Chromatin Deregulation in Cancer

John E Lazar

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

John A. Stamatoyannopoulos, Chair

Marshall S. Horwitz

Evan Eichler

Program Authorized to Offer Degree:

Genome Sciences

University of Washington

Abstract

Chromatin Deregulation in Cancer

John E Lazar

Chair of Supervisory Committee:
John A Stamatoyannopoulos
Department of Genome Sciences

During cancer development, committed cells are reprogrammed to malignant cell states. Genetic and epigenetic changes to chromatin complexes contribute to the reprogramming by disrupting normal cell identity and activating oncogenic phenotypes. How the complexes regulate these phenotypes, however, is often unclear: perturbing chromatin modifying complexes affects restricted—and often context dependent—sets of genes despite causing global changes to the chromatin state. To understand how general chromatin modifiers regulate specific cancer phenotypes, we altered chromatin remodeling complexes linked to cancer development and studied the relationship between their effects on chromatin and their control of gene expression.

We first mapped chromatin accessibility in two dissimilar cancer types, small cell lung cancer(SCLC) and acute myeloid leukemia(AML), to measure the spectrum of regulatory changes in cancer and identify factors underlying the oncogenic chromatin state. We observed frequent expansion in chromatin accessibility at cell type specific DHSs that contain low affinity motifs of diverse lineage specifying transcription factors (TFs). We next in-

hibited a histone demethylase, LSD1, in a panel of SCLC and AML lines cancer lines to study how different chromatin states direct the effects of a chromatin modifier. Despite a shared mechanism of LSD1 recruitment in both cancers, LSD1 inhibition affected cancer specific expression programs due to the preferential response of labile regulatory elements whose activity was most tightly linked to TF dosage. Finally, we tested the effect of a recurrent somatic mutation in a chromatin remodeler by reactivating the BAF complex member SMARCA4 in a SMARCA4-null lung adenocarcinoma cell line. We observed low affinity TF binding sites were preferentially affected by SMARCA4 reactivation and the domain level organization of such sites directs the effects of global chromatin remodeling to key developmental regulators.

Across experiments, we find a common class of regulatory elements is affected by multiple unrelated regulatory complexes. The affected elements contain homotypic, low affinity binding motifs, which make them inherently sensitive to variation in the concentration of bound regulatory factors. During normal development, low affinity sites allow for the precise activation of regulatory elements in response to environmental cues. We therefore propose the intrinsic lability of developmental regulatory elements makes them preferentially susceptible to the changes in the regulatory environment during cancer development and targets the global effects of chromatin remodelers to highly specific developmental phenotypes.

Dedication

For Eileen and George Lazar

Acknowledgements

There are many people who have made this thesis possible. I owe them a great deal.

I would first like to thank all my colleagues at the Altius Institute for Biomedical Sciences. The work presented in this thesis was a collaborative endeavor and would not have been possible without the assistance of many. In particular, I would like to acknowledge Shane Neph and Scott Hansen, who helped me understand the intricacies of working with genomic data when I first joined the lab. And I would like to thank Sandra Stehling-Sun for her extensive contributions to the SMARCA4 project.

My time in graduate school would not have been as productive, or enjoyable, without help from three senior members of the lab, Jeff Vierstra, Alister Funnell, and Wouter Meuleman. Their support, whether suggesting ideas that might be valuable to pursue, helping me to revise figures and manuscripts, or simply talking about the important questions in gene regulation, was essential for my progress through graduate school and development as a scientist.

I would also like to thank my thesis advisor, John Stamatoyannopoulos, for his guidance. Throughout graduate school his encouragement to understand the historical context of scientific questions proved invaluable when grappling with the challenges that arise during research. His guidance on how to distill complex situations down to the essential unknowns is something I will take with me both in medicine and to future research.

Finally, I'm grateful for all the friends and family who have supported me during graduate school, inside the lab and out. I'd like to thank Greg Findlay for the extensive discussions about science and support through the ups and downs of graduate school. My brother and sister, Daniel Lazar and Caitlin Wynne, have constantly been there for me and I'm grateful for their friendship, love, and encouragement.

My parents, George and Eileen Lazar, have always prioritized my education. I am thankful for the sacrifices they made to provide me with opportunities available to few other people. They have encouraged me to pursue my interests wherever they lead me and have modeled the work ethic necessary to turn those goals into reality. Their support, now and growing up, made this possible.

Contents

List of Figures

Chapter 1: Epistasis between lineage specifying TFs and chromatin modifiers in development and disease

Abstract

Development of an organism from a single cell embryo into millions of terminally differentiated cell types requires exquisite control of the gene regulatory programs that define cell state. Cells must remain sensitive to subtle differences in developmental cues when facing cell fate decisions but otherwise remain stably committed to a given cell type. The phenotypes of cancer cells suggest escape from the processes that enforce cell fate is a critical step in the development of cancer. Recent advances in high throughput regulatory assays allow for the genome-wide characterization of regulatory activity and offer the opportunity for an unbiased look at how normal regulatory mechanisms are altered in malignant cell states. In this work, we map the regulatory landscapes of cancer cells and their response to perturbing factors implicated in cancer development in order to understand the connection between the mechanisms regulating normal cell state and cancer phenotypes.

1.1 Structure of regulatory elements

Expression of genes in the appropriate cellular context requires sequence specific transcription factors (TFs) to bind DNA and recruit a series of protein complexes that culminates in the initiation of transcription by RNA polymerase. In eukaryotes, DNA is tightly packaged with proteins into chromatin, which blocks TFs and other transcriptional machinery from accessing the DNA. A critical step in gene activation, therefore, is the recruitment of chromatin modifying complexes to regulatory elements to alter the structure of chromatin to control the access of factors to DNA.

At the first level of chromatin organization, ~147 bp of DNA are wrapped around the core nucleosome, a histone octamer containing two copies each of the histones H2A, H2B, H3, and H4. In this state, individual nucleosomes can slide along the DNA allowing limited access to regulatory factors(Zlatanova, Seebart, and Tomschik 2008). Chromatin is further condensed from an initial "beads-on-a-string" organization with limited interaction between neighboring nucleosomes into higher order chromatin structures by the linker histone H1 and other structural proteins, including HP1 and Polycomb, that stabilize interactions between neighboring nucelosomes(McBryant, Adams, and Hansen 2006). In this condensed state, regulatory factors are further restricted from the DNA, resulting in stable repression of gene

expression at the locus.

