10 gene transfer format (.GTF) file. Differential expressed genes were analyzed with R (v 2.13.0) using DESeq (v1.4.1) [100] available in Bioconductor (v 2.8). The resulting p-values were adjusted using the Benjamini-Hochberg correction and only genes that were significant at a false discovery rate (FDR) of 0.05 were considered as expressed. For other comparisons between different experimental conditions, the statistical significance was assessed by paired Student t test and a p value less than 0.05 was considered significant.

Chapter 3
Brd4-bound enhancers drive critical sex differences in glioblastoma

Published in BioRxiv. (2018)

3.1 Abstract

The incidence and severity of many diseases is strongly dependent on sex. New approaches to treatment could be revealed by dissecting the molecular pathways that control sexually dimorphic phenotypes. In an established model of glioblastoma, we discovered that pharmacological inhibition or genetic depletion of Brd4 decreased clonogenicity and tumorigenicity in male cells while increasing clonogenicity and tumorigenicity in female cells, thus abrogating the sex differences in phenotype. The sex differences in tumorigenic phenotype correlated with transcriptome-wide sexual dimorphism in gene-expression, H3K27ac marks and large Brd4-bound enhancers. Finally, sexual dimorphism in Brd4 function was also suggested by the differential effect of low Brd4 expression on survival in glioblastoma patients. Thus, for the first time, sex is established as an intrinsic element of cellular identity that is driven by Brd4 activity, which renders male and female cells differentially sensitive to BET inhibitors treatment. This has important implications for the clinical use of BET inhibitors and provides new molecular targets for glioblastoma treatment.

3.2 Introduction

Until recently, most basic and clinical research has focused on investigating factors that influence disease susceptibility and progression without regard to sex. However, a mounting body of evidence has revealed significant sex-specific differences in incidence, age of onset, and outcome of numerous human diseases. These include cardiovascular diseases, asthma, autoimmune diseases, birth defects, neurological diseases, psychiatric disorders and cancers [101]. This preponderance of sex differences in disease incidence and outcome led to the implementation of new guidelines by the NIH regarding inclusion of sex as a biological variable in all research.

Glioblastoma (GBM), the most common and aggressive form of brain cancer [102, 103], is more common in males regardless of race or region of the world (male to female incidence of 1.6:1) [104-107]. In addition to the sex difference in GBM incidence, a recent study by Ostrom et al. also revealed a sex difference in prognosis between male and female GBM patients wherein females have a significant survival advantage [108]. The reasons for sex differences in GBM incidence and outcome are largely unknown. While sex differences in disease are often mediated through acute sex hormone actions, sex differences in malignant brain tumor rates are evident at all ages, suggesting that factors other than circulating sex hormones underlie this discrepancy [109]. These can include organizational or epigenetic effects of transient *in utero* sex hormones and extra-gonadal effects of sex chromosome encoded genes. Recently, our lab discovered that an established model of GBM involving combined loss of neurofibromin (NF1) and p53 function in murine neocortical astrocytes exhibits sex differences in *in vivo* tumorigenicity mimicking those observed in patients with GBM [19]. Importantly, these male and female GBM astrocytes display sex differences in response to serum deprivation, CDKi

treatment and chemotherapy [110]. Together, the human and mouse data suggest that male and female cells may be differentially sensitive to the transforming effects of specific oncogenic events.

Sexual differentiation is in large part an epigenetic phenomenon [111]. A preponderance of studies have shown that inhibiting epigenetic modifiers can block normal sexual dimorphism in the brain. Understanding whether sex differences in GBM are mediated by epigenetic mechanisms involving epigenetic readers and/or writers will be imperative to understanding the key processes that impart females with relative resistance and males with relative susceptibility to cancer. Here we show that the previously observed phenotypic differences between male and female GBM cells are abolished by targeting the Bromodomain and extra-terminal (BET) family proteins using a panel of BET inhibitors currently in clinical trials [112]. Genetic depletion of the BET family members indicated that the effects are mediated by the epigenetic reader Brd4. Mapping of Brd4 genomic localization revealed that sex differences in tumor phenotype arise through differential Brd4-bound enhancer usage in male and female GBM cells. Consistent with these data, we observe sex differences in the effects of low Brd4 expression on survival in male and female GBM patients. This is the first demonstration that differential Brd4 activity mediates cell intrinsic sex identity and sex differences in a cancer phenotype. Together with sex differences in BET inhibitors effects, these results strongly indicate that sex differences in disease biology translate into sex differences in therapeutic responses. This has broad implications for medicine.

3.3 Results and discussion

3.3.1 BET family protein (Brd4) inhibition has opposing effects on the tumorigenic phenotype in male and female GBM astrocytes

Based on our previous observation of sex-specific differences in the tumorigenic phenotype and response to chemotherapy between male and female GBM cells [19], we first sought to investigate whether these GBM astrocytes also exhibit sex-specific treatment responses to BET inhibitors currently in clinical trials. We treated male and female GBM astrocytes with a panel of BET inhibitors that are presently being tested in clinical trials and then performed extreme limiting dilution assay to measure clonogenic cell (stem-like cell) frequency (ELDA). These BET inhibitors are either selective for Brd4 or target all members of the Brd family. Treatment with BET inhibitors reproducibly decreased clonogenic frequency in male GBM cells while increasing the functional clonogenic cell fraction in female cells **(Figure 3.1A)**. Next, and to determine which BET family protein was mediating these effects, we performed genetic depletion of Brd2, 3 and Brd4 family members. We used shRNAs specific to either Brd2, Brd3 or Brd4 then evaluated the effect on the tumorigenic phenotype of male and female GBM cells. As shown in **Figure 3.1C**, mRNA expression levels of Brd2, Brd3 and Brd4 were partially silenced in male and female GBM cells after infection with lentiviral shRNAs. Of note, Brd4 mRNA4 levels were equivalent between male and female GBM cells under basal conditions. Knockdown of Brd2 with shRNAs did not affect clonogenic frequency in either male or female cells, which is consistent with our earlier RVX-208 (a Brd2/3 selective inhibitor) results **(Figure 3.1B)**. Interestingly, male GBM cells with knockdown of Brd4 exhibited a decrease in clonogenic frequency whereas female cells displayed an increase in clonogenic frequency. This opposing sex-specific response to the inhibition of Brd4 is

consistent with published data in breast and prostate cancer, wherein ectopic expression of Brd4 in breast cancer cells decreased invasiveness and tumor growth, while Brd4 inhibition decreased viability of prostate cancer cells [113-115]. A similar effect on clonogenic frequency was observed following knock down of Brd3 in male but not female cells.

To investigate the effect of Brd2, 3 and 4 expression on sex-specific survival outcomes in human GBM, we obtained gene expression data from 151 (98 males and 53 females) TCGA glioblastoma samples, converted the expression value of Brd into a z-score specific to the sex of the patient and stratified the patients into a high-expression group (z-score > 1.0) and low-expression group (z-score < -1.0). We analyzed the effects of Brd expression on patient overall survival (OS) using the Kaplan-Meier method [116]. No difference in survival was detected in low Brd2 or Brd3 expression groups (**data not shown**). While high Brd4 expression group had no effect on survival, low Brd4 expression was associated with shortened survival in females (OS median - 5.39 months) compared to males (OS median - 16.59 months, $p = 0.05$) (**Figure 3.1D)**. This is consistent with our findings in our mouse GBM model, in which treatment with the BET inhibitor abrogated the tumorigenic phenotype in male cells, but enhanced it in female cells. Additionally, this data is consistent with previously published breast, endometrial and prostate cancer studies revealing that in women with estrogen receptor positive breast cancer [113] or endometrial cancer [115] low Brd4 expression is correlated with worsened survival. This is in contrast to men with prostate cancer in whom low levels of Brd4 are associated with improved survival [114, 115]. Altogether, and with the multitude of evidence for the deregulation of Brd4 in numerous cancers, these data affirm a role of Brd4 as a mediator of sex differences in GBM tumorigenic phenotype.

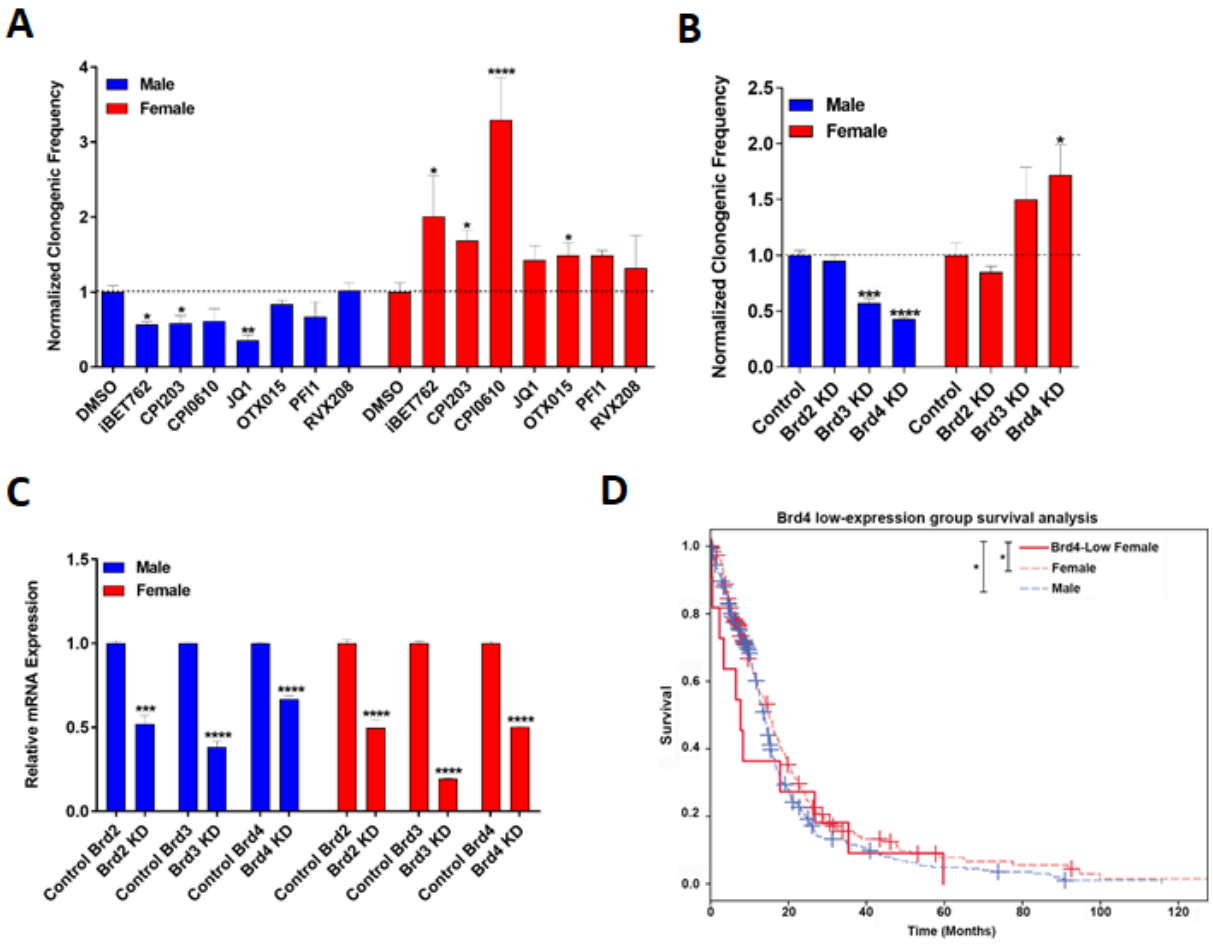


Figure 3.1 Brd4 inhibition has opposing effects on the tumorigenic phenotype in male and female GBM astrocytes. (A) Clonogenic cell frequency (ELDA) assays were performed in GBM cells treated with DMSO or BET inhibitors. BET inhibitors significantly reduced clonogenic cell frequency in male cells while female cells exhibited a significant increase in their stem-like cell frequency. Clonogenic cell frequency were unaffected by RVX-208, a Brd2/3 selective antagonist. **(B)** Clonogenic cell frequency assays were performed in shRNA knockdown and control GBM cells. Knockdown of Brd4 and Brd3, but not Brd2, suppresses clonogenic frequency in male GBM cells while female cells showed an increase in clonogenic frequency following Brd4 knockdown only. Brd2 depletion was without effect. All treatment groups were normalized to male control clonogenic frequency levels (*=p <0.05, **=p <0.01,

***=p <0.001, ****=p <0.0001 as determined by two-tailed t-test and one-way ANOVA). **(C)** Low expression of Brd4 has sex-specific effects on GBM patient survival. Female GBM patients with low Brd4 expression (z-score < -1.0) have decreased survival, (median OS - 5.39 months), compared to male GBM patients with low Brd4 expression (z-score < -1.0, median OS - 16.59 months; p=0.05).

3.3.2 Male and female GBM cells utilize different sets of Brd4-bound enhancers

In order to understand the role played by Brd4 in mediating sex differences in GBM, we sought to map Brd4 genomic localization in male and female GBM astrocytes (highly active Brd4-bound enhancers and typical Brd4 bound enhancers). Highly active Brd4-bound enhancers have been shown to play key roles in establishing cell-identity and thus may be important in establishing sex differences. Furthermore, these enhancers bind up to 90% of Brd4 protein in the cell [47, 117-121]. Brd4 is an epigenetic reader that binds acetylated histones H3 and H4 throughout the entire cell cycle and is known to be deregulated in numerous cancers [122]. Brd4 is thought to play a central role in cancer by promoting epithelial-to-mesenchymal transition, stem cell-like conversion, and pluripotency [31, 123, 124], and the pharmacological inhibition of this protein has shown therapeutic activity in a number of different cancer models [32, 112, 114, 125, 126]. To investigate whether these highly active Brd4-bound enhancers might play a role in establishing cell-intrinsic sex differences, we used transposon Calling Cards [2, 78] to identify enhancers differentially bound by Brd4. To do so, we fused a *piggyBac* (PB) transposase to the C terminus of the Brd4 protein, endowing it with the ability to direct the insertion of the PB transposon into the genome close to Brd4 binding sites. Three biological replicates were carried out and the correlation between replicates was r > 0.9 for all pairwise

comparisons **(Figure 3.S1)**. Using this protocol, we mapped ~700,000 unique insertions directed by the Brd4-PBase fusion for male and female samples and identified 3014 enhancers that bound significantly more Brd4 protein in males and 3122 enhancers that bound significantly more Brd4 protein in females **(Figure 3.2A)**.

To determine whether the enhancers that bound Brd4 in a sex specific manner also displayed differential H3K27ac, we performed Chromatin Immunoprecipitation Sequencing (ChIP-seq) in male and female GBM cells to identify genomic regions enriched for H3K27 acetylation, a well-known marker of active enhancers. Three biological replicates were carried out and the correlation between replicates was $r > 0.9$ for all pairwise comparisons **(Figure 3.S1)**. Using established methods [127], we identified a total of 48881 and 51232 H3K27ac-enriched peaks in male and female GBM cells respectively. Of these, 10861 (22%) were male-specific and 13212 (26%) were female-specific **(Figure 3.2B)**. We then performed differential analysis between male and female cells and identified an additional 15976 differentially H3K27ac-enriched regions, as depicted in the heat map clustered by male and female **(Figure 3.2D)**.

Next, we analyzed the distances of sex-specific Brd4 binding sites to the nearest sex-specific H3K27ac-enriched enhancer regions. As shown in **Figure 3.2E and F**, sex-specific Brd4 binding sites are significantly enriched at sex-specific H3K27ac enhancer regions ($p < 0.01$). Representative examples of two biologically relevant loci, Nkx2.1 and Zic1/4, which correspond to male-specific and female-specific Brd4-marked enhancer regions respectively, are depicted in **Figure 3.2G and H**. This is the first demonstration of differential Brd4-bound enhancer usage by male and female cells of any kind and is consistent with earlier observations that these enhancers are key regulators of cell identity and fate [47, 117, 118, 121, 128].

Using sex-specific Brd4 binding sites that correlated with sex-specific H3K27ac-enriched enhancer regions within 1kb distance, we identified the closest upstream and downstream genes from the center of the Brd4 binding sites. This analysis revealed 351 male-specific genes and 384 female-specific genes. Pathway enrichment analysis on male-specific genes regulated by these sex-specific Brd4-marked enhancers revealed functional enrichment for glioma, neoplasm metastasis, metabolism, cell proliferation, chromosome aberrations and integrin signaling. Similar analysis on female-specific genes showed an enrichment in pathways involved in regulation of metabolic process, DNA repair-deficiency disorders, and semaphorin signaling **(Table 3.S1)**.

Based on our observation of sex-specific enhancers usage between male and female GBM cells [19], we sought to characterize the male and female transcriptome. We profiled male and female GBM cells with RNA sequencing (RNA-seq). For each condition, three biological replicates were performed; the data were highly reproducible (Pearson $r \geq 0.96$ for all pairwise comparisons) (**Figure 3.S1**) and indicated differential expression of 3846 transcripts (FDR < 0.05) (**Figure 3.2I**). Pathway enrichment analysis for the top 400 differentially regulated genes was performed using a combination of KEGG pathway and Genomatix Pathway System (GePS). Classification of these genes according to function revealed a significant number of relevant and important pathways including cell differentiation, cell proliferation, glioblastoma, tumor angiogenesis, metabolism, and DNA binding-transcription factors **(Table 3.S2)**. Thus, there are global differences in male and female GBM cell transcriptomes in cancer relevant pathways, which are due to differences in sex-specific enhancer activity.