Active regulatory elements are characterized by disruption of the chromatin structure when sequence specific TFs bind DNA in the place of the canonical nucleosome(Gross and Garrard 1988). The TFs can then recruit polymerase to initiate gene expression as well as recruit complexes that further modify chromatin to stabilize or repress TF binding. The chromatin modifying complexes can act on multiple aspects of chromatin organization. Several of these complexes, the SWI/SNF complex being the canonical example, directly remodel chromatin by mobilizing nucleosomes and thereby factor binding to DNA(Narlikar, Sundaramoorthy, and Owen-Hughes 2013). Other complexes covalently modify histones, and the chemical modifications to histones can directly affect chromatin structure by altering interactions between neighboring nucleosomes(Shogren-Knaak et al. 2006) or indirectly by serving as platforms to recruit regulatory machinery(Hecht et al. 1995). However, some care must be taken when interpreting the functional effect of histone modifications. Mutations abrogating the catalytic activity of chromatin modifying complexes often have minimal phenotypes compared to loss of function mutations(Dorighi et al. 2017; Rickels et al. 2017), suggesting histone modifying complexes play important roles independent of their modifying activity. Chromatin modifying complexes can also modify the DNA itself by methylating(or demethylating) CpG dinucleotides. DNA methylation can modify regulatory activity directly by changing the binding affinity of TFs for DNA(Schübeler 2015) or indirectly by recruiting methyl binding proteins that condense chromatin to create a repressive environment(McBryant, Adams, and Hansen 2006).

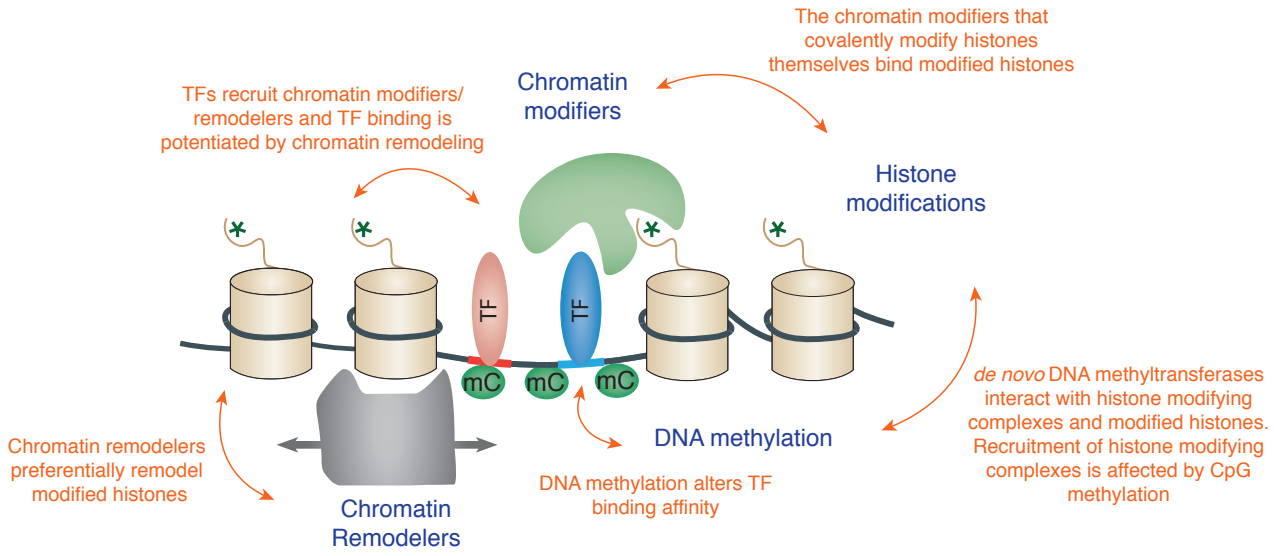


Figure 1.1: Interactions between chromatin modifications

The numerous modifications to chromatin do not occur in isolation, but rather as a network of reinforcing interactions (Figure 1.1). Chromatin remodeling complexes—which throughout this work we use as a general term rather than to imply any specific mechanism—often contain multiple distinct biochemical activities such as both nucleosome positioning and covalent histone modification(Basta and Rauchman 2015). The complexes also include subunits that recognize other histone modifications and, so the deposition of one histone modification may either potentiate or repress the addition of another mark(Du et al. 2015). Despite this complexity, regulatory element activation likely does not depend a on a stereotyped series of modifications, but instead reads out the balance of activating and repressive factors: loss of one activating complex can be rescued by the increasing a distinct activation pathway(Burns and Peterson 1997), and phenotypes due to loss of repressive chromatin complexes can be suppressed by removing activating complexes(Ringrose and Paro 2004).

The central step of regulatory elements is the recognition of the DNA sequence of a TF. However, the strength of this binding and the effect on downstream activation of gene expression is finely tuned by a series of interacting complexes that modify different features of chromatin.

1.2 Chromatin encodes cell state

During development, cells undergo extensive turnover of active regulatory elements. As fate is progressively restricted from multipotent cell types to terminally differentiated cell types, elements from alternative lineages are repressed and cell type specific elements are activated to drive the highly specialized phenotypes of differentiated cells (Hawkins, Hon, and B. Ren 2010; Stergachis et al. 2013; Zhu et al. 2013). This turnover is driven by cascades of transcription factor expression, which activate and repress the regulatory elements programs that define the organism's body plan and specify different cell types.

The commitment of cells to a specific fate can be conceptualized by a series of decisions between mutually exclusive TFs wherein the activation of one TF reinforces its own expression and represses the activation of the alternative lineage(Thomas Graf and Enver 2009). This TF antagonism can be seen both in patterning of the developing embryo where the positioning information is established by the ratio between countervailing anterior and posterior morphogen gradients(Wolpert 1994; Zagorski et al. 2017) as well is during differentiation where cells can be committed to a specific lineage either by the addition of a lineage specifying TF(Nerlov and Graf 1998; Visvader et al. 1992) or the repression of the TF that

specifies that alternative lineage(Galloway et al. 2005; Rhodes et al. 2005). The ability for activation of a single TF to make cell state decisions is perhaps best highlighted by reprogramming experiments, where the expression of a handful of key lineage specifying factor(s) is sufficient to establish a particular cell state(Davis, Harold Weintraub, and Lassar 1987; Takahashi and Yamanaka 2006).