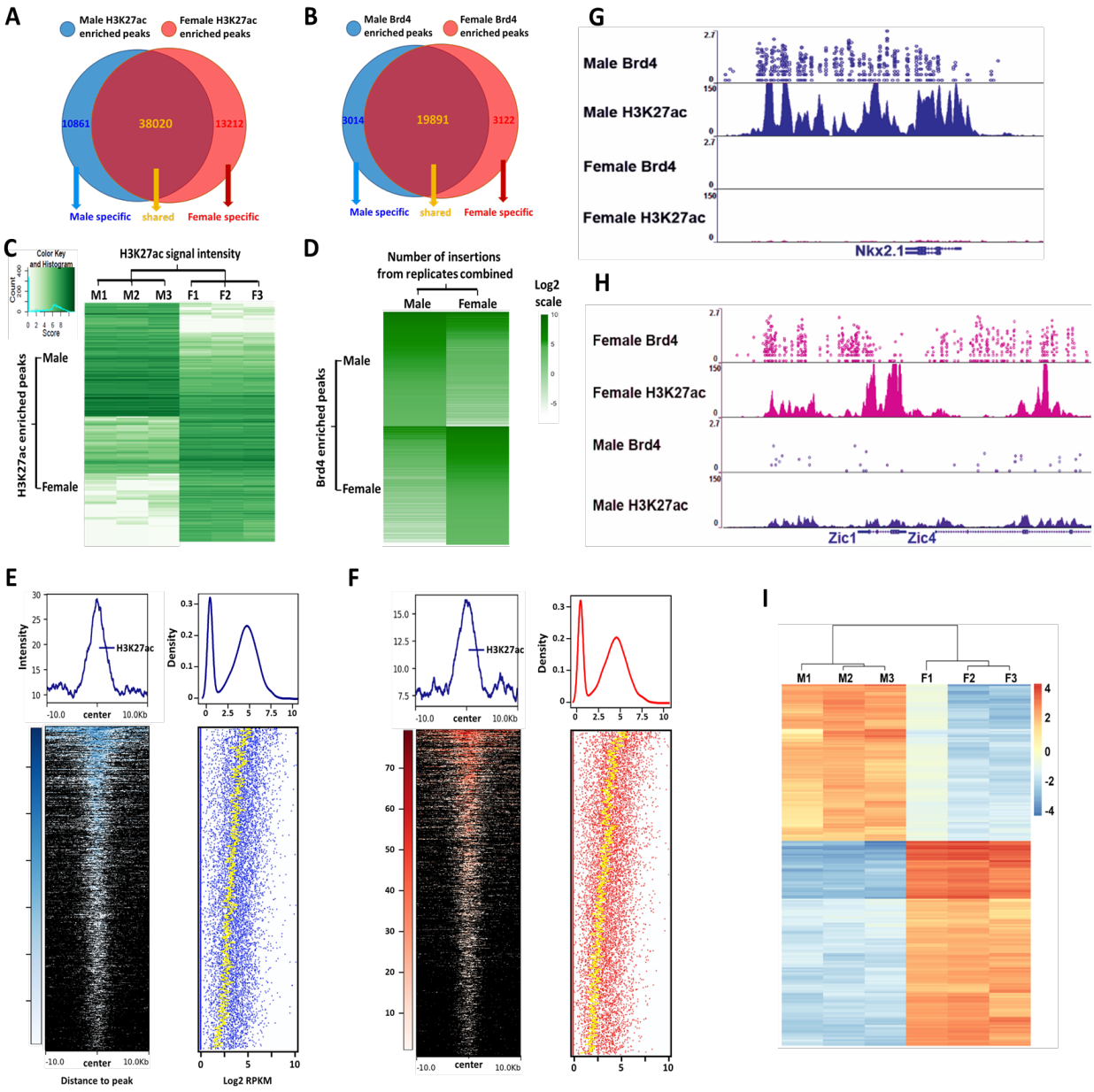


Figure 3.2 Male and female GBM cells have sexually dimorphic Brd4-bound enhancers. (A) 3014 male-specific and 3122 female-specific Brd4-enriched peaks were identified. **(B)** 10861 male-specific and 13212 female-specific H3K27ac-enriched peaks were identified. **(C)** Heatmap of Brd4 binding affinity (number of insertions) for sex-specific Brd4 binding sites in male and female GBM cells **(D)** Heatmap of H3K27ac signal intensity (read depth) for sex-specific H3K27ac-enriched regions in male and female GBM cells. Samples are clustered first by sex,

then by replicate. Heatmap and intensity plot of the H3K27ac signal among sex-specific Brd4 binding sites (+/- 10kb from the center) in male **(E)** and female **(F)**. The expression values of the nearest gene(s) from the corresponding Brd4 binding sites was also plotted side by side. **(G)** Nkx2.1 and **(H)** Zic1/4 illustrate male and female-specific genes, respectively, associated with differential Brd4 binding affinity and H3K27ac enrichment. The x-axis (blue arrows) of all tracks corresponds to genomic location of the gene. The y-axis of Calling Card tracks represents the log10 scale of sequencing reads for each insertion as indicated by circles. The y-axis of ChIP-seq tracks represents the number of uniquely mapped reads). **(I)** RNA abundance (RNA sequencing) in male and female GBM cells (top 200 upregulated and 200 downregulated male/female genes (n=3)).

3.3.3 Brd4-bound enhancers regulate sex differences in GBM

To further validate the transcriptional activation of sex-specific regulatory genes and networks by Brd4-bound enhancers, we treated male and female GBM cells with the Brd4 antagonist JQ1. JQ1 is a thieno-triazolo-1,4-diazepine that displaces Brd4 from chromatin by competitively binding to the acetyl-lysine recognition pocket [112, 125]. Treatment of acute myeloid leukemia cells with JQ1 caused a rapid release of Mediator 1 (Med1) from a subset of enhancer regions that were co-occupied by Brd4, leading to a decrease in the expression of neighboring genes [129]. To investigate the functional activity of Brd4 at sexually dimorphic enhancer sites regulating expression of cancer-relevant genes, we treated male and female GBM cells with either vehicle (0.05% DMSO) or 500 nM JQ1 for 24 hours, and then isolated genomic material for RNA-seq. Gene expression analysis on JQ1 treated cells revealed that Brd4 proximal genes are significantly downregulated compared to genes that are distal to Brd4

binding sites (p <0.01) indicating that JQ1 has a specific and directed effect on genes whose expression is driven by Brd4-bound enhancers **(Figure 3.3 and 3.4).** Pathway analysis for sex-specific Brd4-bound enhancers' genes downregulated following JQ1 treatment in male and female GBM cells revealed functionally important pathways such as cancer stem cell pathway regulated by sex-specific Brd4-bound enhancers **(Table 3.S1 and Figure 3.5)**.

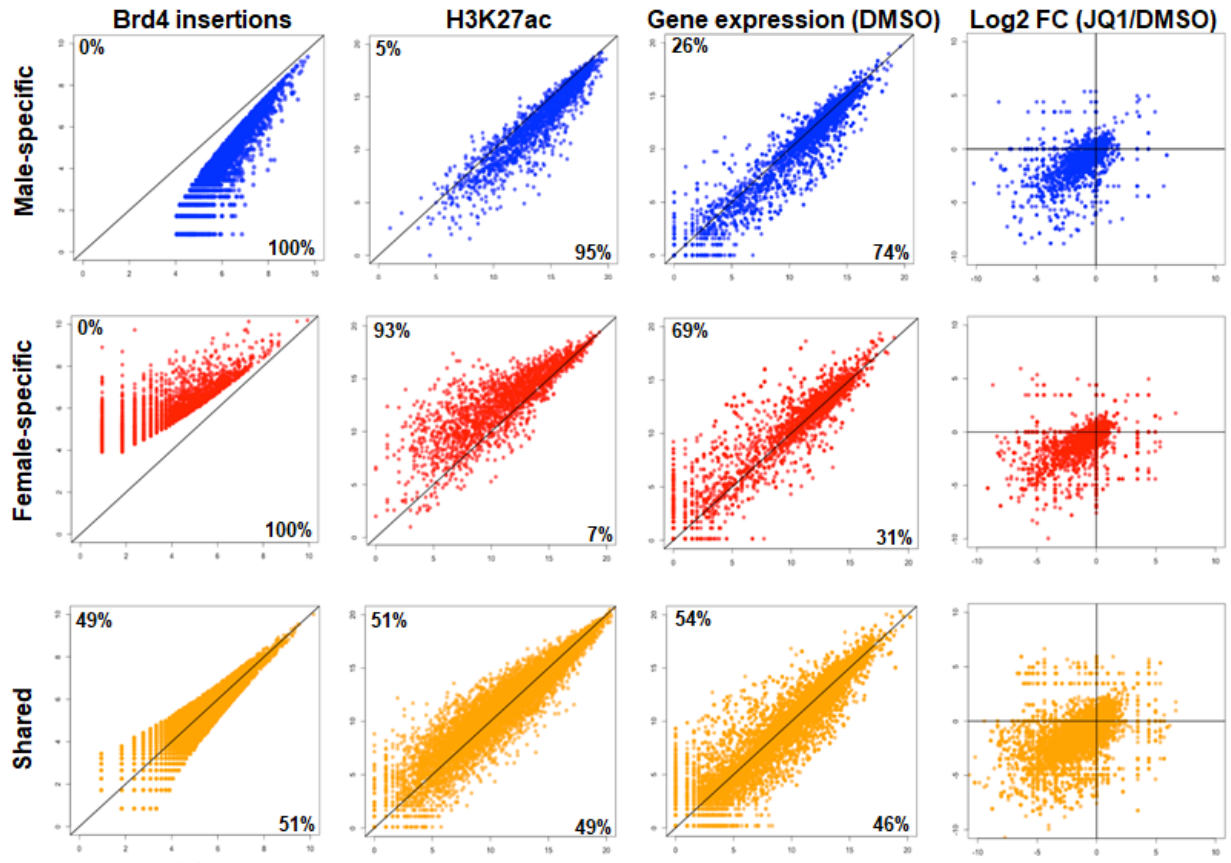


Figure 3.3 Male and female GBM cells have sexually dimorphic Brd4-bound enhancers and concordant gene expression. Sex specific and shared Brd4-bound enhancers are analyzed under the data category of 1) number of insertions from Calling Card method, 2) H3K27ac signal from ChIP-seq, 3) expression value from nearest gene(s) from RNA-seq and 4) expression changes (log2) upon JQ1 treatment.

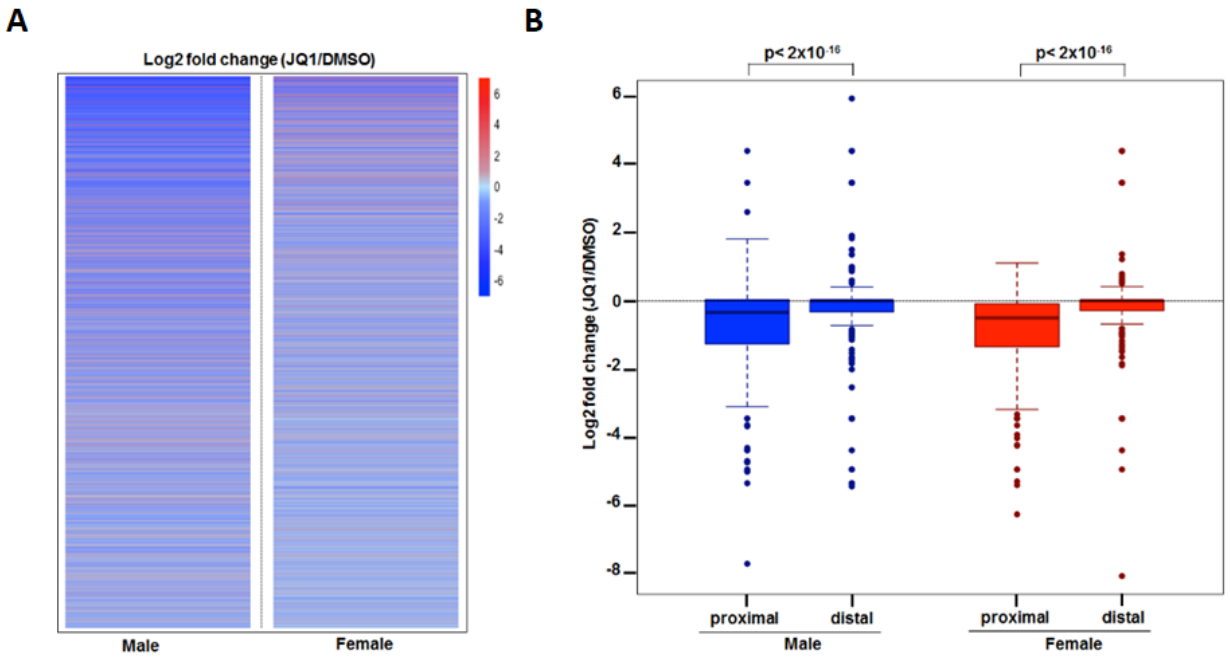


Figure 3.4 JQ1 effect on gene expression in Male and female GBM cells. (A)Heatmap of gene expression changes (log2) upon JQ1 treatment in male and female GBM cells (B) Boxplot of gene expression changes (log2) of Brd4 proximal and distal genes following JQ1 treatment of male and female mouse GBM cells. The gene expression profile of genes in close proximity to Brd4 binding sites is compared to distal genes by a Mann-Whitney-Wilcoxon test. Brd4 proximal genes are significantly downregulated compared to Brd4 distal genes in both male and female GBM cells following JQ1 treatment.

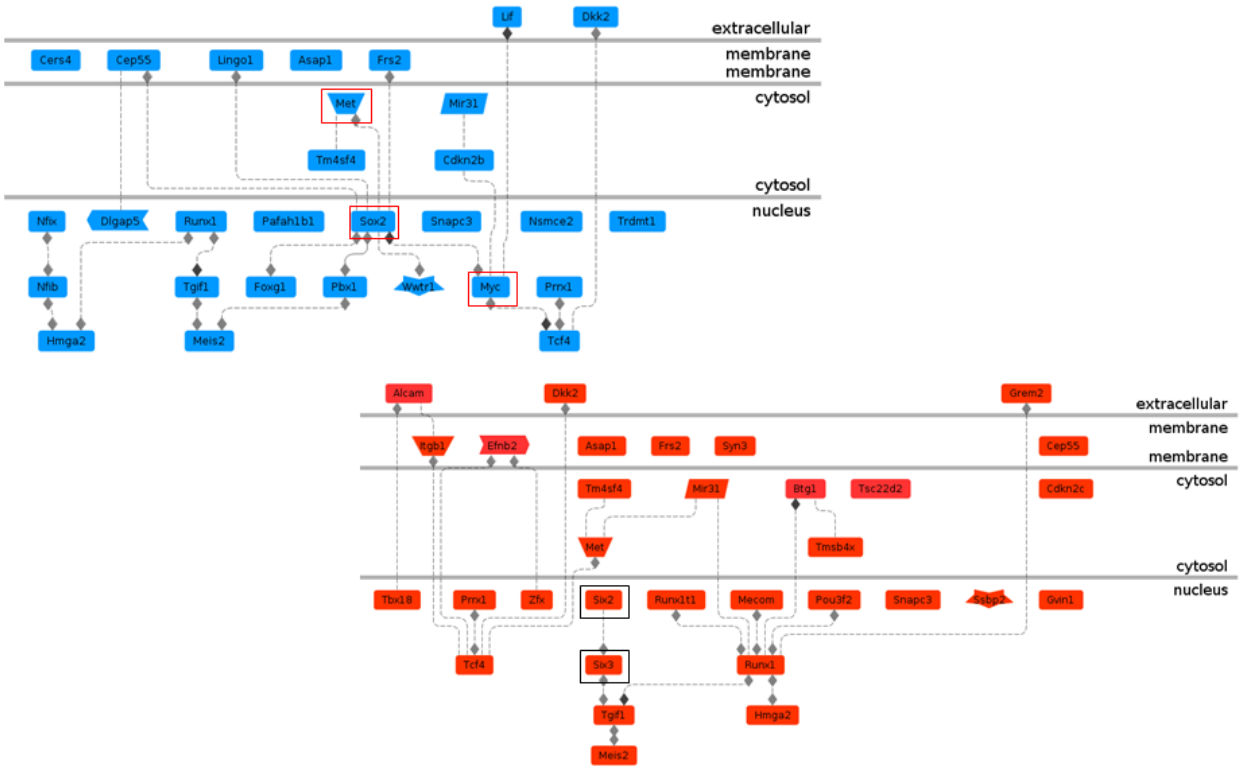


Figure 3.5 Stem cell pathways regulated by sex-specific Brd4-bound enhancers. Blue-labeled genes are male-specific Brd4 bound enhancer regulated genes and red-labeled genes are female-specific Brd4 bound enhancer regulated genes. As depicted, male-specific genes include oncogenes such as Met, Myc and Sox2 while female-specific genes consist of tumor suppressors such as Six 2 and Six 3.

We next integrated our Brd4 Calling Cards, H3K27ac ChIP-seq and RNA-seq data to identify a set of sex-specific, Brd4 enhancer regulated genes. Candidate genes were identified by proximity to both a sex-specific Brd4 binding site and a sexually dimorphic H3K27ac enhancer region, and sensitivity to JQ1 (downregulation). These analyses identified 34 male-specific genes and 30 female-specific genes, which we refer to as sex-specific JQ1-sensitive genes. Similar to our previous pathway analyses, male-specific JQ1-sensitive genes

demonstrated functional enrichment for neoplasm metastasis, tumor angiogenesis, integrin signaling pathway, and metabolic process, in addition to DNA-repair-deficiency disorders. Female-specific JQ1-sensitive genes showed an enrichment in pathways involved in semaphorin signaling, chromosome aberrations, glioblastoma, regulation of transcription, and glucose metabolism disorders **(Table 3.1)**. These results are indicative of sex-specific transcriptional programs regulated by Brd4-bound enhancers. Identifying which specific pathways are critical to sex differences in GBM will require further functional studies.

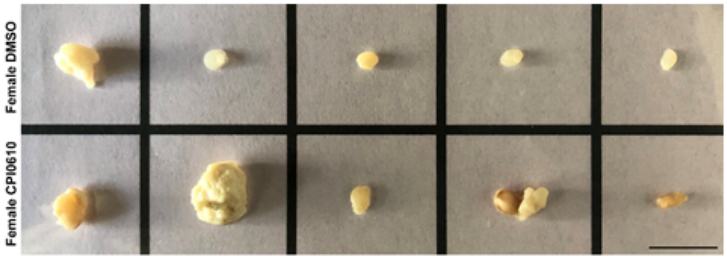
To address the sex-specific effects of BET inhibitors treatment on *in vivo* tumorigenesis, we performed *in vivo* flank implantation studies. Based on our *in vitro* ELDA studies, we chose to use JQ1 and CPI0610, the two BET inhibitors with the most dramatic effect in male and female GBM astrocytes respectively. Each mouse received DMSO, JQ1 or CPI treated male (5000 cells) or female cells (1.5 million cells) injected in the flanks and tumor growth was monitored blindly for 7-12 weeks with thrice weekly micrometer measurements in 3 dimensions. Flank implantation of JQ1-treated transformed male cells produced smaller and less number of tumors than control DMSO male implants **(Figure 3.6)**. This effect was not observed with CPI0610-treated transformed male cells. Importantly, CPI0610-treated transformed female cells produced larger and more tumors than control DMSO female implants **(Figure 3.6)**. This effect was seen following only one dose of BET inhibitors prior to implantation, indicating that this robust response is maintained and manifested at the epigenetic level. Taken together, these results demonstrate for the first time that the sex differences in the tumorigenic phenotype we observe in GBM cells are mediated by differential Brd4-marked enhancers, and that the response to BET inhibition is sex-dependent.

Pathway description	Genes	Adjusted *p*-value	# of Genes
Male GBM Cells: 34 genes			
Glucose metabolism disorders	Nupr1, Rasgrf1, Slc16a7, Dock8, Trhr, Akap6, Sh2d4a, Lgals3, Slc35b3, Fbln1, Ahr, Sucnr1, Ank2, Slc7a11, Btn1a1, Pi15, C1ql3, Pter, Arntl, Cp	1.00E-03	20
Neoplasm Invasiveness	Nupr1, Rasgrf1, Slc16a7, Akap6, Cd109, Lgals3, Csgalnact1, Sema4a, Slc35b3, Fbln1, Ahr, Ank2, Slc7a11, Crabp1, Pter, Bcar3, Arntl, Cp	1.00E-03	18
Glioma	Nupr1, Rasgrf1, Dock8, Trhr, Cd109, Lgals3, Fbln1, Ahr, Ank2, Slc7a11, Crabp1, Pi15, Pter, Arntl, Cp, Snhg18	2.00E-03	16
Cell Regulation of transport	Rasgrf1, Slc16a7, Akap6, Nalcn, Lgals3, Slc35b3, Fbln1, Ank2, Slc7a11, Clic6, Btn1a1, Crabp1, Arntl, Cp	1.00E-03	14
Cell proliferation	Nupr1, Rasgrf1, Dock8, Cd109, Lgals3, Fbln1, Ahr, Btn1a1	2.00E-03	8
Female GBM Cells: 30 genes			
Inflammation	Sema3a, Adm, Slco1a5, Inhba, Cxcl1, Sema3e, Dkk2, F2rl1, Bmpr1b, Clmp, Has2, Map4k3, Cmah, Igf1, Steap2, F3	0.00E+00	16
Glioma	Sema3a, Adm, Pcdh18, Cxcl1, Sema3e, Dkk2, F2rl1, Bmpr1b, Fblim1, Has2, Cmah, Igf1, Otx2, F3	1.00E-03	14
Programmed cell death	Sema3a, Adm, Inhba, Bmpr1b, Map4k3, Igf1, F3	3.00E-03	7
Cell differentiation	Sema3a, Adm, Inhba, Bmpr1b, Has2, Igf1	2.00E-03	6
Growth factor activity	Inhba, Cxcl1, Igf1	0.00E+00	3

* Subset of enriched pathways for sex-specific genes in male and female GBM cells

Table 3.1 Pathway analysis for sex-specific highly Brd4-bound genes downregulated following JQ1 treatment male and female GBM cells.