Cell reprogramming experiments suggest cell state is, under certain circumstances, path independent, depending solely on the induction of expression of key master regulators. Many developmental decisions, however, are path dependent as they rely on a historical memory of previous regulatory events. For example, the same TF may induce differentiation into different cell types depending on the expression of TFs before it. In the myeloid lineage GATA2 expression causes differentiation into basophils if expressed after CEPBA, but if the order of expression of the two TFs is swapped the cell differentiates into an eosinophil(Iwasaki et al. 2006). Also, when cells from one location in the developing embryo are transplanted to an alternative location, they may not respond to the developmental signals at the transplanted location but instead develop as if they were in their previous location(Simcox and Sang 1983). Even forced reprogramming is partially dependent on the development path, or lineage, of a cell: TF combinations that are able to reprogram cells of the same lineage often fail to convert cells in alternative lineages(Ladewig, Koch, and Brüstle 2013).

For previous developmental decisions to affect the cell's response to later regulatory cues, the cell must maintain a memory of previous developmental decisions. This "epigenetic" memory exists during normal development as cells maintain patterns of gene expression across multiple cell divisions even after the stimulus that initiated the expression is removed(Mallo and Alonso 2013) and when exposed to new environmental cues(Simcox and Sang 1983). The mechanisms by which cells' epigenetic memory is propagated are not well understood. But several lines of evidence point to modifications to chromatin structure supporting the epigenetic state. First, there is direct evidence chromatin states can be stably inherited across multiple cell divisions even after the stimulus that induced them is removed(Cavalli and Paro 1999; Groudine and Harold Weintraub 1982). There is also indirect evidence from the phenotypes caused by mutations in chromatin remodeling complexes. Polycomb and Trithorax genes were initially defined based on phenotypes resulting from a failure to correctly maintain HOX gene expression after the initial stimulus that establishes HOX expression disappears. However these terms were then repurposed to apply to chromatin modifying complexes since Polycomb genes were found to be almost uniformly involved in establishing condensed chromatin (heterochromatin), while the Trithorax group genes encode proteins that maintain active chromatin(Ringrose and Paro 2004). Mutations

to chromatin remodeling complexes also affect the inheritance of expression for a subset of genes, such as olfactory receptors, that are stochastically expressed in a progenitor cell and whose expression is then deterministically maintained throughout that cells lineage(Elgin and Reuter 2013). Chromatin remodelers' role in the maintaining stable expression patterns appears to give them a role in enforcing memory of cell state: in cellular reprogramming experiments chromatin remodelers have not been identified as reprogramming factors, but screens of genes that that improve the efficiency of cellular reprogramming often uncover factors that maintain heterochromatin or DNA methylation(Costa et al. 2013; Onder et al. 2012).

In order to understand how chromatin state might maintain cellular memory, it is necessary to understand how the states can be maintained across the cell cycle. There are two points in the cell cycle where chromatin is disrupted and therefore the regulatory element must be reconstructed in the absence of the inducer that caused the initial change in activity: during DNA replication when DNA is unraveled from histones and during mitosis when most factors are evicted from DNA when the chromosomes condense to form mitotic chromatin. Patterns of both DNA methylation and histone modifications can be inherited at specific regulatory elements across the cell cycle offering potential mechanisms for epigenetic memory. Due to semiconservative replication of DNA, methylation state can be faithfully copied from the maternal to daughter DNA strand by a maintenance DNA methyltransferase that specifically recognize and methylate hemimethylated DNA(Song et al. 2011). However, while DNA methylation can be faithfully inherited, effect of this inheritance may be relatively modest. Although DNA methylation can alter the binding affinity of TFs for the DNA, in most cases the effects are not strong enough to specify gene expression(Schübeler 2015), and DNA methylation often appears to be consequence rather than a cause of TF binding(Brandeis et al. 1994; Lienert et al. 2011; Macleod et al. 1994). In differentiation models where the maintenance of methylation is lost in a specific lineage through knockout of the maintenance methyltransferase DNMT1, cell types typically differentiate normally but do not fully silence genes from alternative lineage. Therefore DNA methylation appears to play a secondary role in reinforcing silencing after cell fate decisions are made(Smith and Meissner 2013).

Compared to DNA methylation, it is less obvious how histone modifications might be inherited through the cell cycle. During DNA replication, parental histones and newly synthesized, unmodified histones are randomly deposited on either DNA strand, and no direct mechanisms for copying the histone marks are known. One model for stable inheritance of histone marks relies on the observation that histone modifying complexes bind with higher

affinity to the modified histones they create. Therefore even if the modified histones are diluted during DNA replication, the partial transmission of histone marks could recruit the modifying complex back to the locus to re-establish the chromatin state. Evidence from a few atypical loci suggest this model can hold, as a condensed chromatin states can be maintained through the cell cycle without DNA sequence dependent recruitment of repressive factors in a manner that depends on the histone methyltransferase activity and the histone modification(Audergon et al. 2015; Coleman and Struhl 2017; Ragunathan, Jih, and Moazed 2015). However, in the absence of sequence specific recruitment, the chromatin state is generally lost within a few cell cycles, suggesting the re-establishment of the chromatin state is not efficient without continuous recruitment of factors(Coleman and Struhl 2017; T. Wang et al. 2017). Nevertheless, when the chromatin state is maintained, it is sufficient to correctly maintain the cellular memory during development(Coleman and Struhl 2017).

What is noticeable from experiments that alter the inheritance of DNA methylation and histone modifications is the subtlety of the phenotypes: most cells differentiate correctly or are merely biased towards one cell type and a modest number of genes are aberrantly expressed(Coleman and Struhl 2017; Smith and Meissner 2013). The ability to properly specific cell states in the absence of chromatin-based inheritance can be explained without invoking additional regulatory mechanisms if global factors, e.g. transcription factors, are sufficient to accurately re-establish the regulatory program de novo in each cell cycle. Networks of self-regulating TFs are able to propagate their own expression and thereby maintain a memory of previous events(Alon 2007). Positive feed forward loops of TFs can also create switch-like transitions between expression states, and many single cell organisms undergo stable cell state transitions that are appear mediated solely by interactions between two TFs that define distinct metastable states(Raser and O'Shea 2005). The relative stability of cell fate decisions to perturbations of chromatin remodelers compared to disruption of transcription factor expression suggests many instances of epigenetic inheritance are likely explained by stable networks of TFs that re-establish the regulatory state.

Given the dominant role for transcription factors in defining cell states, it is reasonable to speculate about the evolutionary role for complex chromatin modifying systems, especially since there has been a large expansion in the number of chromatin remodeling complexes during vertebrate evolution(Ho and Gerald R Crabtree 2010). One potential reason for the expansion of chromatin based regulatory mechanisms is as a buffer against TF expression stochasticity. Metastable states driven by self-reinforcing TF expression are highly sensitive to variability in the expression of the factor, as seen in the lambda phage system where environment perturbation of one factor switches the system's state. In multicellular organisms,

stochastic fluctuations in the expression of lineage specific TFs in stem cells appear necessary for cell fate selection(Abranches et al. 2014; Chang et al. 2008; Warren et al. 2006). The level of noise in TF networks necessary for cell fate decisions may, however, be detrimental after cell commitment if it is high enough for additional stochastic cell state transitions to occur. The addition of chromatin based transcriptional memory could act to prevent state transitions due stochastic TF activity by reinforcing existing expression programs. Experimental evidence suggests chromatin remodeling does in fact play a role in controlling variability in expression. Stochasticity in single cells correlates with transitions in chromatin states(Krom et al. 2002; Raj et al. 2006) and mutations in chromatin remodelers increase expression stochasticity(Raser and O'Shea 2004; Xu, Zawadzki, and Broach 2006). Therefore, while TF based mechanisms are often sufficient to establish cellular identity, by reinforcing chromatin states, chromatin remodelers may add an additional layer of stability to the regulation of key genes and buffer cell states to fluctuations in the regulatory environment.

1.3 Loss of cellular identity drives cancer phenotypes

During cancer development the stable identity of normal cell types is apparently overthrown: cancer cells lose the phenotypes of differentiated cells and adopt characteristics of cell types from the early stages of normal development. These observations have inspired a model that cancer reverses development and represents an unwinding of the epigenetic clock to more primitive cell states(Johnston, S. B. Pai, and R. B. Pai 1992; Waddington 1935). The concept that cancer is a reprogramming of a cell from its normal lineage to an alternative developmental state has important implications for understanding cancer phenotypes, but it is not clear the extent to which the observed similarities between cancer cells and early developmental timepoints represent analogous regulatory states or simply convergent phenotypes due to cancers' uncontrolled proliferation.

The proposed link between cancer behavior and early developmental programs was originally based on a number of shared phenotypes including proliferation(Beard 1902), morphology, metabolism(Warburg 1956), and migration(Thiery 2002). Observations of cancer cells reactivating genes expressed at earlier stages in their development (i.e. fetal antigens) provided more direct support for the idea that cancer cells rely on activation of regulatory programs earlier in the lineage(Abate-Shen 2002; Ting et al. 1972). However, the expression of individual genes from fetal states might be a byproduct of increased transcriptional noise rather than re-establishment of complete regulatory programs.

Recently genomic technologies have allowed for a more comprehensive view of the extent

to which the expression and regulatory profiles that underlie malignancy in cancer contain features of normal development. Genome-wide expression data from a subset of cancers has shown the global expression profiles of tumors best matches early stages of the normal tissues' development(Borczuk et al. 2003; Kho et al. 2004), and mapping regulatory changes in cancer cells suggests they reactivate regulatory elements used at earlier timepoints in their development(Irizarry et al. 2009; Roe et al. 2017; Stergachis et al. 2013). The similarity of transcriptional and regulatory programs is supported by transplantation experiments—although as of yet shown only for a single type of cancers—that demonstrate cancer cells can recapitulate the behavior of the progenitors of their cell type of origin when engrafted into embryos(Kulesa et al. 2006).

The model of cancer cells as reprogrammed towards earlier developmental states is not in direct conflict with the genetic model of cancer, in which cancer phenotypes result from the progressive accumulation of mutations. Changes in cell state can be caused by mutations that activate or repress the requisite regulatory pathways. But certain characteristics of cancer cells are better understood by the reprogramming model than a purely genetic model. Cancers from different cell types can vary widely in the number of driver mutations required. The reprogramming model would predict that more primitive cell types would require a smaller number of mutations to convert to a malignant cell state. In fact, pediatric cancers tend to have a much smaller mutational burden, and in some pediatric cancers a single a single mutation that blocks differentiation is sufficient for cancer development(Johann et al. 2016; Lee et al. 2012). The adoption of cancer phenotypes through reprogramming also suggests cancers will frequently adopt novel phenotypes without a genetic change. Multiple cancer phenotypes including metastasis(Thiery and Sleeman 2006), vascularization(Scully et al. 2012; Soda et al. 2011), and drug resistance(Sharma et al. 2010) can be caused by purely epigenetic changes to the regulatory state. Finally, if genetic mutations promote cancer by causing a reprogramming-like event, we expect the mutations to have an effect in certain cellular contexts. The spectrum of cancer mutations provides indirect evidence for this, as even mutations to classic tumor suppressors and oncogenes occur in a highly cell type specific manner(Kandoth et al. 2013). Additionally there is direct evidence that cancer driver mutations are only functional within a certain cellular context: cancer cells that are reprogrammed to induced pluripotent stem cells lose their malignant state and revert to a malignant state only when differentiated into the original lineage of the cancer(Hochedlinger et al. 2004; Kim et al. 2013).

As a whole, these findings stress the link between cancer and development is not a facile one. Lineage specific alterations to developmental programs are central to the oncogenic

process.

1.4 Oncogenic reprogramming by lineage specific transcription factors

Given the link between malignancy and developmental transitions, it is perhaps not surprising that the key players that establish cell state during normal development are frequently altered in cancer cells. A common way cancer cells' regulatory programs are modified is by rewiring the transcriptional targets of the lineage specific TFs from the normal cell of origin to drive oncogenic phenotypes.

The transcription factor ASCL1 is a prototypical example of how the regulatory activity of lineage specifying TFs can be co-opted to drive cancer development. ASCL1 induces differentiation of neural and neuroendocrine cell types(Chanda et al. 2014) and is required for the development of a neuroendocrine cell population in the lung(Borges et al. 1997). Despite regulating a differentiated cell type, overexpression of ASCL1 drives formation of lung neuroendocrine tumors(Linnoila et al. 2000). Small cell lung cancers (SCLC), a malignancy of pulmonary neuroendocrine cells, express ASCL1 at high levels and are dependent on its expression for survival(Augustyn et al. 2014; T. Jiang et al. 2009; Linnoila et al. 2000). Unlike traditional oncogenes, ASCL1 is not associated with cancer outside of neuroendocrine malignancies. Therefore ASCL1 is believed to act as a lineage survival dependency oncogene, promoting oncogenesis by activating lineage specific regulatory programs necessary for cancer cell survival(Garraway and Sellers 2006). Similar relationships exist between master regulators and malignancy in other cell types: IRF4 is the master regulator of B-cell identity and drives multiple myeloma pathogenesis(A. L. Shaffer et al. 2008); NKX2-1 is regulates lung development and is an oncogene in lung adenocarcinoma(Yamaguchi et al. 2013); and MITF is essential for development of the melanocyte lineage and overexpression in melanomas promotes cancer progression(Garraway, Widlund, et al. 2005).