A



B

	JQ1			CPI0610		
	DMSO	JQ1	*p*-value	DMSO	CPI0610	*p*-value
Male	9/10	2/9	**0.0055**	5/5	4/4	
Female	1/9	1/9		0/4	4/5	**0.0476**

Figure 3.6 BET inhibitors have opposing effects on in vivo tumorigenicity in male and female GBM astrocytes. (A) Representative flank tumors from DMSO or CPI0610 treated transformed female GBM astrocyte-initiated tumors. **(B)** Male tumors were significantly inhibited in their *in vivo* growth following JQ1 treatment compared to DMSO treatment (Chi-Square Fisher's exact test $p < 0.0055$). CPI0610-treated transformed female cells significantly produced larger and more tumors than control DMSO female implants (Chi-Square Fisher's exact test $p < 0.0476$).

In conclusion, these data demonstrate for the first time that cell intrinsic sex identity is mediated by sex differences in Brd4-marked enhancer usage. In the model utilized here, the differential usage of Brd4-bound enhancers mediated important sex differences in a GBM tumorigenic phenotype. Additionally, the response to BET inhibitors treatment was shown to be sex-dependent. Altogether, the consistency between our mouse data, the GBM TCGA data, and published breast and prostate cancer studies, provides strong evidence for the context-

dependent, sex-specific dual role of Brd4 in regulating gene expression programs in oncogenesis. Since bromodomain inhibitors are currently being evaluated in a number of clinical trials, understanding this phenomenon is of critical importance, both for the interpretation of existing trials and to guide better application of these drugs. Increasing our knowledge of these sex-specific genetic and epigenetic mechanisms will lead to a greater understanding of cancer biology and its relationship to normal development, as well as identify novel therapeutic targets to improve outcome for all patients with GBM and potentially other cancers that exhibit substantial sex differences in incidence or outcome.

3.5 Methods

3.3.1 RNA-sequencing

Male and female GBM cells (Nf-/-;DNp53 astrocytes) were generated as previously reported (Sun et al., 2015) and grown in DMEM/F12 media supplemented with 10% FBS and 1% penicillin-streptomycin. Total RNA was isolated from male and female GBM cells that were treated with DMSO (0.05%) or JQ1 (500 nM for 24 hours) using the RNeasy Mini Kit from Qiagen, following the kit protocol (Hilden, Germany). PolyA Selection was performed to create RNA Seq libraries. Cell mRNA was extracted from total RNA using a Dynal mRNA Direct kit. The quantity of RNA was measured using a spectrophotometer (NanoDrop 2000c; Thermo Scientific). Samples with an RNA concentration A260/A280 ≥1.8 ng/µl, and purity A230/A260 ≥ 2.0 ng/µl were selected. The Agilent 2100 Bioanalyzer was used to determine the RNA integrity number. The degradation level was identified using the RNA 6000 Nano LabChip kit (Agilent). Samples with RNA integrity number > 9.8 were further processed using

TruSeq mRNA Library Preparation Kit (Illumina) and then sequenced on a HiSeq 25000 (Illumina).

3.3.2 Chromatin immunoprecipitation sequencing (ChIP-seq) for H3K27ac

Male and female GBM cells were grown to 50% confluence (200K cells/T25 flask). On the day of transfection and 30 minutes before applying complexes, media was replaced with DMEM without FBS and antibiotics. Transfection complexes containing 3 μg of PB helper DNA with a PB-Brd4 fusion and/or 3 μg of PB donor DNA with a puromycin selection marker and Lipofectamine LTX/Plus Reagent (Invitrogen) were applied to cells for 12-18 hours at a culturing condition of 37 °C and 5% CO_2. Cells were then allowed to recover in fresh medium containing DMEM/F12 with 10% FBS and 1% penicillin-streptomycin for another 24-48 hours. Cells were then expanded into T75 flasks and selected using puromycin at a concentration of 2.5 μg/ml for 3 days. DNA was extracted from puromycin resistant cells and processed by transposon calling card protocol as previously described [2]. Briefly, DNA sample was divided into three 2-mg aliquots, each digested by MspI, Csp6I, or TaqI individually. Digested DNA was ligated overnight at 15° C in dilute solution to encourage self-ligation. After ethanol precipitation, self-ligated DNA was resuspended in 30 ml ddH2O and used as template in an inverse PCR. Primers that anneal to PB donor sequences were used to amplify the genomic regions flanking PB, and adapter sequences that allow the PCR products to be sequenced on the Illumina genome analyzer were added. The PCR products were purified using the QIAquick PCR purification kit (Qiagen) and diluted into 10 nM concentration. For each sample, the same amount of PCR product from digestion with each restriction endonuclease was pooled and submitted for Illumina sequencing (HiSeq 2500).

3.3.3 Sequencing data alignment and analysis

RNA-seq data sets were aligned to the transcriptome and the whole-genome with STAR [98]. Genes or exons were filtered for just those that were expressed. A gene count table for all the genes in each sample was generated using HTSeq [99]. Differential gene expression between pairs of samples was computed using DESeq2 and was filtered by FDR < 0.05 for differentially expressed genes [130].

ChIP-seq data sets for H3K27ac were aligned to the murine genome build mm10 using Bowtie2 and only uniquely aligning reads were used [131]. Regions of enrichment of H3K27ac over background were calculated using the MACS version (2.1.0) peak finding algorithm [127]. An adjusted *p*-value threshold of enrichment of 0.01 was used for all data sets. The resulting peak files were used as inputs for DiffBind (version 3.5) to derive consensus peak sets [132, 133]. The differential enrichment of H3K7Ac signals between male and female analysis was carried out with Diffbind using DESeq2 (method = DBA_DESEQ2) with libraries normalized to total library size.

Transposon calling cards data sets were aligned to the murine genome build mm10 using Bowtie2 [131]. The Brd4 binding site was identified by using an established algorithm [2]. Briefly, the Brd4-directed PB insertions were clustered using a hierarchical clustering algorithm to identify insertions within 2500 bp and then modeled as a Poisson distribution, with the number of independent insertions in the "transposase-alone" experiment in the same genomic window setting the expectation. The *p*-value is then calculated from cumulative distribution function given the observed number of independent insertions in the Brd4-directed experiment.

To identify the sites with an excess of Brd4 insertions in male GBM cells relative to female cells, we used the algorithm from ChIP-Seq peaks caller MACS [127] but modified for the analysis of calling card data. First, the transposon insertions were grouped by hierarchical clustering. Then, peaks with an excess of insertions in the male sample were identified by computing lambda, the number of insertions per TTAA expected from the female sample by taking the maximum lambda calculated from the number of insertions in the female peak, or in a 1kb, 5kb or 10kb window centered at that peak. We computed a *p*-value based on the expected number of insertions (lamda × number of TTAAs in peak × number of insertions in peak). To identify the sites with an excess of Brd4 insertions in female GBM cells relative to male cells, we performed the same analysis, substituting the male and female data sets.

3.3.4 Pathway analysis

Pathway enrichment analysis for differentially regulated genes was performed using a combination of KEGG pathway and Genomatix Pathway System (GePS). GePS uses information extracted from public and proprietary databases to display canonical pathways and or to create and extend networks based on literature data. These sources include NCI-Nature Pathway Interaction Database, Biocarta, Reactome, Cancer Cell Map, and the ENCODE Transcription Factor project data. All data for pathway analyses are presented with adjusted corrected *p*-values.

3.3.5 Growth assays

Growth kinetics of male and female GBM cells (Nf-/-;DNp53 astrocytes) treated with DMSO (0.05%), JQ1 (500 nM), or RVX-208 (5μM) for 24 hours or shRNAs were examined by

counting live cell number using an automated T4 cell counter as previously described with minor modifications (Sun et al., 2015). Briefly, cells were harvested and plated in a 6-well plate at a density of 2 x 10^4 cells/well (2 technical replicates per treatment/genotype/time point). 4 hours post plating, cells were harvested by trypsinization and counted in the presence of trypan blue. This time point was designated as the starting point (T0) of the time course. Cells were then harvested and counted every 24 hours for a total of 4 days (24, 48, 72 and 96 hours). This experiment was repeated three times.

3.3.5 Clonogenic cell frequency assay: Extreme Limiting Dilution Assays (ELDA analysis)

Clonogenic capacity of male and female GBM cells (Nf-/-;DNp53 astrocytes) treated with DMSO (0.05%), JQ1 (500 nM), or RVX-208 (5μM) for 24 hours or shRNAs was assayed by the Extreme Limiting Dilution Assay (ELDA). The frequency of clonogenic stem cells was evaluated by the cells' ability to form tumor-sphere in low-adherent conditions as previously reported (Sun et al., 2015). Briefly, cells were harvested into a single cell suspension and plated in neurosphere media containing EGF and FGF on 96-well ultra-low attachment plates in a serial dilution ranging from 3000 cells/well to 1 cell/well (3000, 600, 120, 24, 5 and 1 cells; n=14/cell density). Sphere formation was measured 7 days after plating. Clonogenic stem-like cell frequency was analyzed using the Extreme Limiting Dilution Analysis (http://bioinf.wehi.edu.au/software/elda/).

3.3.6 *In vivo* tumorigenesis: flank implantation

Flank tumors were generated by implanting GBM astrocytes at various cell numbers subcutaneously into left and right side flanks (top and bottom). These cells were treated with

EGF for one week (50ng/ml) followed with a 0.05% DMSO or 500 nM JQ1 treatment for 24 hours. 1 million, 500,000, 100,000 or 5000 cells were then harvested and resuspended in 100 µl of 1:1 media to matrigel (BD Biosciences) and injected into the four flanks of mice. Mice were monitored weekly and tumor growth was monitored blindly for 7-8 weeks with thrice weekly micrometer measurements in 3 dimensions. Animals were used in accordance with an animal studies protocol (no. 20150177) approved by the Animal Studies Committee of the Washington University School of Medicine per the recommendations of the Guide for the Care and Use of Laboratory Animals (NIH).

3.3.7 shRNAs lentiviral infection and knockdown of Brd2 and Brd4

Brd2 and Brd4 knockdown lines were generated by infecting male and female GBM cells (Nf-/-;DNp53 astrocytes) with lentiviral shRNAs against Brd2 or Brd4. We used a pool of 5 shRNAs for Brd4 as well as Brd2. Both Brd2 and Brd4 Knockdown lines were selected with puromycin (2.5 µg/ml) in media for 1-2 weeks and the survivors were expanded for downstream target knockout analysis.

3.3.7 Quantitative Real-Time PCR

Total RNA was isolated using Trizol RNA extraction method (Invitrogen, CA) from male and female GBM cells (Nf-/-;DNp53 astrocytes) infected with shRNAs lentivirus against Brd2 or Brd4. cDNA was generated using the QuantiTect Reverse Transcription Kit (Qiagen). Quantitative RT-PCR was performed using gene-specific primers and iTaq SYBR Green PCR master mix (Biorad, CA). Data was analyzed by standard ΔCq method (2-ΔΔCq) where ΔCq is the difference between the gene of interest and GAPDH control Cq value.

3.3.7 Statistical analysis

All experiments in this study were carried out at least three times. ANOVA and two-tailed Student's t-test were used to compare the differences in all functional measurements between JQ1, RVX-208 and control group (DMSO), and a p-value < 0.05 was considered statistically significant.

To test if upregulated genes in male cells compared to female cells and vice versa are significantly enriched for H3K27ac binding, the normalized H3K27ac signal intensities (reads) from 1kb upstream of the gene start site to 1kb downstream of the end of the gene were used for the correlation analysis. A paired Mann-Whitney-Wilcoxon test was used to compare normalized H3K27ac signal intensities between male and females cells and a p-value less than 0.01 was considered statistically significant.

To investigate if Brd4-proximal genes are significantly downregulated upon JQ1 treatment compared to Brd4-distal genes, we first defined Brd4 proximal genes as the closest genes to Brd4 binding sites and Brd4 distal genes as genes located near sites that are not enriched for Brd4 binding sites(353 male genes and 292 female genes). The expression profiles before and after JQ1 treatment of Brd4 proximal and distal genes were used for a Mann-Whitney-Wilcoxon test for male and female respectively and a p-value less than 0.01 was considered statistically significant.

3.3.7 TCGA human GBM data analysis

Level 3 RNA-seq gene expression data for TCGA GBM samples were obtained from the Broad GDAC Firehose data portal. Clinicopathologic data for the GBM samples were obtained

from the cBioPortal for cancer genomics. Only tumor samples that represented primary tumors were used and recurrent tumor samples were excluded from the analysis. In total, gene expression and clinicopathologic data of 98 males and 53 females were used in this analysis.

To identify sex differences in overall survival (OS) outcomes for Brd4 in GBM patients, male and female patients were stratified into different expression groups and differences in survival outcomes among groups were then assessed [116]. The Brd4 gene expression value was first transformed into a z-score that was specific to the sex of the patient, and the patients were grouped into high expression group ($Z > 1.0$) and low expression group ($Z < -1.0$). The patient groups of each sex were associated with survival endpoints by the Kaplan-Meier method, and log-rank test was used to compare survival difference among groups. High expression and low expression groups were analyzed separately.

3.3.7 Data availability

All Illumina sequencing reads and processed file have been deposited in the Short Read Archive/GEO database (http://www.ncbi.nlm.nih.gov/geo/).

Chapter 4
High-throughput functional identification of enhancers using a CRE recombinase-mediated reporter assay

4.1 Abstract

Enhancers are DNA sequences that regulate the timing, tissue-specificity, and level of gene expression. Developmental enhancers are critical determinants of normal organismal development and cellular differentiation [134], and changes in their functions likely underlie biological processes ranging from cell fate decisions to disease susceptibility [37, 46, 135]. Despite their importance, genome-wide discovery of developmental enhancers remained challenging [136]. Here, we developed a CRE recombinase-mediated method for the functional identification of active enhancers at different time periods of development, named as Developmental Enhancer Sequencing (DevEn-seq). Genomic integration of a single copy enhancer and a CRE reporter assay are coupled with high-throughput sequencing to enable parallel screening and tracking of large numbers of enhancer candidates. We demonstrated that DevEn-seq is able to detect enhancers more efficiently than regular reporter methods and trace enhancer activities without being disturbed by the gene silencing effect caused by lentiviral sequences. By functionally interrogating >500 kilobases (kb) of mouse sequence in mouse embryonic stem cells for enhancer activity, we identified 38 enhancers at pluripotency loci with a positive predictive value of 80%. With an *in vitro* neural differentiation protocol, we identified two neural progenitor-

specific enhancers around HB9 and Olig2 genes. DevEn-seq is a high throughput method for functional identification and tracking of enhancer activities throughout cellular differentiation.

4.2 Introduction

The ability to create distinct cell types is fundamental for the development of multicellular organisms. Since all cells in an organism contain the same genes, cellular diversity is achieved through the distinct regulatory DNA elements or enhancers throughout the genome that control the spatial and temporal expression patterns of specific sets of genes, leading to a cell's specialized role[134]. Developmental enhancers are frequently tissue-specific or cell type-specific[35-37] and provide the genomic landscapes required for the binding of key TFs that drive cell-specific gene expression programs[135]. Despite their importance, developmental enhancer discovery has remained challenging and their regulatory role has not been characterized in detail because understanding these regulatory mechanisms requires the global functional identification of differentially active enhancers in development[137].

Genome-wide approaches for enhancer identification are mostly based on the indirect assessment of enhancer-associated markers such as histone modifications and binding of transcription factor p300 by ChIP-seq or DNase digestion followed by sequencing (DNase-seq)[138]. Currently, there are two general strategies for the genome-wide prediction of enhancers: 1) Deep sequencing of DNase I-hypersensitive sites (DNaseI-HS seq) [139] or formaldehyde-assisted isolation of regulatory elements sequencing [140] (FAIRE-seq), which allow for the mapping of open chromatin; 2) Chromatin immunoprecipitation followed by deep sequencing [141] (ChIP-seq) enables the detection of regulatory (e.g., TF or cofactor) binding sites and enhancer associated histone modifications [e.g., histone 3 (H3) Lys4 monomethylation (H3K4me1) or H3

Lys27 acetylation (H3K27ac)]. These methods have provided an expanded view of chromatin landscapes as well as insights into the relationship between active enhancers and the dynamics of chromatin modification and remodeling. However, one limitation of these approaches is that they are associated with false positive and false negative errors because they lack a direct functional or quantitative readout of enhancer activity and putative enhancers predicted this way requires further functional validation by reporter assays[142, 143]. In addition, the genomic loci defined by these marks typically span several kilobases (kb) and are generally too broad to define the specific DNA sequences mediating enhancer function.

Functional approaches for identifying active enhancers rely on functional assays of individually transfected reporter plasmids harboring putative regulatory regions[144], but most of these assays do not scale to the millions of tests required for global identification of enhancers. Recently, several groups have developed massively parallel reporter assays (MPRAs) that permit the simultaneous analysis of hundreds of thousands of reporter plasmids and, thereby, a functional assessment of the transcriptional-activation properties of large numbers of genomic regions [142, 143, 145]. However, MPRAs have primarily been used as a way to dissect the functional components of previously identified transcriptional regulatory DNA elements rather than as a tool for the discovery or screening of enhancer in mammalian cells. Even with these advances, the enormous size and complexity of mammalian genomes, and the concomitant number of required reporter plasmids, remain among the primary challenges to using functional approaches for the *de novo* discovery of enhancers.

More recently, two high-throughput functional approaches for enhancer identification have been developed: self-transcribing active regulatory region sequencing (STARR-seq) and enhancer-FACS-seq (eFS) [146, 147]. In STARR-seq, sheared DNA fragments are inserted within a

noncoding portion of the reporter plasmid transcription unit, and enhancer activity is detected in transfected cells by the ability to self-transcribe, permitting identification of 5,499 elements [146]. eFS instead is a highly parallel functional screen for identifying developmentally relevant, mesodermal-specific enhancers within developing Drosophila embryos [147]. Although both STARR-seq and eFS are elegant systems for enhancer discovery in Drosophila, their successful application to the analysis of large mammalian genomes is uncertain. Either human or mouse genome is roughly 23 times larger than that of D. melanogaster, suggesting that initial STARR-seq libraries would require over 200 million unique plasmids, and a preliminary study using STARR-seq analysis of human DNAs identified only six enhancers from a plasmid library consisting of 1.3-million unique genomic regions derived from 1 Mb of human DNA. In addition, STARR-seq relies on transient delivery of enhancer-reporter plasmids, limiting the use of these methods to a small number of easily transfected cell types. Furthermore, the relevance of STARR-seq identified enhancers to endogenous gene expression is less clear, as many correspond to closed chromatin regions. eFS focuses on potentially relevant portions of the genome, but an eFS-like approach using transgenic mouse embryos for screening genomic segments would be prohibitively expensive for most researchers.

Here, we developed a high-throughput functional method to identify enhancers and track their activity changes at different stages of cellular development. We first utilized a CRE-reporter embryonic stem cell (ESC) line that has a single copy transgene integrated at the Rosa26 locus of the mouse genome[148]. The transgene contains a strong and ubiquitously expressed CAG promoter followed by a loxP-flanked (‘floxed’) stop cassette-controlled red fluorescent gene (RFP) (Fig. 4.1). We position the enhancer fragments upstream of a CRE recombinase gene in a lentiviral transfer vector to produce lentivirus, which permits the individual assessment of

enhancer activity of millions of fragments at the lentiviral integrated genomic loci. If the integrated lentiviral transgene harbors an active enhancer that is able to drive the CRE expression, CRE will flip out the stop cassette and in turn activate the expression of RFP within the CRE-reporter ESC (Fig. 4.2). The ubiquitously expressed RFP at Rosa26 locus then becomes a permanent marker for that enhancer element activity in the lentiviral transgene and survives the cellular division and differentiation. After *in vitro* differentiation of the transduced CRE-reporter ESCs, the RFP positive and negative cells along each cell lineage are collected by fluorescence-activated cell sorting (FACS). The enhancer elements within the RFP positive cells and the corresponding non-enhancer elements in RFP negative cells are recovered by PCR and sequenced by Illumina HiSeq, which produces a genome-wide map of lineage-specific enhancers in development.

To control the tracing time window, we created a chimeric protein consisting of CRE fused to the ligand binding domain of the estrogen receptor (ERT2, Fig. 4.2). The fusion protein is constitutively expressed but sequestered outside of the nucleus by binding to the cytoplasm-anchored protein HSP90. The treatment of ER antagonists (4-hydroxy tamoxifen; 4OHT) causes the fusion protein to translocate into the nucleus and recombine the loxP sites[16]. To obtain a high temporal resolution of the enhancer activity along development, we used a degradation domain of ERT2 that confers short protein half-life and attached it to both N and C terminus of the CRE (dERT2:CRE:dERT2). The lentiviral transgenes carrying candidate enhancers corresponding to each time window were recovered and sequenced.

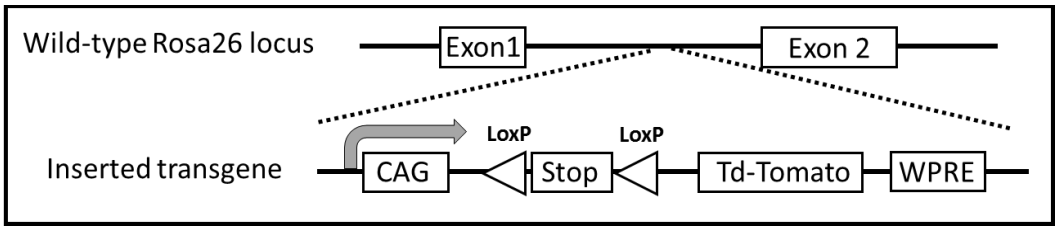


Figure 4.1 Schematic diagram of the gene targeting strategy to generate CRE-reporter cell line. The CRE-reporter cassette was inserted into the Rosa26 locus, in the intron between endogenous exons 1 and 2.

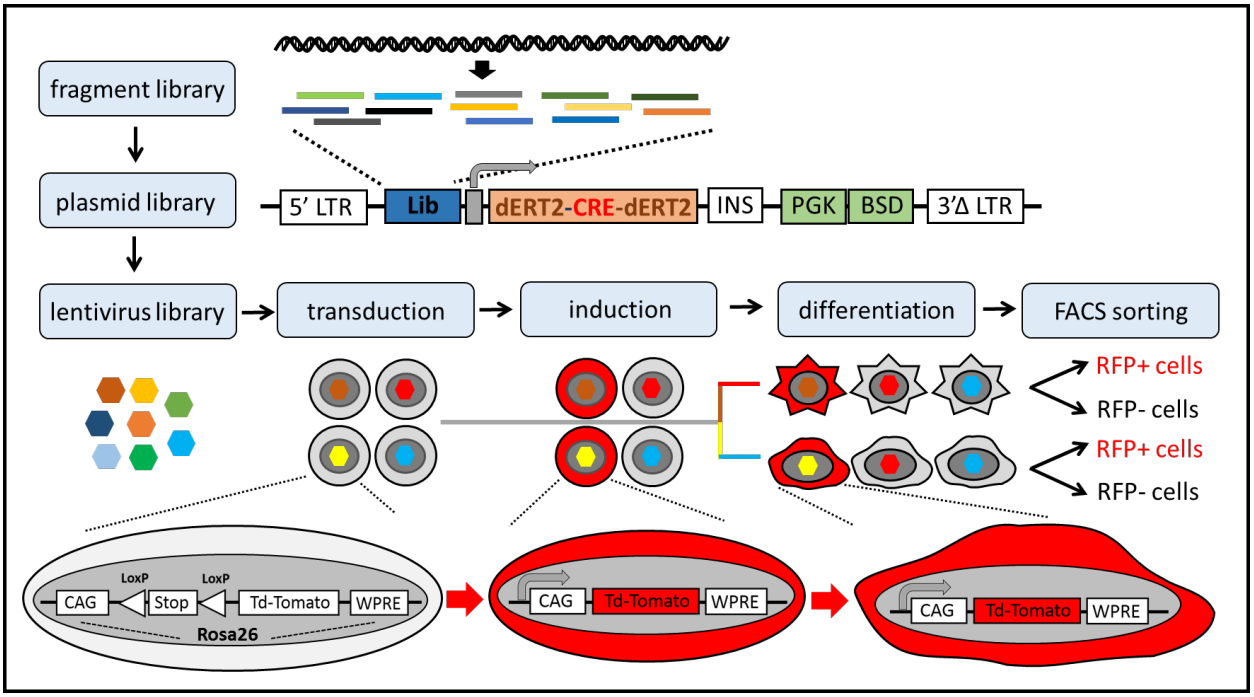


Figure 4.2 Overview of CRE-mediated enhancer screen. DNA test fragments containing putative enhancers are linked to an inducible CRE and then are integrated into genome by lentiviruses. After drug selection, each cell in the library contains no more than one candidate fragment. If the enhancer is active in a cell, CRE recombinase will be expressed and remove a stop codon in an RFP reporter and permanently generate red cells in itself and its progeny. RFP-

expressing cells are isolated by flow cytometry, and presumptive enhancer sequences from these cells are amplified and sequenced. ES cell libraries can be differentiated *in vitro* before sorting.

4.3 Results

4.3.1 Active enhancers can be efficiently detected using a CRE reporter assay.

Current methods for functional identification of enhancers use a basal vector in which a minimal promoter is positioned in front of a reporter. The enhancer fragments are cloned upstream of this minimal promoter to test their transcriptional activity. To develop our CRE-mediated enhancer identification methodology, we first sought to engineer a minimal promoter from which the basal expression of the CRE protein in the absence of an enhancer should be below the level of recombination required to trigger reporter (RFP) expression. A total of five minimal promoters that have been successfully applied for fluorescence or LacZ-based enhancer detection were evalutated: heat shock protein 68 (HSP68)[149], β-globin (HBB)[150], oncogene c-fos (cFOS)[151], fibroblast growth factor 4 (FGF4)[152], and super core promoter (SCP)[146]. These minimal promoters were cloned upstream of CRE with and without CMV enhancer (CMVE). The CRE-reporter ESCs were transfected with these constructs and the percentage of RFP positive cells were measured by FACS. As shown in Fig 4.1, the basal transcriptional activity of all five minimal promoters is different: SCP, HBB, and FGF4 are all below 0.5% while cFOS and HSP68 are 2% and 5%, respectively. The addition of CMV enhancer increases the RFP positive cells significantly for all five minimal promoters and the percentage ranges from 30% to 45% ($p>0.05$). For an enhancer identification assay, a lower basal activity and higher signal are optimal. Therefore, we calculated the induction fold change as the ratio of RFP positive cells observed in the CMV enhancer-driven construct to those in the corresponding non-enhancer-

driven construct and chose FGF4 minimal promoter for the CRE-mediated assay because its ratio is the highest among all five.

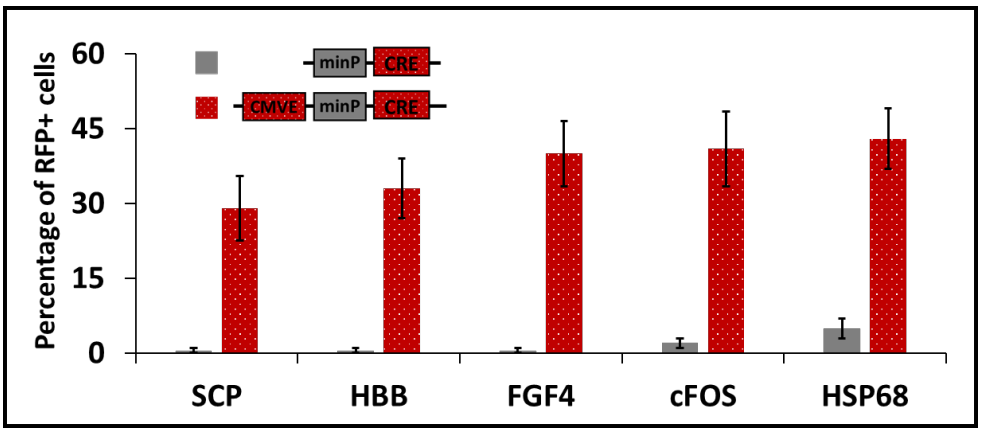


Figure 4.3 Optimization of minimal promoters for the CRE-mediated enhancer identification. Five minimal promoters were cloned upstream of CRE gene with (red) and without (grey) a CMV enhancer (CMVE).

4.3.2 CRE activity can be induced by a degradation domain of ERT2

To assess the enhancer activity in a specific time window during a differentiation process, we created a chemically inducible CRE by fusion to a regular or degradable ERT2 (dERT2) domain. The ERT2 fusion protein is sequestered outside of the nucleus by binding to the cytoplasm-anchored protein HSP90 and is released from the inactive complex in the presence of 4-OHT[153]. To have an optimal inducible CRE, three CRE variants were created. Single and double copies of the ERT2 domain were fused to CRE (ERT2:CRE and ERT2:CRE:ERT2), and double copies of degradation domain of ERT2 (dERT2) were fused to CRE (dERT2:CRE:dERT2). These CRE variants were evaluated on two parameters: 1) the enzymatic activity of the fusion protein relative to that of the unfused one and 2) the fold induction between induced and non-induced states. As shown in Fig 4.2, three CRE variants have similar activities

to unfused CRE ($p>0.05$), suggesting that the attached ERT2 domain does not significantly affect the CRE activity, regardless of the N and C-terminal fusions. In terms of inducibility, ERT2:CRE:ERT2 and dERT2:CRE:dERT2 fusion proteins showed similar induction fold and a much higher than the CRE:ERT2 variant ($p<0.05$), indicating that two copies of the ERT2 domains increase the inducibility significantly than single copy. Taken together, ERT2:CRE:ERT2 and dERT2:CRE:dERT2 fusion proteins both have good enzymatic activity and inducibility and are therefore optimal candidates for CRE-mediated reporter assay.

We next evaluated the half-life of the fusion protein dERT2:CRE:dERT2 and ERT2:CRE:ERT2 because the rate at which the inducible CRE degrades defines the temporal resolution of the CRE-mediated reporter assay. To have a high temporal resolution, the inducible CRE synthesized earlier should degrade quickly to an extent that CRE "carryovers" would not interfere with newly synthesized CRE upon induction. To compare the half-life of the fusion protein of ERT2:CRE:ERT2 and dERT2:CRE:dERT2, mRNAs from both CRE variants were synthesized by *in vitro* transcription and transfected into the CRE reporter ESCs treated with cell cycle inhibitor (abl kinase inhibitor). Chemical inducer 4OHT was added into the culture medium to activate the CRE-mediated system at different time points ranging from 0h to 72h posttransfection and the percentage of RFP positive cells was measured. As shown in Fig 4.2 B, the percentage of RFP positive cells transfected with dERT2:CRE:dERT2 mRNA drops significantly compared to those transfected with ERT2:CRE:ERT2 mRNA, suggesting that the half-life of dERT2:CRE:dERT2 is shorter than ERT2:CRE:ERT2. This result is expected considering that degradation domain is unstable[154] and demonstrates that dERT2:CRE:dERT2 is able to offer a higher temporal resolution for the CRE induction system.

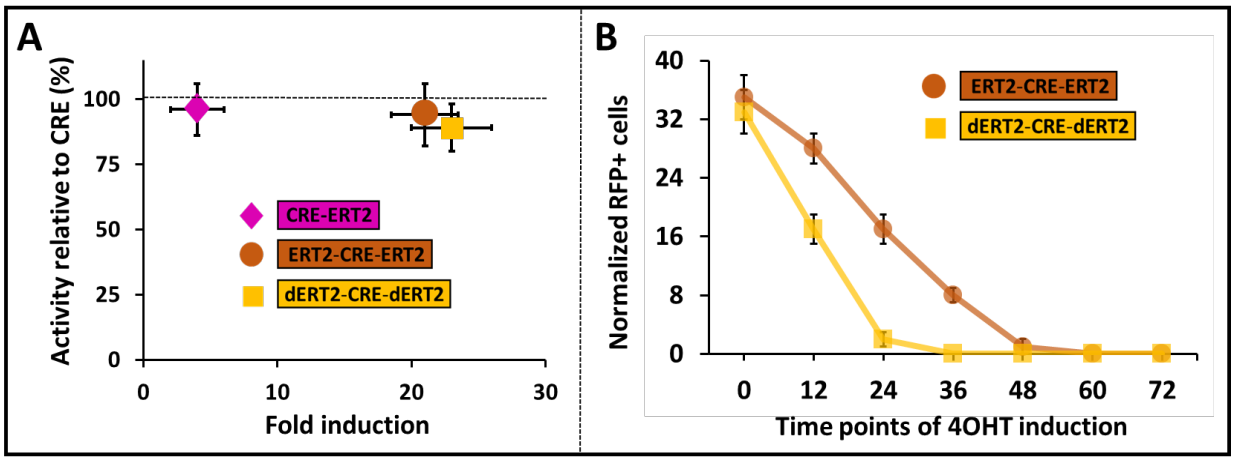


Figure 4.4 Development of an inducible CRE system a degradation domain of ERT2. (A) Evaluation of the inducible CRE systems made from different CRE-ERT2 fusion variants. (B) Estimation of the degradation time of inducible CRE variants.

4.3.3 CRE-mediated enhancer identification has a higher sensitivity than a regular fluorescence-based assay

We next investigated whether enhancers placed upstream of the FGF4 minimal promoter and CRE variants could be efficiently identified compared to a regular reporter assay which places enhancer fragments upstream of fluorescent protein. We selected three active enhancers: SP1 enhancer (SP1E), CMV enhancer (CMVE), SV40 enhancer (SV40E) and one non-enhancer (NE) marked by intensely dense chromatin in a gene desert region. CRE reporter ESCs were transiently transfected with the enhancer-driven and non-enhancer-driven constructs and the percentage of RFP positive cells were measured. As shown in Fig 4.3, non-enhancers showed no signal using either method, while active enhancers showed a significantly higher percentage of positive cells in all variants of CRE-mediated enhancer identification than a regular RFP based

assay ($p<0.05$), indicating that CRE-mediated enhancer identification has a higher sensitivity than a regular fluorescent-based assay.

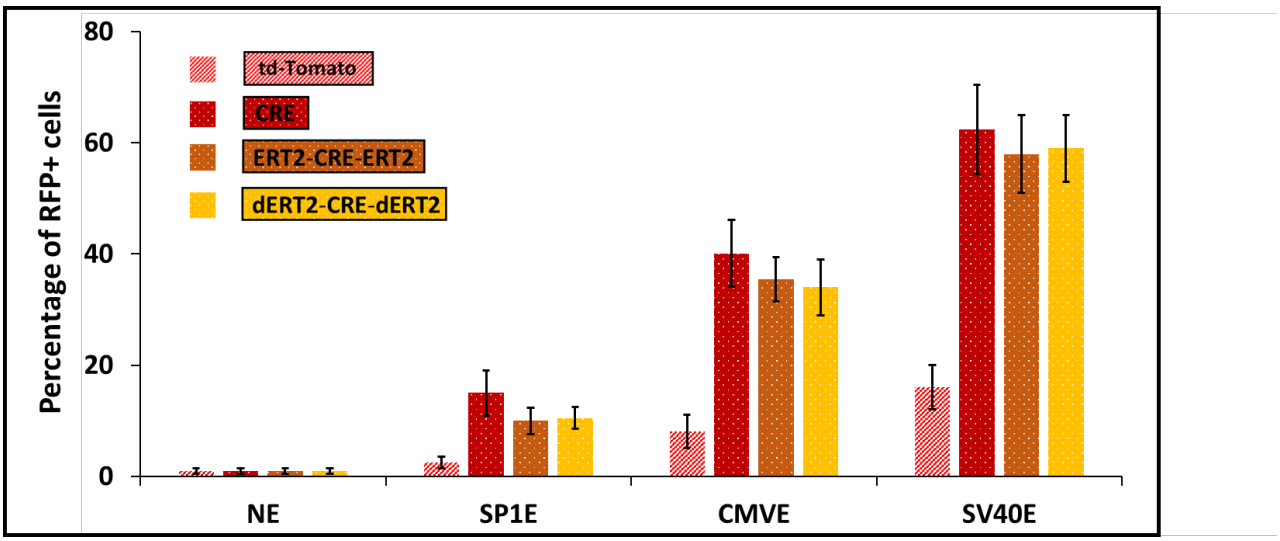


Figure 4.5 Comparison of the regular fluorescent based enhancer identification with CRE-mediated enhancer assay. One non-enhancer sequence (NE) and three enhancer sequences (SP1E, CMVE and SV40E) were cloned upstream of RFP reporter gene to measure their transcriptional activity. In the meanwhile, these sequences were positioned upstream of three CRE variants individually to tested enhancer activity. The readout is the percentage RFP positive cells for both methods.

4.3.4 Neural differentiation of CRE-reporter ESCs generates motoneurons, astrocytes, and oligodendrocytes

To develop an *in vitro* cell culture differentiation system, we adapted a protocol from the Sakiyama-Elbert lab to differentiate the ESCs to neural lineages: motorneurons, oligodendrocytes, and astrocytes[155, 156]. This *in vitro* cell differentiation protocol consists of culturing whole embryoid body (EB) aggregates on gelatinized plates in DFK5 media containing retinoic acid (RA) and sonic hedgehog agonist (SAG), dissociating induced EBs (day 6), and

plating the single cells on poly-ornithine plates for 7-10 days (Fig 4.4 A). The fully differentiated cells were stained with cell-type specific antibodies against beta-tubulin class II to mark neurons, Olig2 to mark oligodendrocytes, or GFAP to mark astrocytes. We observed all three neural lineages in roughly in equal amounts (Fig4.4 B), suggesting we are able to obtain an unbiased representation for each cell type in a single differentiated culture.