The TF oncogenes can be activated through diverse mechanisms. Most straightforwardly, amplification of the gene(Kwei et al. 2008) or activating mutations in the regulatory elements controlling the TF expression(Mansour et al. 2014) genetically increase factor expression. Alternatively, mutations in the signaling pathways upstream of the TF indirectly increase expression; for example, in SCLC, recurrent mutations in the Notch signaling pathway cause ASCL1 overexpression(George et al. 2015). Additionally, fusions between the TF and other proteins may modify the activity of the TF without changing expression: in Ewings sarcoma

a fusion between the EWSR1 and FLI1 proteins causes aberrant oligomerization of FLI1, which activates additional target sites(Boulay et al. 2017). However, in many cases there is no obvious genetic lesion leading to dysregulation of the TF. In such instances, purely epigenetic changes to the cell state may play a role in the elevated levels of TF activity: within a genetically homogeneous population of melanoma cells, subpopulations are observe to stably express lineage specific TF, including the melanocyte specific lineage survival oncogene MITF, at high levels(S. M. Shaffer et al. 2017).

Unlike oncogenes that directly regulate cell proliferation or survival, in most cases it is unclear how these upregulation of lineage specifying TFs drives cancer phenotypes. One possibility is increased activity allows the TFs to regulate targets not found in the normal cell state. For example, IRF4 upregulates MYC in multiple myeloma but does not do so in normal plasma cells (multiple myeloma's cell of origin), and the activation of MYC is essential for IRF4's role in myeloma pathogenesis(A. L. Shaffer et al. 2008). Examples from a limited number of factors suggests the increased activity of the lineage specific TFs can expand the number of TF target sites: in SCLC, amplification of NFIB promotes metastasis by activating additional NFIB binding sites(Denny et al. 2016) while overexpression of androgen receptor (AR) cofactors in prostate cancer increases the number of AR targets to activate the expression of genes outside of the targets found in normal epithelium(Pomerantz et al. 2015). Notably many of these new sites arise at elements active earlier in development of the relative cell type: NFIB amplification activates regulatory elements active in the developing brain, where NFIB is involved in neural specification(Denny et al. 2016). Since several lineage specifying TFs, including ASCL1(Castro et al. 2011) and MITF(Garraway, Widlund, et al. 2005), regulate survival and proliferation at early stages in the normal development of their lineage, reactivation of the primitive programs through increased TF expression may be one way to produce proliferative phenotypes.

An interesting feature of the oncogenic TFs is that, though they act as oncogenes, they often appear to also have tumor suppressive functions. Complete loss of expression of the oncogenic TFs is often associated with aggressive cancer phenotypes and a worse prognosis(Hartman and Czyz 2015; Yamaguchi et al. 2013). This dual behavior could be explained if activation of the TF beyond levels found in the normal cell type activates novel phenotypes, but at normal levels of expression the TFs acts as a check on progression by maintaining phenotypes of the cell of origin.

It is clear that lineage specifying TFs from the cell type of origin play a central role in defining cancer cell states. More work, however, is needed to understand the differences in the TFs' regulatory activity between cancer and normal cell types and the mechanisms that

might underlie the rewiring of the TFs' regulatory networks in the malignant cell state.

1.5 Recurrent somatic mutations in chromatin modifiers affect diverse developmental programs

While lineage specific TFs are associated with cancer development in the cell type or restricted set of cell types where they are expressed, the chromatin modifying complexes they recruit are broadly expressed and often play a role in multiple cancer types. These factors, however, have highly pleotropic effects and regulate distinct gene sets in different cell types, making it difficult to understand their role in a specific oncogenic context.

Chromatin modification was first linked to cancer development by a number of the complexes that acted as classical proto-oncogenes (Lohuizen et al. 1991) or tumor suppressors(Versteege et al. 1998). But the importance of chromatin modifiers in oncogenesis was not fully appreciated until widescale sequencing projects identified mutations in chromatin components at the same frequency as the classic tumor suppressors and oncogenes, such as TP53 and RAS, that helped define the hallmarks of cancer(Hanahan and Robert A Weinberg 2000; Network, Weinstein, et al. 2013). Like lineage restricted TFs, chromatin remodelers appear to exert their oncogenic effects by altering regulatory programs related to cell differentiation(Dawson and Kouzarides 2016). However, the connection between the regulatory mechanisms of these complexes (i.e. their direct effects on chromatin) and the expression programs they affect is not clear.

Many chromatin modifying complexes have highly specific effects on gene expression despite broad effects on chromatin structure. One illustration of the complex relationship between the impact on chromatin and gene expression is the effect of deleting the linker histone H1. H1 binds DNA as it exits the core nucleosome, stabilizes the position of nucleosomes, and facilitates for higher order folding of chromatin(Thomas 1999). Deleting histone H1 causes a global change in nucleosome spacing and chromatin compaction(Fan et al. 2005). Despite the global chromatin remodeling, deletion of H1 has a highly specific transcriptional effect with fewer than 1% of genes showing greater than 2-fold change in gene expression(Fan et al. 2005). The discordance between a global chromatin response and specific transcriptional response is typical of altering chromatin components: deletion of factors involved in DNA methylation(Blattler et al. 2014), nucleosome positioning(Alver et al. 2017), histone modification(Lovén et al. 2013), and three-dimensional chromatin structure(S. S. P. Rao et al. 2017) all cause attenuated transcriptional changes compared to their global effects on

chromatin. Understanding the source of this transcriptional specificity is important since the phenotypes of chromatin remodelers are often linked to the deregulation of a specific set of genes. It is not uncommon for reactivation of a single target gene to suppress the developmental effects of the loss of a chromatin remodeler(Ho and Gerald R Crabtree 2010; Stankunas et al. 2008; Voncken et al. 2003).

Adding to the regulatory complexity of chromatin modifiers, the specific transcriptional effects of the same chromatin remodeler often vary widely between different cell types. highly pleotropic effects. For example, the BAF chromatin remodeling complex controls the expression of key pluripotency genes in stem cells, cardiac specific expression programs in cardiac progenitors, neural gene expression in developing neurons, and globin genes in erythrocytes(Hota et al. 2019). The cell type specific regulation of chromatin remodelers results in highly pleotropic phenotypes. Mutations in BAF have diverse effects on embryonic, cardiac, and neurological development: SMARCA4 loss of function mutations are embryonic lethal(Bultman et al. 2000), cause heart defects when conditionally knocked out in the developing heart(Stankunas et al. 2008), and cause defects in neural differentiation when removed from the neural stem cells(Matsumoto et al. 2006).