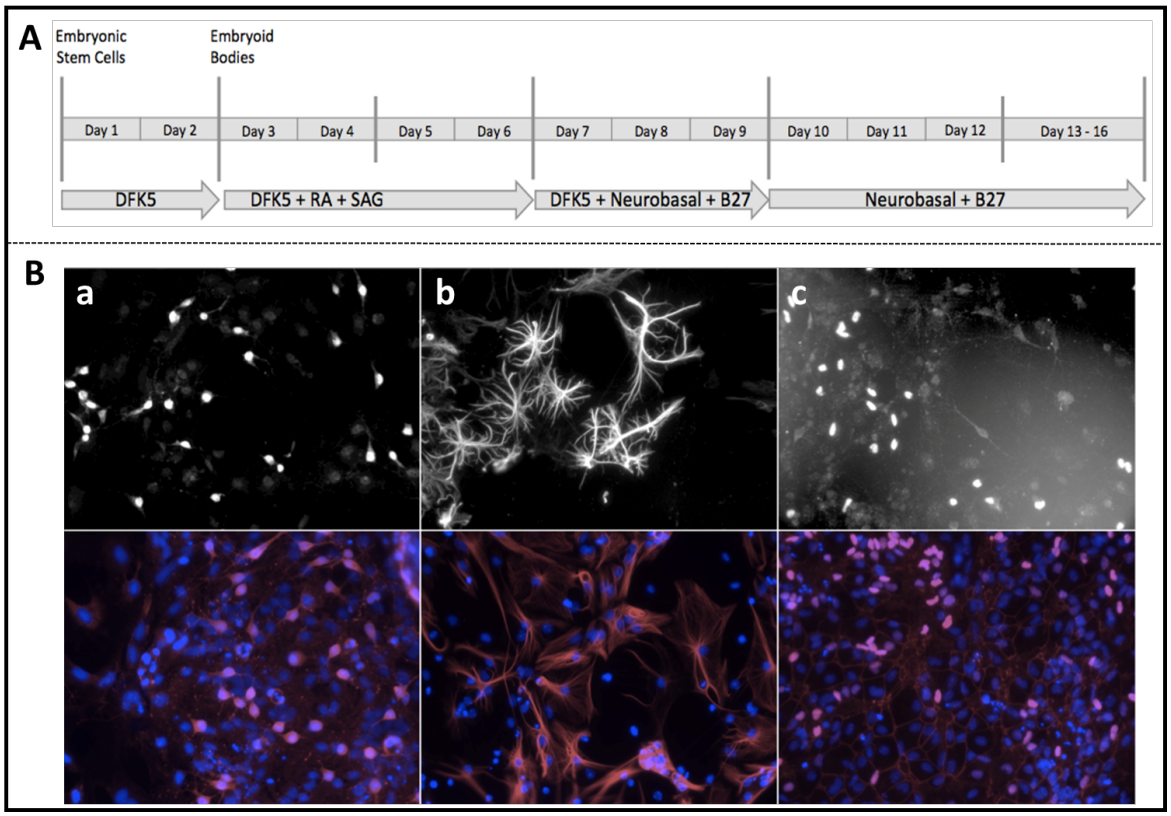


Figure 4.6 Neural differentiation of CRE-reporter ESCs. (A) Timeline of the neural differentiation protocol. (B) Antibody staining of neurons, astrocytes, and oligodendrocytes. Differentiated ESCs were stained with antibodies against beta-tubulin class II to mark neurons (a), GFAP to mark astrocytes (b), or Olig2 to mark oligodendrocytes (c). Nuclei were stained with DAPI. Phase contrast and fluorescent images were positioned at the top and bottom, respectively.

4.3.5 Enhancer activity can be tracked during neural differentiation of ESCs

To test if CRE-mediated enhancer identification can track a given enhancer's activity through neural development, we selected three genes that are transcriptionally active at three different stages of neural differentiation: Nanog, Nestin, and ChAT(choline acetyltransferase), corresponding to the ESC, embryoid body (EB), and mature neuron stages, respectively[90, 157-159]. The enhancer element of each candidate gene was previously identified[160-162]. We then PCR-amplified these enhancers and individually positioned them upstream of dERT2:CRE:dERT2 in a basal lentiviral screen vector. We also included EF1 promoter and non-enhancer sequence as positive and negative controls. CRE reporter ESCs were transduced with these lentiviruses individually and selected under blasticidin S deaminase (BSD) for 5 days to obtain polyclonal cell populations for each transduction. The polyclonal cell populations were further induced with 4-OHT at three developmental stages: ESC, EB, and mature neuron respectively as shown in Fig 4.5. The cells transduced with positive and negative enhancer sequences showed consistently high and low percentages of RFP positive cells in all three developmental stages, respectively, indicating validity of the assay (Fig 4.5). At the ESC stage, sequence from Nanog showed a significantly higher RFP positive percentage than any other enhancers ($p<0.05$). At EB and mature neuron stages, cells with Nestin and CHAT enhancer sequences demonstrated the highest percentages of RFP positive cells, respectively ($p<0.05$). These results are consistent with previous findings[160-162] and suggest that enhancers active at different stages or time periods can be efficiently detected by the CRE-mediated assay.

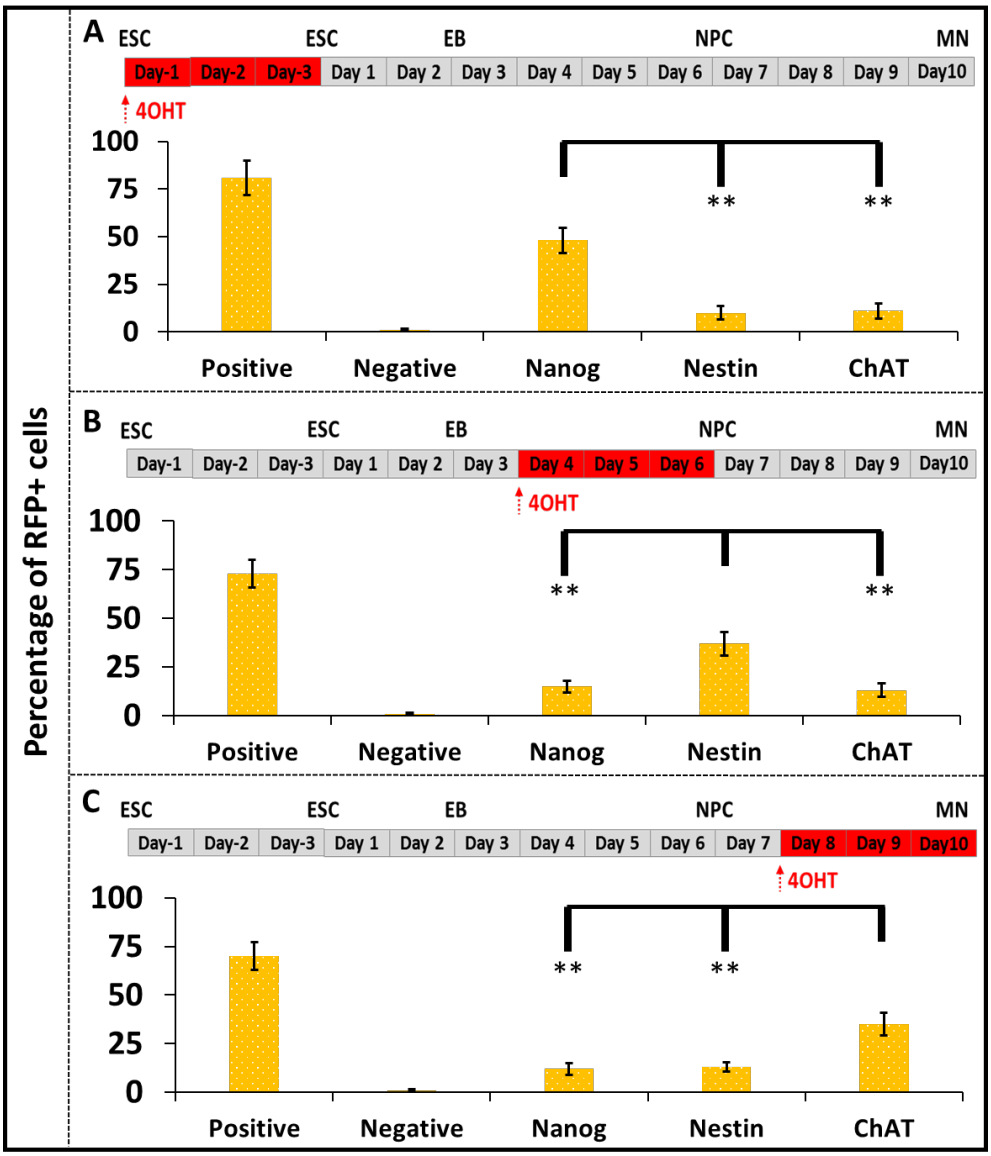


Figure 4.7 Enhancer strength assayed at three stages of the neural differentiation. Positive, negative and three enhancer sequences (Nanog, Nestin, and ChAT) were tested by CRE-mediated enhancer assay at the three stages of neural differentiation: (A) ESC, (B) EB and (C) MN. The RFP positive cells were measured at each stage. "**" indicates a p value less than 0.01.

4.3.6 DevEn-Seq accurately identifies mouse ES cell enhancers

To identify mouse ESC enhancers, we constructed an enhancer library by shearing a pool of 6 bacterial artificial chromosomes (BACs) containing genes of interest into ~500bp fragments (Fig 4.2). The BACs contain the pluripotency genes Sox2, Oct4, and Nanog that all have a high density of sites marked with H3K27ac or the histone acetyltransferase p300 at the ESC stage, both strong predictors of enhancer activity[160]. The BACs also cover HB9 and Olig2 genes. The HB9 gene is active at the neural progenitor cell (NPC) stage and promotes motoneuron cell[155, 163]. The Olig2 gene is expressed in motoneuron progenitor cells in two waves. The first wave of expression acts to produce motoneurons, the second, oligodendrocytes[164-166]. We cloned the sheared BAC fragments upstream of dERT2:CRE:dERT2 under the control of FGF4 minimal promoter in a basal lentiviral vector and produced the lentiviruses from the lentiviral plasmid library. CRE reporter ESCs were transduced with the lentiviral library and selected under BDS for 3 days to remove cells that do not have lentiviral integration, while maintaining leukemia inhibitory factor (LIF) in the medium to maintain pluripotency. To identify the active enhancers present at the ESC stage, the chemical inducer 4OHT was applied immediately after BSD selection to activate the CRE-mediated enhancer detection system and RFP positive and negative cells were sorted. To calibrate the sorting process, we used cell populations transduced with lentiviruses that have non-enhancer and EF1 promoter driving inducible CRE as negative and positive controls, respectively. Fig. 4.6 showed that cells from the negative control showed universally low reporter expression, in contrast to the positive control, in which the majority of cells showed strong RFP expression. The BAC enhancer library from randomly sheared BACs contained a small population of cells with robust reporter expression and a large population with negligible reporter expression, which is expected considering that any given genomic locus is

likely to harbor only a few enhancers active in any given cell type[137]. We collected the RFP positive cells which are expected to contain an enhancer activating the CRE-reporter system and RFP negative cells that are expected to have a transcriptionally inactive non-enhancer. We then amplified the enhancer candidate sequences in these harvested cells by PCR using universal primers that recognize the sequences flanking the enhancer site, sequenced enhancer amplicons using high-throughput sequencing technology, and mapped the reads to the BAC reference sequence. To check the coverage or distribution of the enhancer fragments, we also amplified and sequenced fragments present in the plasmid library and mapped it back to the BAC sequences. We found that the enhancer fragments from the plasmid library cover ~90% of all BAC sequences and the locus shown specifically in Fig. 4.7 showed 100% coverage of the selected region. Functionally active enhancers (Fig. 4.7 row4) were defined as those sequences that showed a significant enrichment in the fluorescent cell population (Fig. 4.7 row2) relative to the negative population (Fig. 4.7 row3). For DNA regions around pluripotent genes, a total of 48 enhancer regions were identified (Fig. 4.8 A).

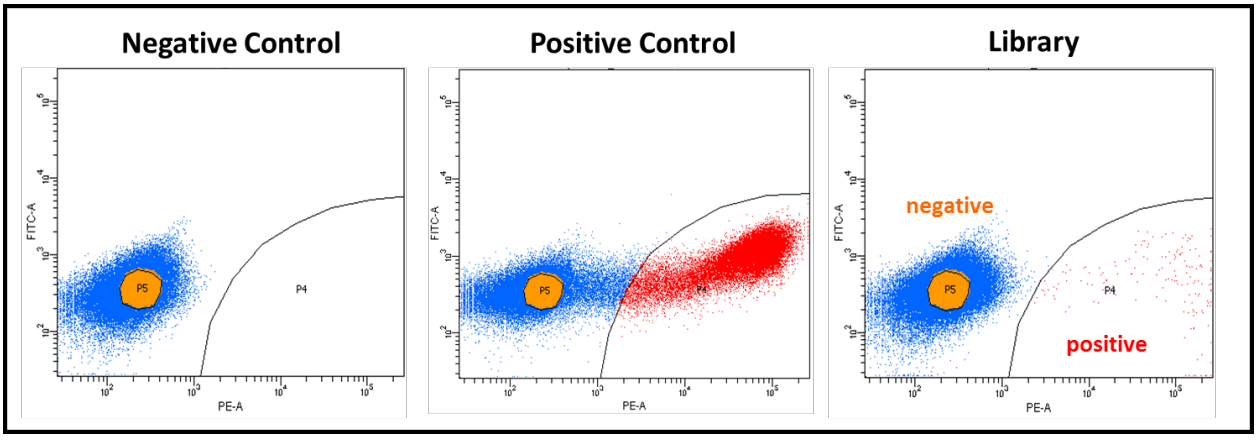


Figure 4.8 Graphs of FACS for cell soring. The plots show fluorescent intensity of RFP reporter gene for negative control (no RFP expression) and positive control (strong RFP

expression), and ES cells transduced by lentiviral library. Points represent unique cells. Red dots in P4 delineate the gate for cells considered positive for reporter gene expression.

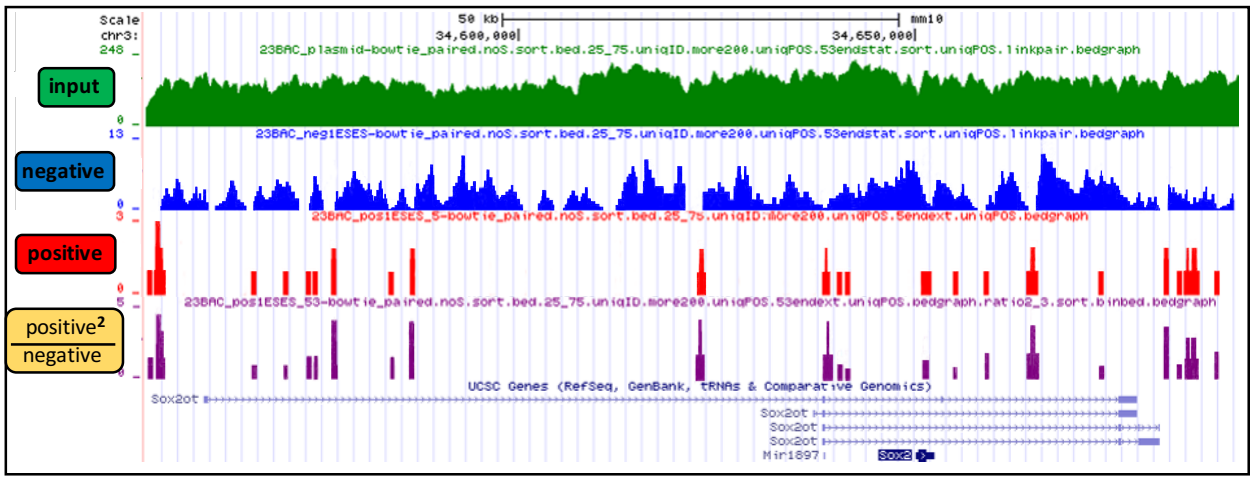


Figure 4.7 Genome browser view of enhancer peaks. Sequencing reads from input plasmid library, RFP negative cells, RFP positive cells were uniquely mapped. To identify enhancer peaks, the ratio of reads from positive over negative cells is calculated.

To confirm the accuracy of the method, we examined the enhancer activities of 20 candidate enhancer sites out of a total of 32 identified by DevEn-seq and an additional 19 sites with no enhancer activity (referred to as 'non-enhancer'). Each site was amplified from the genome and then cloned upstream of unfused CRE under control of the FGF4 minimal promoter for validation. The CRE-reporter ESCs were transfected with these constructs and the percentage of RFP positive cells was measured. We observed robust enhancer activity for 16 of the 20 (80%) DevEn-seq predicted enhancers (Fig. 4.8 B). This is in contrast to the sequences predicted to have no enhancer activity by DevEn-seq, 16 of 19 (84%) had negligible reporter expression. Collectively, the enhancers predicted by DevEn-seq drove significantly higher reporter

expression than those that were predicted by DevEn-seq to have no enhancer activity ($p<0.05$). The high validation rate in these complementary assays demonstrates the accuracy of identification of enhancers by DevEn-seq. To further estimate the positive predictive value, we overlapped the active enhancer peak regions with evidence of DNaseI hyper-sensitivity (HS) and histone modifications H3K27ac. Almost ~85% of the active enhancer peaks harbored at least one of these two enhancer markers. Taken together, these results showed that DevEn-seq is able to identify active enhancers at the ES cell stage with a high positive predictive value of ~85%.

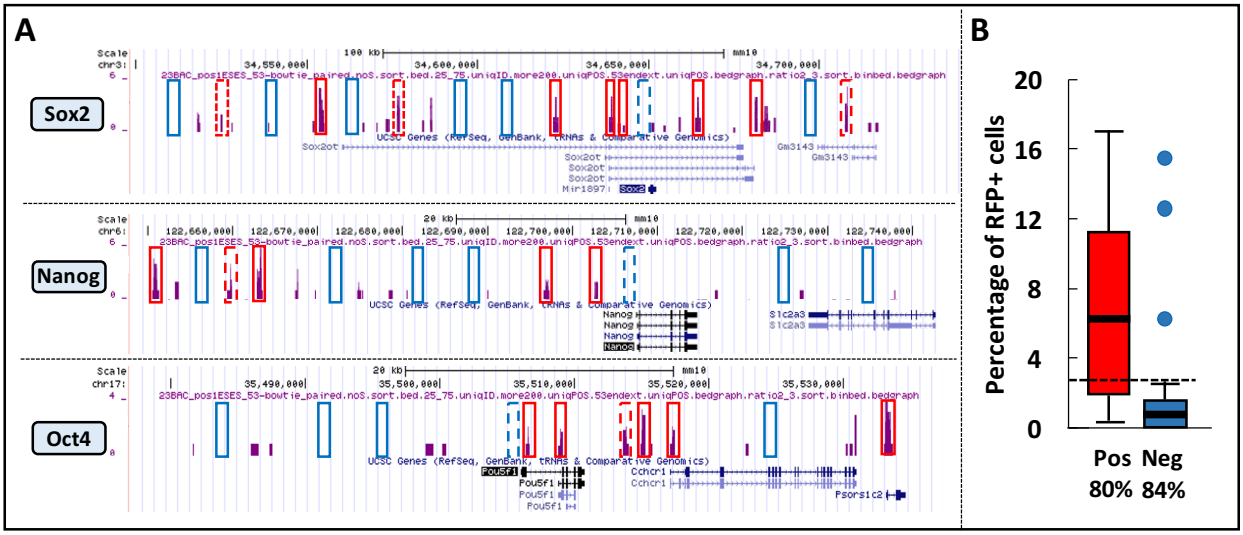


Figure 4.9 Mouse ESC enhancers identified by DevEn-Seq. (A) Mouse ESC enhancers (purple peaks) from pluripotent genes Sox2, Nanog and Oct4 loci. Enhancer peak represents (normalized RFP positive read depth)^2 /normalized RFP negative read depth. Peak signals >1 represent sequences enriched in the RFP positive population relative to the negative ones. For simplicity, only DevEn-seq signal is shown. (B) Validation of mouse ESC enhancers and non-enhancers identified by DevEn-seq. Candidate enhancer and non-enhancer regions were marked by red and blue rectangles in A. A total of 20 enhancers (red dots in B and red rectangles in A) and 19 non-enhancers (blue dots and blue rectangles in A) regions were tested by CRE-mediated reporter assay individually. The dash line indicates the cutoff for positive and negative results.

Validated and non-validated sequences were marked by solid and dash rectangles respectively in (A).