The diverse sets of target genes in different cellular contexts means the disrupting the same chromatin complex will have distinct effects in different cancer types. The frequency of mutations across different cancer types highlights these cell type specific effects as even chromatin components expressed in every tissue drive cancer in a specific subset of cell types. For example, mutations in the histone H3.3 are specific to pediatric neural tumors(Schwartzentruber et al. 2012) and pediatric sarcomas(Behjati et al. 2013), and the histone mutations cause increased proliferation and block markers of differentiation only in the progenitor cells of the tissues where they are found to be recurrently mutated(Funato et al. 2014; Lu et al. 2016). While the H3.3 mutations are a particularly striking example of the regulatory specificity of chromatin components, similar mutational patterns are observed for other chromatin remodelers such as EZH2 (a Polycomb complex member), which is specifically mutated in B-cell malignancies(Sashida and Iwama 2017), and the de novo methylase DNMT3A which is mutated specifically in acute myeloid leukemia and related myeloid malignancies(Ley et al. 2010; Walter et al. 2011). Even chromatin remodelers are mutated across multiple cancers show increased specificity for certain lineages: while mutations in SMARCA4 occur in many different cancers, the gene is uniformly mutated in certain cancer subtypes(Witkowski et al. 2014).

Not only do chromatin remodelers only promote cancer development in specific cellular contexts, the same chromatin complex may have pro-oncogenic or tumor suppressive effects

in different tissues. For example components of the BAF complex are typically believed to act as tumor suppressers, but in synovial sarcoma it acts an oncogene(McBride et al. 2018). Similarly EZH2 acts as an oncogene through gain of function mutations in B-cells but functions as a tumor suppressor in the myeloid lineage(Sashida and Iwama 2017). There is even evidence that chromatin remodelers can have opposite oncogenic and tumor suppressive effects at different stages of the same cancer(Glaros et al. 2008; Roy et al. 2015; Sun et al. 2017). The apparently contradictory behavior of the chromatin remodelers is linked to different sets of target genes at different stages in cancer progression, so understanding the relationship between cell type and the chromatin complexes target genes will be necessary to better understand their role in cancer development.

Given the connection between chromatin remodelers and cancer development, they have become popular targets for drug development(Helin and Dhanak 2013). However, the complex relationship between chromatin remodelers and critical target genes has hampered our ability to understand the critical functional consequences of small molecule intervention and predict which cancers might respond to epigenetic therapies. In multiple instances, small molecule inhibitors applied with a certain rationale were found to be effective through different mechanisms. For example, the sensitivity of multiple cancers to bromodomain (BRD), a protein domain that binds acetylated histones, inhibitors was first rationalized as a result of downregulation of the MYC oncogene by BRD inhibition(Delmore et al. 2011). But, while BRD inhibitors do downregulate MYC in certain contexts, in other cancers there was no correlation between MYC downregulation and sensitivity to MYC inhibition. Instead the overall transcriptional response to BRD inhibition is highly cancer specific and appears to affect differentiation programs in a cell type specific manner(J. Shi and Vakoc 2014). Similarly, DNA demethylating agents are frequently rationalized as anti-cancer agents through reversal of the DNA hypermethylation that occurs at the promoters of some tumor suppressors(Jabbour et al. 2008). However, in myeloid malignancies, demethylating agents also affect differentiation programs more broadly and results from clinical trials suggests their effects are independent of demethylation of specific tumor suppressors(Fandy et al. 2009; Shen et al. 2010).

The high mutation rate in chromatin remodelers indicates the complexes plays a central role in cancer development. This importance makes them attractive targets for therapeutic intervention. However, since their specific effects on gene expression are not clearly related to their chromatin remodeling activities, it will be necessary to understand the source of their regulatory specificity and how cellular context influences their effects before we can fully take advantage of cancers' dependence on changes in chromatin state.

1.6 Epistasis between lineage specifying TFs and chromatin modifiers suggest a common target for epigenetic dysfunction

Several lines of evidence suggest modulating the activity of a cell's lineage specific TFs might be the main mechanism by which chromatin remodelers affect cancer phenotypes. Many of the classic developmental phenotypes that prompted the study of chromatin remodeling (i.e. Polycomb and Trithorax phenotypes) are the result of mis-expression of TF master regulators. In multiple different cancers, the effect of mutations in chromatin remodelers has been directly linked to mis-expression of lineage specific TFs(Kadoch and Gerald R. Crabtree 2013) or a chromatin remodeler with recurrent mutations directly regulates the expression of an oncogenic TF that is itself frequently mutated(Vachtenheim, Ondrušová, and Borovanský 2010).

In addition to regulating the expression lineage specifying TFs, the chromatin remodelers linked to cancer development directly recruited by oncogenic TFs and are necessary the regulatory activity of the TF. For example, the EWSR1-FLI fusion protein, which drives Ewing's sarcoma, is dependent on the frequently mutated BAF complex for its regulatory activity(Boulay et al. 2017). Additionally, the fusion protein PML-RAR in acute myeloid leukemia aberrantly targets the DNA methylase DNMT3A to silence elements(Di Croce et al. 2002), and DNMT3A is itself is recurrently mutated in AML(Ley et al. 2010). A similar phenomenon also occurs outside of fusion proteins, where the oncogenic behavior of MITF in melanoma is dependent on its ability to recruit BAF(Serna et al. 2006), and the BAF subunit ARID2 is recurrently mutated in melanoma. In many signaling pathways, such as the MAPK pathway, mutations in different components of the pathway lead to similar downstream activation. Perhaps a similar effect might be seen between TFs and chromatin remodelers where various mechanisms of epigenetic dysregulation can funnel to a shared set of regulatory changes.

Chemical inhibition of chromatin remodelers also specifically affects the regulatory programs of the lineage specific TFs in a cancer specific manner. Inhibition of the same chromatin complex in different cancers is linked to the regulatory programs of the cancers' respective lineage specific TFs. In multiple myeloma, sensitivity to bromodomain inhibition is linked to the suppression of the IRF4 oncogenic programs(Conery et al. 2016; Ishiguro et al. 2019), while in AML, bromodomain inhibition specifically targets the activity of the hematopoietic TFs govern myeloid differentiation(Bhagwat et al. 2016) and in small cell lung cancer bromodomain inhibition affects the neuroendocrine TF ASCL1 that is essential for SCLC activity(Lenhart et al. 2015). In addition to the same chemical inhibition disrupting

a different key TF in different cancer types, chemical inhibition of different regulatory complexes has been shown to affect the same TF in a single cancer type. In SCLC, not only is the behavior of bromodomain inhibitors dependent on its effect on ASCL1, inhibition of transcriptional elongation(Christensen et al. 2014), and inhibition of the histone demethylase LSD1(Takagi et al. 2017) have similar ASCL1 dependent effects as bromodomain inhibition.