4.3.7 DevEn-Seq can identify developmental enhancers during neural differentiation

To identify enhancers that are active at neural progenitor cell (NPC) stage during neural differentiation, we used the polyclonal ESCs obtained from previous experiment and initiated the CRE-mediated enhancer assay by 4OHT when EBs were induced to generate motoneurons from Day4 to Day6 (Fig4.9 A NPC-MN). As for control and comparison, we included two additional time windows for induction and harvest: 1) induction from ESC Day -1 to ESC Day -3 and harvest at ESC Day -3 (Fig 4.9 A ES-ES) and 2) induction from ESC Day -1 to MN Day 10 and harvest at MN Day 10 (Fig 4.9 A ES-MN). For each experiments, both RFP positive and negative cells were collected by FACS for PCR amplification and sequencing. The sequencing reads were mapped back to BAC reference and peaks for each condition were called. To identify NPC specific enhancers, we compared peaks in condition of NPC-MN to the other two conditions and found that a cluster of peaks around HB9 gene were enriched for the read depths (Fig. 4.9 B, red rectangle), suggesting that this enhancer element may play a role in specifying NPC fate. Previous studies[155, 163] on HB9 gene indicate that this NPC specific enhancer region matches the highly conserved enhancers of HB9 and is used for purifying NPC-derived motoneurons from a mixture of other neural cell types including oligodendrocytes and astrocytes. In addition, we found two peaks enriched around Olig2 gene (Fig. 4.9 C, red rectangle), which are consistent with what previously found to be NPC-specific regulatory sequences[167, 168]. These results suggest that DevEn-seq can identify cell type specific enhancers during neural differentiation throughout development.

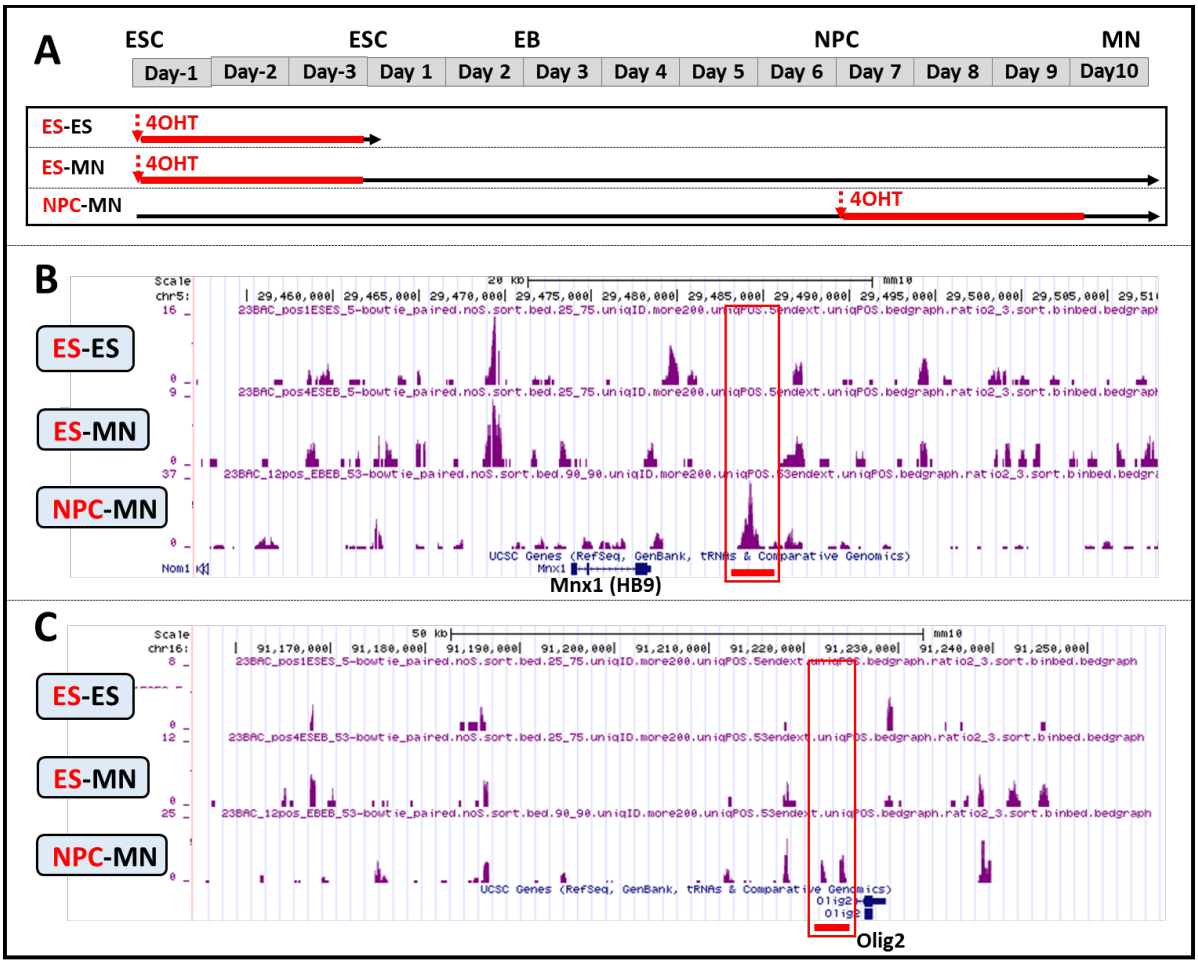


Figure 4.10 NPC-specific enhancers identified by DevEn-seq. (A) Timeline of neural differentiation. Red line indicates the induction time window and the black arrow denotes the starting and ending points for each experiment. (B and C) Genome browser view of the enhancer peaks from three conditions. The peaks in the red rectangle at condition of NPC-MN indicates the NPC specific enhancers around HB9 gene (B) and Olig2 gene (C).

4.4 Discussion

Regular methods for functional identification of enhancers place enhancer fragments upstream of a minimal promoter and a reporter to test its transcriptional activity. One major

concern is that the integrated lentiviral sequences become methylated and repressed during differentiation, regardless of the sequences driving the reporter, causing the initial positive signals to disappear from the differentiated cells [152]. DevEn-seq positions the enhancer fragments upstream of a CRE gene that initiates the cascade reactions to activate the 'real' reporter. By adopting CRE/loxP system and a CRE-reporter cell line, the reporter gene is not in the lentiviral transgene but was transferred to the Rosa26 locus, a ubiquitously expressed locus that survives the methylation-driven gene silencing in development [148].

DevEn-seq also has a higher sensitivity than regular methods that use fluorescent protein as reporter because the amount of CRE recombinase required to activate the CRE/loxP system is lower than that of fluorescent reporter protein to be detected by FACS. In addition, once the CRE/loxP system is activated, the reporter of DevEn-seq is under a strong CAG promoter at Rosa26 locus, which can provide a strong and ubiquitous expression of the reporter gene.

By functionally interrogating >500 kilobases (kb) of mouse sequence in mouse embryonic stem cells for enhancer activity, we identified 38 enhancers at pluripotency loci with a positive predictive value of 80%. With an *in vitro* neural differentiation protocol, we identified two neural progenitor-specific enhancers around the HB9 and Olig2 genes. This demonstrated that DevEn-seq can be used to trace enhancer activity without being disturbed by the gene silencing effect caused by lentiviral sequences and is limited only by currently available stem cell differentiation methods.

The need for high-throughput assays to directly interrogate enhancer activity has led to the recent development of multiplex methods to functionally assess genetic regulatory elements [142, 143, 146, 147, 169-171]. However, most of these methods rely on transient delivery of enhancer-reporter plasmids, limiting the use of these methods to a small number of easily transfected cell types.

Furthermore, many enhancers have been shown to have negligible or different activity when tested in transient assays but robust activity when integrated into the genome [172-174]. This suggests that transient delivery of enhancer-reporter constructs may not recapitulate the native chromatin environment found in chromosomes, which may be necessary for proper gene regulation. DevEn-seq improves substantially on these previous methods by assessing mammalian enhancer activity in a genomic context and in a potentially much wider variety of cell types.

The number of putative enhancers assessed in a single DevEn-seq experiment is currently limited by the transduction efficiency in mouse ES cells. For the experiments described, lentiviral genome integration occurred in approximately one in 10^5 mouse ES cells at an MOI of 0.3. Due to the enormous size of mammalian genomes, performing unbiased enhancer discovery across an entire genome may therefore prove too experimentally arduous for DevEn-seq. New technologies to improve stem cell permissiveness to HIV-derived lentiviral vectors and relieve host lentiviral restriction blocks may improve this integration efficiency and thereby increase the throughput of this approach[175]. Alternatively, limiting the library or search space to the accessible cellular chromatin that are most relevant for endogenous transcriptional regulation could potentially allow for a more efficient enhancer screening by DevEn-seq. It was previously shown that the efficiency of identifying biologically relevant transcriptional regulatory elements can be dramatically increased by focusing on DNA isolated from nucleosome-free regions (NFRs) [176, 177] (i.e., genomic regions in which nucleosomes are relatively depleted and/or highly destabilized). Notably, NFRs represent only 2% of chromatin. Thus, by focusing functional analysis on NFR-derived DNA reduces the search space to the most relevant portions of the genome and eliminates the need for a whole genome screen.

As mounting evidence for enhancer's contributions to development and human disease, the identification of enhancer elements in different cell types and under different biological conditions is currently of high priority in biomedical research. DevEn-seq will help to overcome the limitations currently curbing the ability to functionally identify or validate large numbers of putative enhancers directly in many disease-relevant cell types throughout development. For example, DevEn-seq has the potential to decrease the need for transgenic mice to test enhancers active in specific cell types. Moreover, future development of DevEn-seq could allow for the more comprehensive study of the roles of enhancers in human disease and ultimately guide the reprogramming of embryonic or induced pluripotent stem cells to produce specific cell types for personalized transplants, such as pancreatic beta cells to treat diabetes.

4.5 Method

4.5.1 Constructing plasmid libraries

BACs that contain the genes of interests were ordered from the BACPAC Resource Center at Children's Hospital Oakland Research Institute. BAC DNA was isolated by MACHEREY-NAGEL NucleoBond® Xtra BAC kit and sheared with a Covaris S220/E220 Focused-ultrasonicator using parameters of generating ~1kb long DNA fragments. The fragmented DNA were further size-selected using 1% agarose gel. Illumina Multiplexing Adapters were ligated to 1µg size-selected DNA fragments using Accel-NGS® 2S Plus DNA Library Kit (Swift Bioscience) following manufacturer's instructions. The screening vector was linearized by a 16-hour digestion with MreI (Thermo Fisher Scientific) and AscI (NEB), followed by agarose gel electrophoresis and QIAquick gel extraction.

The adapter-ligated fragments were recombined to the linearized screening vector by In-Fusion HD kit (Clontech) in a total of 2 10ul reactions. The In-Fusion HD reactions were pooled, ethanol precipitated and eluted in 20µl EB [10mM Tris-HCL, pH 8]. Two aliquots (40 µl each) of MegaX DH10B Electrocompetent Bacteria (Invitrogen) were transformed with 10µl eluted DNA each, according to the manufacturer's protocol. After one hour recovery at 37 °C, two transformation reactions were pooled and grown on six LB-AMP 245mm x 245mm plates (Corning). The plasmid libraries were extracted using Plasmid Maxi kit (Qiagen).

4.5.2 Preparation of lentiviral libraries

The lentiviral libraries were prepared by the Hope Center viral core at Washington University School of Medicine as previously described [Production of Lentiviral Vectors for Transducing Cells from the Central Nervous System]. Lentiviral titres will be determined in triplicate by transducing 1×10^5 HT1080 cells with serial dilutions of concentrated viruses (1:100 to 1:10000). After 72h incubation, the genomic DNAs of each transduced samples will be extracted. The six point standard curves from 10^6 to 10^2 copies will be generated by serial dilution of lentiviral gene and human gene (Albumin) fragments respectively. The number of lentiviral gene copies will be determined by quantitative PCR (qPCR) with primers recognizing the lentiviral LTRs while number of mouse genomes will be determined using primers recognizing a unique human gene (Albumin). Titres will be determined according to the following formula (Fig 4.11) and expressed as transducing units/mL (TU/mL). The titres will be used to balance the differences of infectious efficiency for each batch of viruses, ensuring the consistency of the experiment.

$$\text{Titer (transducing units/mL):} \quad \frac{\text{number of hygromycin genes}}{\text{number of genomes}} \times \frac{\text{number of cells on day of transduction}}{\text{total volume of viral inoculum}} \times \text{Dilution factor}$$

Figure 4.11 The formula for calculating the lentiviral titre.

4.5.3 Transduction of lentiviral libraries and drug administration

Mouse ESCs were transduced at a MOI of 0.2 to ensure that a majority of cells (~90% according to Poisson distribution) was transduced with a single-copy viral integration. To increase the likelihood that any given BAC-lentiviral fragments would be represented, the number of cells transduced was equivalent to more than 30 times each library's complexity. Considering the MOI of 0.2, xxx × 10^6 ESCs were plated in complete ESC medium plus LIF in feeder-free conditions on a 10-cm gelatin-coated dish 1 d before transduction. The cells were then transduced overnight in 10 ml of complete ESC medium plus 10 μg/ml Polybrene. The following day, the medium was replaced with fresh ESC medium plus LIF. Blasticidin selection was initiated 4 d post transduction in ESC medium/LIF containing 5 μg/ml Blasticidin S. Cells were selected for Blasticidin resistance for 5 d, with medium changed daily. To activate the cre-mediated enhancer identification, cells were treated with 1 μM 4-hydroxy tamoxifen (4OHT, Sigma) and maintained for at least one day for full activation.

4.5.4 Neural differentiation

Mouse ESCs were differentiated into neurons, oligodendrocytes and astrocytes using a retinoic acid (RA) and smoothened agonist (SAG) induction protocol as described [93]. Briefly, ESCs were cultured in suspension on low attachment plates (Corning) in modified DFK5 media consisting of DMEM/F12 base media (Gibco) containing 5% knockout serum replacement, 1 ×

insulin transferrin selenium, 50 μM of nonessential amino acids, 100 mM of BME, 5 mM of thymidine, and 15 mM of the following nucleosides: adenosine, cytosine, guanosine, and uridine. During this process, ESCs aggregate into multi-cellular embryoid bodies (EBs). After the first 2 days, the EBs were moved to a 15 mL of conical and allowed to settle for 5 min. The media was aspirated and replaced with 10mL of fresh DFK5 containing 2 mM of RA and 600nM of SAG. EBs were then cultured on the adhesive plates (Corning) for an additional 4 days for further differentiation, and media was replaced every 2 days.

4.5.5 Fluorescence-activated cell sorting (FACS)

Before sorting, cells were washed with PBS and collected using trypsin. Cells were pelleted by centrifugation, the trypsin was removed and the cells were washed with PBS. Cells were resuspended in 1% w/v saline by repeated pipetting and passed through a 0.4-μm strainer to ensure single-cell suspension. Cells were sorted on an iCyt Reflection HAPS2 cell sorter at the Washington University Siteman Flow Cytometry Core. The gate was set relative to the cells transfected with non-fluorescent control plasmids to eliminate background. Cells transfected with a positive control fluorescent reporter plasmid were also used to eliminate false positive singles. Flow cytometry metrics were analyzed using FlowJo Version 7.6.3 (TreeStar).

4.5.6 PCR amplification of inserts and sequencing

Genomic DNA was isolated from both RFP positive and negative populations of cells using the QIAamp DNA Mini kit (Qiagen). The enhancer position sites were amplified from the genomic DNA by primers that contains both Illumina sequencing adaptors and indexes, permitting one-step amplification and sequencing library preparation. For each populations, two

PCR reactions were performed with 100 ng gDNA and 23 cycles of amplification(10 µl Phusion polymerase buffer, 1 µl 10 mM dNTP, 2.5 µl 10 µM forward and reverse primer, 1.5 µl DMSO, 0.5 µl (NEB) 50 ng DNA, 31 µl H2O; 16 cycles with 55 °C annealing temperature). The 2 reactions were then pooled and purified, and sequenced on Illumina MiSeq platform at the Genome Technology Access Center (GTAC) at Washington University in St. Louis.

4.5.7 Bioinformatic data analysis

Paired-end sequences were aligned to the mouse reference genome (mm9) using bowtie2. Read pairs were filtered from the final data set if either read failed to map to the genome, if both reads did not map in the proper orientation, if the mapping quality score of both reads was less than 25, or if neither read had a unique map location on the genome. Target sites were identified as loci where paired reads both aligned entirely within a 1200-bp genomic region. To pinpoint the active enhancer regions enriched for reads from positive libraries, the fold enrichment was calculated as the normalized reads density from the RFP positive cells divided by that from RFP positive negative cells.

Chapter 5
Conclusion

For multicellular organisms to develop, stem cells must differentiate into a broad range of specialized cells, all containing the same DNA. This remarkable feat is made possible by the interaction of TFs and cis regulatory elements that control the spatial and temporal expression patterns of specific sets of genes[134]. Although progress has been made toward understanding of how transcription are regulated to generate distinct cell types, it could be greatly accelerated with better molecular tools.

Chapter 2 describes a degradation domain based induction system for "Calling Cards" method which maps the binding sites of TFs using *piggybac* transposons. This induction system satisfies five important criteria: 1) The system has a low basal transposition activity when 'off ' and maintains high transposition activity when 'on'; 2) The PBase fusion protein shows high transposition activity almost equal to that of the unfused "wild type" PBase; 3) The system can be applied across different cell lines with high performances; 4) The induction is reversible and responds in a dose-dependent manner; 5) The chemical inducer does not interfere with general cellular function. This induction system is successfully applied to PB transposon-mediated "Calling Cards" method, which offers an alternative to chromatin immunoprecipitation (ChIP) methods and furthermore the ability to record TF binding through cell division and at different time periods of the development.

Chapter 3 describes the application of the "Calling Cards" method to study the role of master regulatory Brd4-bound enhancers for sex differences in glioblastoma. Integration of Brd4 Calling Cards, H3K27ac ChIP-seq and RNA-seq data revealed a set of sex-specific regulatory

genes and networks. Male-specific JQ1-sensitive targetable genes demonstrated functional enrichment for glioblastoma, tumor angiogenesis, integrin signaling pathway, metabolic process in addition to DNA-repair-deficiency disorders and cell proliferation. Female-specific JQ1-sensitive targetable genes showed an enrichment in pathways involved semaphorin signaling, chromosome aberrations, positive regulation of transcription, cell differentiation and tumor progression. These results are indicative of sex-specific transcriptional programs regulated by Brd4-bound enhancers. Identifying which specific pathways are critical to sex differences in GBM will require further functional studies.

To record the activity of regulatory elements or enhancers, in Chapter 4, we developed a high-throughput method for functional identification of active enhancers at different time periods of development, named as Developmental Enhancer Sequencing (DevEn-seq). We demonstrated that DevEn-seq is able to: 1) identify active enhancers more efficiently than regular reporter methods; 2) trace enhancer activities along a cell lineage at a high temporal resolution without being disturbed by the gene silencing effect caused by lentiviral sequences; 3) discover potential makers that can be used to purify the progenitors of a given cell type. The knowledge obtained through the application of this method would greatly expand our understanding of neural cell fate specification and transcriptional network dynamics in development.

In summary, this dissertation contributes to the field of developmental biology by providing useful methods for recording transcription factor binding and enhancer activity during development. Because of the mounting evidence for transcriptional network's contributions to development and human disease, the identification of its two key elements TF and enhancers in different cell types and under different biological conditions is currently of high priority in biomedical research. The "Calling Cards" method and DevEn-seq will help to overcome the

limitations currently to trace curbing the ability the activity of TFs and enhancers in many disease-relevant cell types throughout development. These methods could allow for the more comprehensive study of the roles of TFs and enhancers in human disease and ultimately guide the reprogramming of embryonic or induced pluripotent stem cells to produce specific cell types for personalized transplants, such as pancreatic beta cells to treat diabetes.