It is not clear what mechanisms might explain the tight correspondence between the regulatory programs affected by oncogenic TFs and those affected by chromatin remodelers. Although direct recruitment of the complexes by TFs might explain part of the phenomenon, mot interactions between TFs and chromatin complexes are driven by relatively non-specific binding so complexes can be recruited by a large set of TFs(Dyson and Wright 2005). Therefore, it is unlikely that recruitment on its own can explain the correspondence. It has been proposed that cooperativity between neighboring elements, termed super-enhancers, might explain the relationship(Lovén et al. 2013), but direct evidence of this is low, and the initial results suggest the phenomenon is much more widespread than can be explained by super enhancers(J. Shi and Vakoc 2014). Therefore, a better understanding of the relationships between lineage specific TFs and chromatin remodelers might help us understand the cell type dependent specificity of chromatin remodelers towards developmental and oncogenic regulatory programs.

1.7 Studying chromatin remodelers

Methods to map regulatory activity genome-wide are valuable tools to study how genes are regulated. Chromatin accessibility mapping, in particular, has proven to be a powerful technique to identify multiple classes of regulatory elements and directly measure changes in regulatory activity.

Regulatory elements are activated when a transcription factor binds DNA in the place of the core nucleosome. The exchange of the core nucleosome for TFs can be measured by any reagent that has differential activity at nucleosome bound and nucleosome free DNA (i.e. accessible chromatin). The cleavage rate of DNA by the endonuclease DNaseI is highly sensitive to the presence of a nucleosome and has traditionally been the most commonly used reagent to measure chromatin accessibility due to its incredibly high sensitivity towards nucleosome free regions and low sequence specificity(Gross and Garrard 1988). Historically, patterns of DNaseI cleavage were measured at an individual locus by Southern blotting(Gross and Garrard 1988) or qPCR with site specific primers(McArthur, Gerum, and G. Stamatoyannopoulos 2001). Measurement of DNaseI cleavage events by DNA microarrays(Sabo et al.

2006) and later by next-generation sequencing (i.e. DNaseI-Seq) allowed for unbiased detection of open chromatin across the genome in a single experiment(Hesselberth et al. 2009). DNaseI-Seq has been successfully applied to map active regulatory elements across a broad range of cell and tissue types(Thurman et al. 2012), identify changes in regulatory activity across developmental and oncogenic states(Stergachis et al. 2013), and study the regulatory role of a single transcription factor(John, Sabo, Johnson, et al. 2008; John, Sabo, Thurman, et al. 2011).

In addition to mapping the location of regulatory elements, DNaseI cleavage rates also provides higher resolution information about TF binding and occupancy. Within the broader regions of DNaseI accessibility formed by the eviction of a nucleosome (~200bp), individual TF binding sites can create focal regions of decreased cleavage by protecting the bound DNA from cleavage by(Galas and Schmitz 1978; Vierstra and John A Stamatoyannopoulos 2016). These "footprints" extend the utility of DNaseI hypersensitivity mapping by allowing for precise quantification of TF occupancy with base-pair resolution.

In this work, we apply DNaseI-Seq to study how lineage specifying TFs and chromatin remodelers interact to establish oncogenic regulatory states. In Chapter #2, we comprehensively map chromatin accessibility across a panel of small cell lung cancer (SCLC) and acute myeloid leukemia (AML) cell lines to understand the spectrum of chromatin changes that occur during cancer development. We use the natural variation between samples to link differences in TF expression to changes in the chromatin accessibility and find, for both cancer types, differences in expression of lineage specifying TFs produce extensive variation in chromatin accessibility specifically at lineage and cell type specific regulatory elements from the cancers' normal developmental lineages. To understand how the variation in chromatin states studied in Chapter #2 modifies the effects of a chromatin remodeling complex, in Chapter #3 we characterize the effect of chemically inhibiting the histone demethylase LSD1 (*KDM1A*) in the same set of SCLC and AML lines. We observe the regulatory elements whose activities are most sensitive to variation in TF expression between cancer samples are also preferentially affected by LSD1 inhibition. The selective response at more labile regulatory elements produces a highly sample specific response determined by the pre-existing expression of lineage specific TFs and targets the effects of perturbing a general chromatin modifier to the developmental programs altered during cancer development. Finally, to understand if the connection between labile regulatory elements and chromatin remodeling applies in a different model system, in Chapter #4 we genetically rescue the expression of SMARCA4, a frequently mutated chromatin remodeler, in a well characterized lung adenocarcinoma model and study the relationship between recruitment of the com-

plex and its effects on gene expression. We observe SMARCA4 binding has heterogeneous effects on gene expression. Consistent with the results from Chapters #2-3, SMARCA4 recruitment to low affinity TF binding sites and to balanced chromatin domains is associated with greater chromatin remodeling and the largest changes in gene expression. The heightened response to SMARCA4 recruitment at these classes of regulatory elements targets the effects of SMARCA4 to key developmental regulators important for lung adenocarcinoma pathogenesis.

In three separate cancer types and for two different chromatin modifying complexes, we observe changes in the *trans* regulatory environment—whether variation in the expression of TFs or modifications to chromatin remodeling complexes—preferentially affect a common class of regulatory elements. The DNA sequences of the affected elements are characterized by homotypic TF motif content and low affinity TF motifs, sequence features that are predicted to cause highly sensitive responses to graded levels of TF expression(J. Jiang and Levine 1993). Developmental enhancers are thought to contain these sequence features in order for the element to activate only in response to a precise *trans* regulatory (i.e. cellular) state (Crocker et al. 2015; Farley et al. 2015). We therefore propose based on the results presented in this work that the intrinsic instability of developmental regulatory elements directs the effects of the global chromatin changes that occur during oncogenesis towards key developmental phenotypes.