References

1. Hardingham, G.E., Pruunsild, P., Greenberg, M.E. & Bading, H. Lineage divergence of activity-driven transcription and evolution of cognitive ability. *Nat Rev Neurosci* **19**, 9-15 (2018).
2. Wang, H., Mayhew, D., Chen, X., Johnston, M. & Mitra, R.D. "Calling cards" for DNA-binding proteins in mammalian cells. *Genetics* **190**, 941-949 (2012).
3. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. & Luscombe, N.M. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**, 252-263 (2009).
4. Nebert, D.W. Transcription factors and cancer: an overview. *Toxicology* **181-182**, 131-141 (2002).
5. Vierbuchen, T. et al. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035-1041 (2010).
6. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).
7. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods* **6**, 917-922 (2009).
8. Hermann, A., Liewald, J.F. & Gottschalk, A. A photosensitive degron enables acute light-induced protein degradation in the nervous system. *Curr Biol* **25**, R749-750 (2015).
9. Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G. & Wandless, T.J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995-1004 (2006).
10. Iwamoto, M., Bjorklund, T., Lundberg, C., Kirik, D. & Wandless, T.J. A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem Biol* **17**, 981-988 (2010).
11. Egeler, E.L., Urner, L.M., Rakhit, R., Liu, C.W. & Wandless, T.J. Ligand-switchable substrates for a ubiquitin-proteasome system. *J Biol Chem* **286**, 31328-31336 (2011).
12. Banaszynski, L.A., Sellmyer, M.A., Contag, C.H., Wandless, T.J. & Thorne, S.H. Chemical control of protein stability and function in living mice. *Nat Med* **14**, 1123-1127 (2008).
13. Armstrong, C.M. & Goldberg, D.E. An FKBP destabilization domain modulates protein levels in Plasmodium falciparum. *Nat Methods* **4**, 1007-1009 (2007).
14. Muralidharan, V., Oksman, A., Iwamoto, M., Wandless, T.J. & Goldberg, D.E. Asparagine repeat function in a Plasmodium falciparum protein assessed via a regulatable fluorescent affinity tag. *Proc Natl Acad Sci U S A* **108**, 4411-4416 (2011).
15. Glass, M., Busche, A., Wagner, K., Messerle, M. & Borst, E.M. Conditional and reversible disruption of essential herpesvirus proteins. *Nat Methods* **6**, 577-579 (2009).
16. Cadinanos, J. & Bradley, A. Generation of an inducible and optimized piggyBac transposon system. *Nucleic acids research* **35**, e87 (2007).
17. Kool, M. et al. Molecular subgroups of medulloblastoma: an international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, Group 3, and Group 4 medulloblastomas. *Acta Neuropathol* **123**, 473-484 (2012).

18. Taylor, M.D. et al. Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol* **123**, 465-472 (2012).
19. Sun, T. et al. Sexually dimorphic RB inactivation underlies mesenchymal glioblastoma prevalence in males. *J Clin Invest* **124**, 4123-4133 (2014).
20. Suva, M.L. et al. Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* **157**, 580-594 (2014).
21. Gangemi, R.M. et al. SOX2 silencing in glioblastoma tumor-initiating cells causes stop of proliferation and loss of tumorigenicity. *Stem Cells* **27**, 40-48 (2009).
22. Forger, N.G. et al. Deletion of Bax eliminates sex differences in the mouse forebrain. *Proc Natl Acad Sci U S A* **101**, 13666-13671 (2004).
23. Murray, E.K., Hien, A., de Vries, G.J. & Forger, N.G. Epigenetic control of sexual differentiation of the bed nucleus of the stria terminalis. *Endocrinology* **150**, 4241-4247 (2009).
24. Iyer, N.G., Ozdag, H. & Caldas, C. p300/CBP and cancer. *Oncogene* **23**, 4225-4231 (2004).
25. Lagger, G. et al. Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression. *EMBO J* **21**, 2672-2681 (2002).
26. Rahman, S. et al. The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Mol Cell Biol* **31**, 2641-2652 (2011).
27. Jung, M. et al. Affinity map of bromodomain protein 4 (BRD4) interactions with the histone H4 tail and the small molecule inhibitor JQ1. *J Biol Chem* **289**, 9304-9319 (2014).
28. Mochizuki, K. et al. The bromodomain protein Brd4 stimulates G1 gene transcription and promotes progression to S phase. *J Biol Chem* **283**, 9040-9048 (2008).
29. Dey, A., Nishiyama, A., Karpova, T., McNally, J. & Ozato, K. Brd4 marks select genes on mitotic chromatin and directs postmitotic transcription. *Mol Biol Cell* **20**, 4899-4909 (2009).
30. Devaiah, B.N. et al. BRD4 is a histone acetyltransferase that evicts nucleosomes from chromatin. *Nat Struct Mol Biol* **23**, 540-548 (2016).
31. Wu, T. & Donohoe, M.E. The converging roles of BRD4 and gene transcription in pluripotency and oncogenesis. *RNA Dis* **2** (2015).
32. Loven, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-334 (2013).
33. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299-308 (1981).
34. Ren, B. & Yue, F. Transcriptional Enhancers: Bridging the Genome and Phenome. *Cold Spring Harb Symp Quant Biol* **80**, 17-26 (2015).
35. Heintzman, N.D. et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112 (2009).
36. Shen, Y. et al. A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116-120 (2012).
37. Ong, C.T. & Corces, V.G. Enhancers: emerging roles in cell fate specification. *EMBO Rep* **13**, 423-430 (2012).

38. Ding, N., Qu, H. & Fang, X. [The ENCODE project and functional genomics studies]. *Yi Chuan* **36**, 237-247 (2014).
39. Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
40. Kleinjan, D.A. & Lettice, L.A. Long-range gene control and genetic disease. *Adv Genet* **61**, 339-388 (2008).
41. Zhang, X., Cowper-Sal lari, R., Bailey, S.D., Moore, J.H. & Lupien, M. Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. *Genome Res* **22**, 1437-1446 (2012).
42. Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. & Shiroishi, T. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**, 797-803 (2005).
43. Yanagisawa, H., Clouthier, D.E., Richardson, J.A., Charite, J. & Olson, E.N. Targeted deletion of a branchial arch-specific enhancer reveals a role of dHAND in craniofacial development. *Development* **130**, 1069-1078 (2003).
44. Shim, S., Kwan, K.Y., Li, M., Lefebvre, V. & Sestan, N. Cis-regulatory control of corticospinal system development and evolution. *Nature* **486**, 74-79 (2012).
45. Attanasio, C. et al. Fine tuning of craniofacial morphology by distant-acting enhancers. *Science* **342**, 1241006 (2013).
46. Maurano, M.T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-1195 (2012).
47. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947 (2013).
48. Martin, D.I., Fiering, S. & Groudine, M. Regulation of beta-globin gene expression: straightening out the locus. *Curr Opin Genet Dev* **6**, 488-495 (1996).
49. Tuan, D. & London, I.M. Mapping of DNase I-hypersensitive sites in the upstream DNA of human embryonic epsilon-globin gene in K562 leukemia cells. *Proc Natl Acad Sci U S A* **81**, 2718-2722 (1984).
50. Sabo, P.J. et al. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods* **3**, 511-518 (2006).
51. Vierstra, J. et al. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* **346**, 1007-1012 (2014).
52. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).
53. Buenrostro, J.D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-490 (2015).
54. Cusanovich, D.A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910-914 (2015).
55. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
56. Mikkelsen, T.S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560 (2007).
57. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837 (2007).

58. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283 (2011).
59. Creyghton, M.P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936 (2010).
60. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L.A. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res* **35**, D88-92 (2007).
61. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106-1117 (2008).
62. Boyer, L.A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-956 (2005).
63. Heintzman, N.D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-318 (2007).
64. Hoffman, M.M. et al. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods* **9**, 473-476 (2012).
65. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-216 (2012).
66. Carlson, C.M. & Largaespada, D.A. Insertional mutagenesis in mice: new perspectives and tools. *Nature reviews. Genetics* **6**, 568-580 (2005).
67. Ivics, Z., Hackett, P.B., Plasterk, R.H. & Izsvak, Z. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**, 501-510 (1997).
68. Moran, J.V. et al. High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917-927 (1996).
69. Fraser, M.J., Brusca, J.S., Smith, G.E. & Summers, M.D. Transposon-mediated mutagenesis of a baculovirus. *Virology* **145**, 356-361 (1985).
70. Wu, S.C. et al. piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 15008-15013 (2006).
71. Fraser, M.J., Ciszczon, T., Elick, T. & Bauser, C. Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect molecular biology* **5**, 141-151 (1996).
72. Wilson, M.H., Coates, C.J. & George, A.L., Jr. PiggyBac transposon-mediated gene transfer in human cells. *Molecular therapy : the journal of the American Society of Gene Therapy* **15**, 139-145 (2007).
73. Ding, S. et al. Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **122**, 473-483 (2005).
74. Wu, S., Ying, G., Wu, Q. & Capecchi, M.R. Toward simpler and faster genome-wide mutagenesis in mice. *Nature genetics* **39**, 922-930 (2007).
75. Woltjen, K. et al. piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458**, 766-770 (2009).
76. Yusa, K., Rad, R., Takeda, J. & Bradley, A. Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *Nature methods* **6**, 363-369 (2009).

77. Rad, R. et al. PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice. *Science* **330**, 1104-1107 (2010).
78. Wang, H., Mayhew, D., Chen, X., Johnston, M. & Mitra, R.D. Calling Cards enable multiplexed identification of the genomic targets of DNA-binding proteins. *Genome Res* **21**, 748-755 (2011).
79. Wang, H., Johnston, M. & Mitra, R.D. Calling cards for DNA-binding proteins. *Genome Res* **17**, 1202-1209 (2007).
80. Pick, E. et al. High HSP90 expression is associated with decreased survival in breast cancer. *Cancer research* **67**, 2932-2937 (2007).
81. Zagouri, F. et al. Decreased Hsp90 expression in infiltrative lobular carcinoma: an immunohistochemical study. *BMC cancer* **10**, 409 (2010).
82. Zhao, H., Yang, H., Zhao, H., Chen, M. & Wang, T. The molecular characterization and expression of heat shock protein 90 (Hsp90) and 26 (Hsp26) cDNAs in sea cucumber (Apostichopus japonicus). *Cell stress & chaperones* **16**, 481-493 (2011).
83. Deb, R. et al. Effect of heat stress on the expression profile of Hsp90 among Sahiwal (Bos indicus) and Frieswal (Bos indicus x Bos taurus) breed of cattle: a comparative study. *Gene* **536**, 435-440 (2014).
84. Nobakht, M., Najafzadeh, N. & Kordestani Shargh, B. Effects of tamoxifen on morphological and ultrastructural aspects of developing hippocampus of rat. *Iranian biomedical journal* **13**, 237-243 (2009).
85. Mehasseb, M.K., Bell, S.C. & Habiba, M.A. The effects of tamoxifen and estradiol on myometrial differentiation and organization during early uterine development in the CD1 mouse. *Reproduction* **138**, 341-350 (2009).
86. Singh, R., Singh, A.K. & Tripathi, M. Effect of a non steroidal tamoxifen on the gonad and sex differentiation in Nile tilapia, Oreochromis niloticus. *Journal of environmental biology / Academy of Environmental Biology, India* **33**, 799-803 (2012).
87. Sellmyer, M.A., Chen, L.C., Egeler, E.L., Rakhit, R. & Wandless, T.J. Intracellular context affects levels of a chemically dependent destabilizing domain. *PloS one* **7**, e43297 (2012).
88. Wilson, M.H. & George, A.L., Jr. Designing and testing chimeric zinc finger transposases. *Methods in molecular biology* **649**, 353-363 (2010).
89. Wang, W., Bradley, A. & Huang, Y. A piggyBac transposon-based genome-wide library of insertionally mutated Blm-deficient murine ES cells. *Genome research* **19**, 667-673 (2009).
90. Mansergh, F.C. et al. Gene expression profiles during early differentiation of mouse embryonic stem cells. *BMC Dev Biol* **9**, 5 (2009).
91. Bradley, E., Bieberich, E., Mivechi, N.F., Tangpisuthipongsa, D. & Wang, G. Regulation of embryonic stem cell pluripotency by heat shock protein 90. *Stem Cells* **30**, 1624-1633 (2012).
92. Mullick, A. et al. The cumate gene-switch: a system for regulated expression in mammalian cells. *BMC Biotechnol* **6**, 43 (2006).
93. Wichterle, H. & Peljto, M. Differentiation of mouse embryonic stem cells to spinal motor neurons. *Curr Protoc Stem Cell Biol* **Chapter 1**, Unit 1H 1 1-1H 1 9 (2008).

94. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57 (2009).
95. Huang da, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**, 1-13 (2009).
96. Keith, J.H., Fraser, T.S. & Fraser, M.J., Jr. Analysis of the piggyBac transposase reveals a functional nuclear targeting signal in the 94 c-terminal residues. *BMC Mol Biol* **9**, 72 (2008).
97. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
98. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
99. Anders, S., Pyl, P.T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
100. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).
101. Ober, C., Loisel, D.A. & Gilad, Y. Sex-specific genetic architecture of human disease. *Nat Rev Genet* **9**, 911-922 (2008).
102. Ostrom, Q.T. et al. The epidemiology of glioma in adults: a "state of the science" review. *Neuro Oncol* **16**, 896-913 (2014).
103. Ang, C., Guiot, M.C., Ramanakumar, A.V., Roberge, D. & Kavan, P. Clinical significance of molecular biomarkers in glioblastoma. *Can J Neurol Sci* **37**, 625-630 (2010).
104. Ostrom, Q.T. et al. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2009-2013. *Neuro Oncol* **18**, v1-v75 (2016).
105. Siegel, R. et al. Cancer treatment and survivorship statistics, 2012. *CA Cancer J Clin* **62**, 220-241 (2012).
106. Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics for Hispanics/Latinos, 2012. *CA Cancer J Clin* **62**, 283-298 (2012).
107. Sun, T., Warrington, N.M. & Rubin, J.B. Why does Jack, and not Jill, break his crown? Sex disparity in brain tumors. *Biol Sex Differ* **3**, 3 (2012).
108. Ostrom, Q.T. et al. CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014. *Neuro Oncol* **19**, v1-v88 (2017).
109. Dorak, M.T. & Karpuzoglu, E. Gender differences in cancer susceptibility: an inadequately addressed issue. *Front Genet* **3**, 268 (2012).
110. Kfoury, N. et al. Cooperative p16 and p21 action protects female astrocytes from transformation. *Acta Neuropathol Commun* **6**, 12 (2018).
111. McCarthy, M.M. & Nugent, B.M. At the frontier of epigenetics of brain sex differences. *Front Behav Neurosci* **9**, 221 (2015).
112. Filippakopoulos, P. et al. Selective inhibition of BET bromodomains. *Nature* **468**, 1067-1073 (2010).

113. Crawford, N.P. et al. Bromodomain 4 activation predicts breast cancer survival. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 6380-6385 (2008).
114. Urbanucci, A. et al. Androgen Receptor Deregulation Drives Bromodomain-Mediated Chromatin Alterations in Prostate Cancer. *Cell reports* **19**, 2045-2059 (2017).
115. Janouskova, H. et al. Opposing effects of cancer-type-specific SPOP mutants on BET protein degradation and sensitivity to BET inhibitors. *Nat Med* **23**, 1046-1054 (2017).
116. Ippolito, J.E., Yim, A.K., Luo, J., Chinnaiyan, P. & Rubin, J.B. Sexual dimorphism in glioma glycolysis underlies sex differences in survival. *JCI Insight* **2** (2017).
117. Dowen, J.M. et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374-387 (2014).
118. Hnisz, D. et al. Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol Cell* **58**, 362-370 (2015).
119. Ounzain, S. & Pedrazzini, T. Super-enhancer lncs to cardiovascular development and disease. *Biochim Biophys Acta* **1863**, 1953-1960 (2016).
120. Parker, S.C. et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 17921-17926 (2013).
121. Whyte, W.A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319 (2013).
122. Filippakopoulos, P. & Knapp, S. Targeting bromodomains: epigenetic readers of lysine acetylation. *Nature reviews. Drug discovery* **13**, 337-356 (2014).
123. Alsarraj, J. et al. Deletion of the proline-rich region of the murine metastasis susceptibility gene Brd4 promotes epithelial-to-mesenchymal transition- and stem cell-like conversion. *Cancer research* **71**, 3121-3131 (2011).
124. Wu, T., Pinto, H.B., Kamikawa, Y.F. & Donohoe, M.E. The BET family member BRD4 interacts with OCT4 and regulates pluripotency gene expression. *Stem cell reports* **4**, 390-403 (2015).
125. Delmore, J.E. et al. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904-917 (2011).
126. Ali, I., Choi, G. & Lee, K. BET Inhibitors as Anticancer Agents: A Patent Review. *Recent Pat Anticancer Drug Discov* (2017).
127. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
128. Najafova, Z. et al. BRD4 localization to lineage-specific enhancers is associated with a distinct transcription factor repertoire. *Nucleic acids research* **45**, 127-141 (2017).
129. Bhagwat, A.S. et al. BET Bromodomain Inhibition Releases the Mediator Complex from Select cis-Regulatory Elements. *Cell reports* **15**, 519-530 (2016).
130. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
131. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
132. Ross-Innes, C.S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-393 (2012).
133. Stark, R.a.B., G. DiffBind: differential binding analysis of ChIP-Seq peak data. (2011).

134. Davidson, E.H. & Erwin, D.H. Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796-800 (2006).
135. Kvon, E.Z. et al. Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**, 91-95 (2014).
136. Yanez-Cuna, J.O., Kvon, E.Z. & Stark, A. Deciphering the transcriptional cis-regulatory code. *Trends in genetics : TIG* **29**, 11-22 (2013).
137. Maston, G.A., Landt, S.G., Snyder, M. & Green, M.R. Characterization of enhancer function from genome-wide analyses. *Annu Rev Genomics Hum Genet* **13**, 29-57 (2012).
138. Hardison, R.C. & Taylor, J. Genomic approaches towards finding cis-regulatory modules in animals. *Nature reviews. Genetics* **13**, 469-483 (2012).
139. Boyle, A.P. et al. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311-322 (2008).
140. Baker, M. Making sense of chromatin states. *Nature methods* **8**, 717-722 (2011).
141. Johnson, D.S., Mortazavi, A., Myers, R.M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497-1502 (2007).
142. Melnikov, A. et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* **30**, 271-277 (2012).
143. Patwardhan, R.P. et al. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat Biotechnol* **30**, 265-270 (2012).
144. Landolin, J.M. et al. Sequence features that drive human promoter function and tissue specificity. *Genome research* **20**, 890-898 (2010).
145. Mogno, I., Kwasnieski, J.C. & Cohen, B.A. Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. *Genome research* **23**, 1908-1915 (2013).
146. Arnold, C.D. et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074-1077 (2013).
147. Gisselbrecht, S.S. et al. Highly parallel assays of tissue-specific enhancers in whole Drosophila embryos. *Nat Methods* **10**, 774-780 (2013).
148. Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature neuroscience* **13**, 133-140 (2010).
149. Kothary, R. et al. Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development* **105**, 707-714 (1989).
150. deBoer, E., Antoniou, M., Mignotte, V., Wall, L. & Grosveld, F. The human beta-globin promoter; nuclear protein factors and erythroid specific induction of transcription. *The EMBO journal* **7**, 4203-4212 (1988).
151. Lucibello, F.C., Ehlert, F. & Muller, R. Multiple interdependent regulatory sites in the mouse c-fos promoter determine basal level transcription: cell type-specific effects. *Nucleic acids research* **19**, 3583-3591 (1991).
152. Murtha, M. et al. FIREWACh: high-throughput functional detection of transcriptional regulatory modules in mammalian cells. *Nature methods* **11**, 559-565 (2014).
153. Casanova, E. et al. ER-based double iCre fusion protein allows partial recombination in forebrain. *Genesis* **34**, 208-214 (2002).

154. Miyazaki, Y., Imoto, H., Chen, L.C. & Wandless, T.J. Destabilizing domains derived from the human estrogen receptor. *J Am Chem Soc* **134**, 3942-3945 (2012).
155. McCreedy, D.A. et al. A new method for generating high purity motoneurons from mouse embryonic stem cells. *Biotechnol Bioeng* **111**, 2041-2055 (2014).
156. Iyer, N.R., Huettner, J.E., Butts, J.C., Brown, C.R. & Sakiyama-Elbert, S.E. Generation of highly enriched V2a interneurons from mouse embryonic stem cells. *Exp Neurol* **277**, 305-316 (2016).
157. Liu, W. et al. BRD4 regulates Nanog expression in mouse embryonic stem cells and preimplantation embryos. *Cell Death Differ* **21**, 1950-1960 (2014).
158. Gurok, U. et al. Gene expression changes in the course of neural progenitor cell differentiation. *J Neurosci* **24**, 5982-6002 (2004).
159. Shimojo, D. et al. Rapid, efficient, and simple motor neuron differentiation from human pluripotent stem cells. *Mol Brain* **8**, 79 (2015).
160. Dickel, D.E. et al. Function-based identification of mammalian enhancers using site-specific integration. *Nat Methods* **11**, 566-571 (2014).
161. Cai, C. & Grabel, L. Directing the differentiation of embryonic stem cells to neural stem cells. *Dev Dyn* **236**, 3255-3266 (2007).
162. Lonnerberg, P. et al. Regulatory region in choline acetyltransferase gene directs developmental and tissue-specific expression in transgenic mice. *Proc Natl Acad Sci U S A* **92**, 4046-4050 (1995).
163. Singh Roy, N. et al. Enhancer-specified GFP-based FACS purification of human spinal motor neurons from embryonic stem cells. *Exp Neurol* **196**, 224-234 (2005).
164. Zhou, Q. & Anderson, D.J. The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification. *Cell* **109**, 61-73 (2002).
165. Mizuguchi, R. et al. Combinatorial roles of olig2 and neurogenin2 in the coordinated induction of pan-neuronal and subtype-specific properties of motoneurons. *Neuron* **31**, 757-771 (2001).
166. Sun, T. et al. Evidence for motoneuron lineage-specific regulation of Olig2 in the vertebrate neural tube. *Dev Biol* **292**, 152-164 (2006).
167. Zhang, X., Horrell, S.A., Delaney, D. & Gottlieb, D.I. Embryonic stem cells as a platform for analyzing neural gene transcription. *Stem Cells* **26**, 1841-1849 (2008).
168. McCreedy, D.A., Rieger, C.R., Gottlieb, D.I. & Sakiyama-Elbert, S.E. Transgenic enrichment of mouse embryonic stem cell-derived progenitor motor neurons. *Stem Cell Res* **8**, 368-378 (2012).
169. Kwasnieski, J.C., Mogno, I., Myers, C.A., Corbo, J.C. & Cohen, B.A. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc Natl Acad Sci U S A* **109**, 19498-19503 (2012).
170. Sharon, E. et al. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat Biotechnol* **30**, 521-530 (2012).
171. Nam, J., Dong, P., Tarpine, R., Istrail, S. & Davidson, E.H. Functional cis-regulatory genomics for systems biology. *Proc Natl Acad Sci U S A* **107**, 3930-3935 (2010).
172. Tuan, D.Y., Solomon, W.B., London, I.M. & Lee, D.P. An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "beta-like globin" genes. *Proc Natl Acad Sci U S A* **86**, 2554-2558 (1989).

173. Hug, B.A., Moon, A.M. & Ley, T.J. Structure and function of the murine beta-globin locus control region 5' HS-3. *Nucleic Acids Res* **20**, 5771-5778 (1992).
174. Inoue, F. et al. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res* **27**, 38-52 (2017).
175. Petrillo, C. et al. Cyclosporin a and rapamycin relieve distinct lentiviral restriction blocks in hematopoietic stem and progenitor cells. *Mol Ther* **23**, 352-362 (2015).
176. Yaragatti, M., Basilico, C. & Dailey, L. Identification of active transcriptional regulatory modules by the functional assay of DNA from nucleosome-free regions. *Genome Res* **18**, 930-938 (2008).
177. Murtha, M., Wang, Y., Basilico, C. & Dailey, L. Isolation and analysis of DNA derived from nucleosome-free regions. *Methods Mol Biol* **977**, 35-51 (2013).

Appendix One

Supplementary Figure for Chapter Two "An optimized, broadly applicable *piggyBac* transposon induction system"

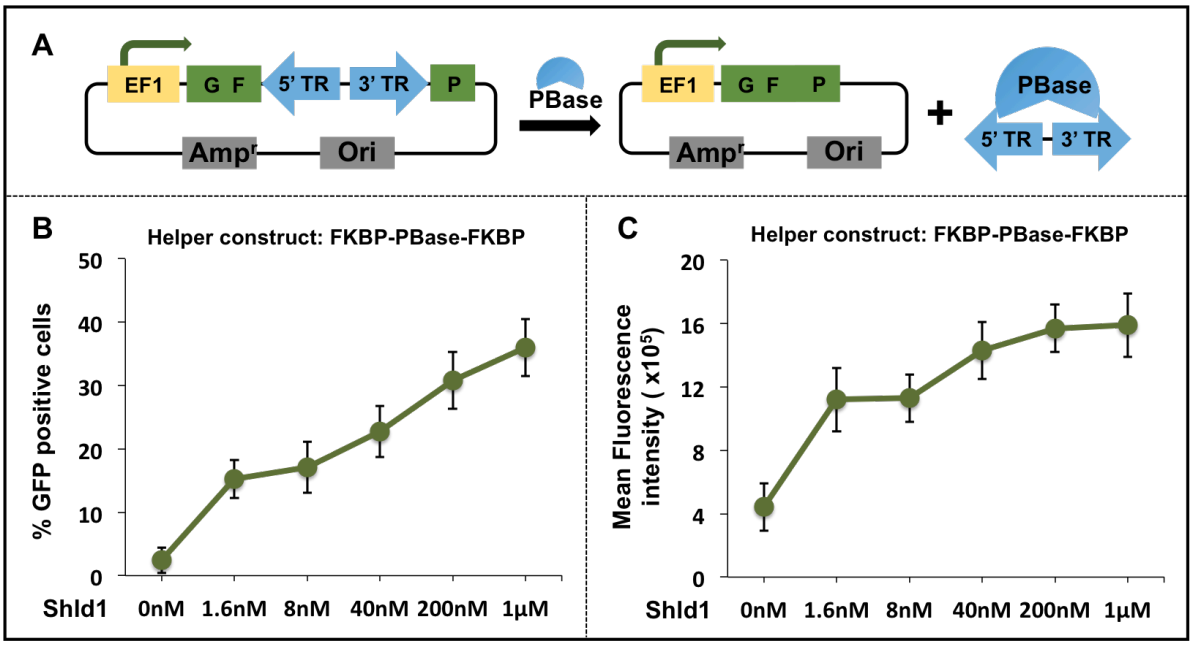


Figure 2.S1 Tunability of the FKBP-based PB transposon induction system measured by GFP-split donors in HEK293 cell lines. (A) Schematic illustration of GFP-split donor mediated transposition. The PB transposon disrupts the GFP gene. When the transposon is excised the GFP coding sequence is reconstituted and the cells fluoresce. (B) The percentage of GFP positive cells from the FKBP-PBase-FKBP fusion at different concentrations of Shld1. Experiments were done in triplicates. (C) The mean fluorescence intensity from the FKBP-PBase-FKBP fusion at different concentrations of Shld1. Experiments were done in triplicates.

Supplementary Figure for Chapter Three "Brd4-bound enhancers drive critical sex differences in glioblastoma"

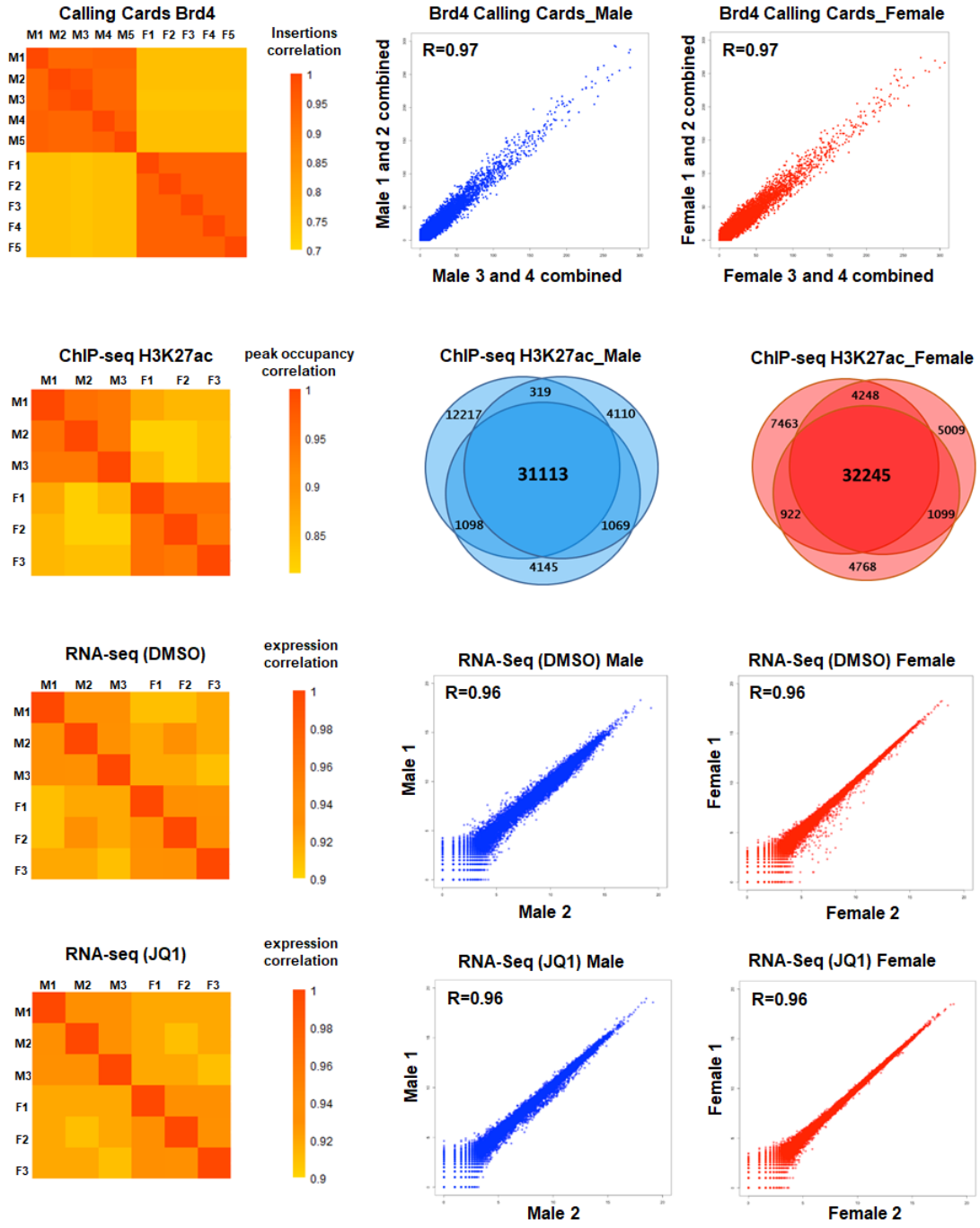


Figure 3.S1 Quality assessment of transposon Calling Card, H3K27ac ChIP-seq and RNA-seq experiments. Correlation heat map was made from replicates in male and female GBM cells. Venn diagram depicts the shared H3K27ac-enriched peaks among replicates in male and female GBM cells.

Table 1. Pathway analysis for sex-specific typical enhancers' genes downregulated following JQ1 treatment in male and female GBM cells*

Pathway description	Subset of Genes	Adjusted *p*-value	# of Genes
Male GBM Cells: 326 genes			
Metabolic Diseases	Nrxn1, Tgfbi, Lgals2, Igfbp3, Mical1, Il33, Prcp, Ly6c1, Bdh2, Crhr1, Slc5a7, Aipl1, Kcnab1, Itgb2, Scd1, Sirt5, Kcnk10, Pax9, Nfatc2, Pcolce2, Pld2, Foxo6, Adh7, Slc16a7, Trhr, Btbd3, Ptx3, Matn2, Shh, Slit3, Fundc1, Foxl1, Ank2, Sspn, Traf1, Cngb1, Apln, C1qtnf1, Decr1, Fbxo7, Fbn2, Chst1, Aqp5, Ptgfr, Ets2, Edil3, Akap6, Nrn1, Lgals3, Mark1, Adam22, Fars2, Ano1, Neu2, Faim2, Fbln1, Kcna4, Rasgrf1, Col25a1, Map2k6, Ar, Arntl, Alk, Car2, Pde1b, Mapk10, Cd38, Aldh3a1, Slc7a7, Aldh1a3, Camk1d, Cdkn2b,Nf1, Cdk6, Cxcl14, Rad51d, Stat6, Pde8a	0.00E+00	196
Chromosome aberrations	Nrxn1, Nalcn, Adssl1, Nup210, Klf12, Crhr1, Itgb2, Scd1, Pax9, Nfatc2, Plscr2, Ctnnd2, Cntnap4, Ptx3, Cradd, Shh, Slit3, Foxl1, Ank2, Traf1, Nhs, Crabp1, Fbn2, Ptgfr, Ets2, Pcyt1b, Akap6, Nrn1, Lgals3, Fars2, Ano1, Neu2, Fbln1, Bcl3, Gria1, Lsp1, Sfrp2, Kcnn3, Chst3, Dap, Dock8, Nov, Fgfbp1, Irak4, Tcn2, Spp1, Alk, Car2, Syn3, Sgca, Mapk10, Tenm4, Cdh23, Nrp2, Nanog, Grpr, Lef1, Zhx2, Flii, Cbr2, Mecom, Nox4, Sema5a, Cyb5a, Fgd2, Socs1, Nf1, Trib2, Cdk6, Rad51d, Stat6, Steap3, Ndp, Eps8, Sh2d4a, Gipc2, Epha3, Ndst3, Nptx1, Ank	2.00E-03	132
Glioma/Glioblastoma	Tgfbi, Igfbp3, Ly6c1, Slc5a7, Eef2k, Itgb2, Nfatc2, Pld2, Adh7, Ctnnd2, Il13ra1, Ptx3, Shh, Apln, Chst1, Ets2, Cd109, Lgals3, Adam22, Ano1, Gria1, Adgrb1, Sall2, Nid1, Sfrp2, Aqp4, Dap, Map2k6, Tcn2, Spp1, Nav3, Ahr, Ar, Il34, Alk, Plxdc1, Nrp2, Nanog, Grpr, Lef1, Xrcc5, Mlkl, Rapgef4, Aldh3a1, Slc7a7, Aldh1a3, Cdkn2b, Nox4, Socs1, Nf1, Lpar1, Cdk6, Thsd4	0.00E+00	126/76
Regulation of transcription	Klf12, Pax9, Nfatc2, Shh, Nrde2, Ets2, Bcl3, Sall2, Sfrp2, Crym, Ahr, Ar, Msx3, Arntl, Mapk10, Nanog, Lef1, Mlip, Zhx2, Xrcc5, Sbno2, Wwp2, Tsc22d3, Bhlhe22, Mecom, Nfatc4, Stat6, Sorbs3	0.00E+00	31
Integrin signaling	Ly6c1, Itgb2, Slit3, Sele, Fbxo7, Edil3, Lgals3, Spp1, Vcam1, Slc7a11, Pde8a, Cd9, Cspg4	7.00E-03	13
Stem cell proliferation	Shh, Fbln1, Sfrp2, Lef1, Mecom, Nf1	4.00E-03	6
Female GBM Cells: 270 genes			
Glioma/Glioblastoma	Rgs4, Sema3e, Mme, Thbs1, Pcdh18, Tnfaip3, Hmmr, Pappa, Vsig2, Ada, Syt1, Acss1, Phex, Galc, Rsad2, Pdgfd, Ngf, Trmt61a, Plcb4, Cdc25a, Ltbp1, Cxcl1, Epha1, Adam12, Wnt5a, Igf1, Grb10, Cxcl12, Lpar4, Trpm3, Angpt2, Ect2, Kif18a, Has2, Adm, Ifih1, F2rl1, Hhip, Ccna1, Nek2, Isg15, S1pr1, Prkaa2, Grin2b, Kif11, Dkk2, Tmem163, Wnt2, Tacr1, Bmpr1b, Usp1, Sulf1, Col3a1, Prps1, Sema3a, Cthrc1, Melk, Hgf, Stc2, Mmp16, Ptprb, Aldh1a1, Usp46, Ptprm, Xbp1, Ptprd, Il7, Pde4b, Ptprz1	0.00E+00	115/75
Chromosome deletion	Mme, Thbs1, Setbp1, Negr1, Pcdh18, Tnfaip3, Pappa, Rhoj, Sox3, Gypc, Ada, Phex, Inhba, Cdh11, Mob3b, Ngf, Pde1a, Plcb4, Epha1, Prepl, Wnt5a, Igf1, Grb10, Gcnt2, Pax2, Alx4, Angpt2, Itga4, En1, Pkp2, Bdh1, Itgb6, Runx1t1, Grin2b, Kif11, Npr3, Wnt2, Boc, Otx2, Bmpr1b, Gja5, Col3a1, Prps1, Sema3a, Magi2, Ripk2, Tg, Tll1, Mmp16, Cftr, Ptprb, Arap2, Aldh1a1, Ptprm, Casp14, Dcn, Ptprd, Cacna1c, Acp6, Etv4, Pde4b, Ptprz1	0.00E+00	71
Immune system process	Raet1e, Thbs1, Sla, Tnfaip3, H60b, Ada, Inhba, Rsad2, Pdgfd, Ifitm3, Oasl2, Il31ra, Cxcl1, Wnt5a, Igf1, Cxcl12, Frk, Angpt2, Ptgs2os, Itga4, Arhgef5, Ifih1, F2rl1, Oasl1, Cysltr1, Masp1, Tacr1, Rrs1, Cftr, Usp46, Ptprd,Acp6, Il7, Pde4b, Ptprz1	0.00E+00	44
Cell cycle	Thbs1, Tnfaip3, Rprd1b, Anapc1, Inhba, Tpx2, Cdc25a, Haus7, Wnt5a, Igf1, Kcnh5, Map9, Kmt5a, Ect2, Kif18a, Cenpw, Map4k3, Cyp26b1, Ckap5, Stk39, Ccna1, Nek2, Kdm8, Smarcd3, Kif11, E2f8, Rrs1, Magi2, Melk, Anapc15	3.00E-03	30
TGF beta signaling	Thbs1, Sema6d, Lox, Inhba, Ltbp1, Adam19, Adam12, Grem2, Gcnt2, Pax2, Cenpw, Has2, Pkp2, Itgb6, Sgcd, Runx1t1, Chrdl1, Bmpr1b, Col3a1, Cthrc1, Melk	3.00E-03	23
Transcription corepressor activity	Sox3, Tbx18, Kmt5a, Runx1t1, E2f8, Jazf1, Lmcd1	1.10E-02	7

* Subset of enriched pathways for sex-specific genes in male and female GBM cells

Table 3.S1 Pathway analysis for sex-specific typical enhancers' genes downregulated following JQ1 treatment male and female GBM cells.

Pathway description	Adjusted *p*- value	# of Genes
Metabolic process	1.50E-02	176
Glioblastoma	0.00E+00	86
Cell proliferation	0.00E+00	69
Cell differentiation	0.00E+00	63
Cell death	0.00E+00	57
DNA binding-Transcription factors	0.00E+00	35
Tumor angiogenesis	0.00E+00	29

*Subset of pathway enrichment analysis

Table 3.S2 Pathway enrichment analysis for differentially regulated genes in male and female GBM cells