Chapter 2: Lineage specific TFs drive extensive regulatory variation in SCLC and AML

This chapter (along with Chapter #3) is adapted with minor modifications from a manuscript in preparation: JE Lazar, S Hansen, S Neph, J Hallow, H Wang, R Sandstrom, D Bates, J Nelson, A Johnson, H Mohammad, C Carpenter, R Kruger, W Meuleman, and JA Stamatoyannopoulos. *Regulatory plasticity links epigenetic therapy to oncogenic and developmental programs.*

Abstract

During cancer development, escape from the epigenetic mechanisms that reinforce cell state leads to loss of cell identity, novel cancer phenotypes, and highly variable regulatory states. Here, to study regulatory variation in cancer, we comprehensively map DNaseI accessibility in two distinct cancer types: small cell lung cancer (SCLC) and acute myeloid leukemia (AML). Within both cancer types, chromatin accessibility is highly plastic with an average of 46% of DNaseI hypersensitive sites (DHSs) in SCLC and 35% of sites in AML present in fewer than half of samples. The variable DHSs are specific to the developmental origin of the cancer, and their activity is tightly linked to the expression of lineage specifying transcription factors (TFs) from the cancers' cells of origin. Higher levels of TF expression are associated with expansion of chromatin accessibility around low affinity TF motifs at sites that are more cell type specific, sensitive to variation in the expression of the TF, and associated with increased stochasticity in single cell expression. Expansion of chromatin accessibility at the motifs of lineage specific TFs is common across multiple cancer types, suggesting activation of plastic, low affinity sites may be a frequent regulatory event in cancer.

2.1 Introduction

Cancer progression requires malignant cells adopt multiple novel phenotypes to escape constraints on growth (Hanahan and Robert A. Weinberg 2011). Many of these oncogenic phenotypes including metastasis(Thiery and Sleeman 2006), vascularization(Scully et al. 2012; Soda et al. 2011), and drug resistance(Sharma et al. 2010; S. M. Shaffer et al. 2017) can be caused by purely epigenetic changes to the cell state. Based on the expression of a limited number of fetal antigens and gene expression profiles that resemble earlier developmental stages, aberrant reactivation of regulatory programs from normal development is thought to

be one way for malignant cells to manifest these phenotypes(Garraway and Sellers 2006). But the nature and extent of the regulatory changes that underlie this oncogenic reprogramming remain unclear.

During development, sequential expression of lineage specifying transcription factors (TFs) drives differentiation from multipotent stem cells to terminally differentiated states (Orkin 1995; Murry and Keller 2008; Weintraub et al. 1989). The TFs activate *cis* regulatory elements by binding DNA in place of a nucleosome (Felsenfeld et al. 1996) and recruit co-factors that further remodel chromatin to reinforce gene expression and cell state (Ringrose and Paro 2004). The cascade of TF expression cause extensive turnover of cells' regulatory landscapes by activating cell type specific regulatory elements and repressing elements from other lineages (Hawkins, Hon, and B. Ren 2010; Stergachis et al. 2013; Zhu et al. 2013).

The regulatory factors that specify chromatin state during development, including sequence specific transcription factors (TFs), nucleosome remodelers, and chromatin modifying enzymes are frequently mutated or dysregulated in cancers, altering the expression of developmental and oncogenic regulatory programs. As a result, the complexes have become popular therapeutic targets (Baylin and Jones 2011). Targeting the complexes produces highly specific transcriptional effects, and understanding the link between the often non-specific mechanisms of chromatin remodelers and their target genes has proved challenging. This challenge is in part due to the highly cell type specificity of the effects. Inhibition of ubiquitously expressed chromatin modifiers affects different subsets of genes in different cell types (J. Shi and Vakoc 2014); recurrent mutations in common regulatory factors are restricted to specific cancer types (Kadoch and Gerald R. Crabtree 2013); and disruption of the same complex may have pro- or anti-tumorigenic effects depending on the cell type or stage of cancer progression (Sashida and Iwama 2017; Sun et al. 2017). A more comprehensive understanding of how the epigenetic state of a cancer interacts with regulatory perturbations is therefore required.

Therefore, to study how chromatin context of a cell might modify the effect of an regulatory perturbation, we first mapped DNaseI accessibility in cancer cell lines of two cancers, the neuroendocrine malignancy small cell lung cancer (SCLC) and the hematopoietic cancer acute myeloid leukemia (AML), to characterize the spectrum of possible regulatory contexts. Cancer cell lines derived from different tumors represent independent paths of cancer development, maintain characteristics of the initiating cancer including tumorigenicity, and are genetically and epigenetically stable (Carney et al. 1985; Consortium et al. 2012; Drexler, Matsuo, and MacLeod 2000; Gerstein et al. 2012). We selected SCLC and AML as examples of cancers from disparate cellular origins with distinct sets of driver mutations; yet they

share share overlapping sensitivity to a set of epigenetic therapies including bromodomain (Lenhart et al. 2015; Zuber et al. 2011), CDK7 (Christensen et al. 2014; Y. Ren et al. 2015), and LSD1 inhibition (Harris et al. 2012; Mohammad et al. 2015; Schenk et al. 2012).

Chromatin accessibility within both SCLC and AML is highly heterogeneous, with frequent variation in accessibility at lineage restricted elements from normal cell types. The variable DHSs are characterized by a tight link between their accessibility and the expression of a small number of lineage specifying TFs active in the cancers' lineages. These variable DHSs contain low affinity binding motifs and are less buffered to changes in TF expression. We propose variable DHSs in a given cancer represents an intrinsically plastic regulatory compartment that preferentially responds to changes in the *trans* regulatory environment.

2.2 Results

High-resolution chromatin profiling reveals core vs plastic cancer regulomes

To map chromatin accessibility across a spectrum of SCLC and AML samples, we performed DNaseI-Seq in 37 SCLC and 20 AML cell lines (Figure 2.1a). For each cancer, we integrated peak calls from individual experiments to create a master list of putative regulatory elements (Methods). In SCLC we identified 708,030 non-overlapping elements (DHSs) with a mean of 201,992 elements per cell line, and in AML we identified 337,687 elements with a mean of 124,358 elements per sample. The number of elements covered at different sized subsets of the cell lines suggests we mapped sufficient samples to identify most DHSs active in the cancers (Figure 2.1b).

Chromatin accessibility is highly heterogeneous within SCLC and AML. The variation between different samples is similar in magnitude to changes that occur during normal development (Supplementary Table #1). Since different epigenetic trajectories over cancer development will be reflected in the variation in chromatin accessibility at the final state (Figure 2.1c), DHSs variable between sample were altered in course of cancer development. We therefore identified the sets of DHSs whose accessibility varied among SCLC and AML samples, and find, on a per sample basis, 49% DHSs in SCLC are active in fewer than half of SCLC samples and 41% of DHSs in AML elements are active in fewer than half of AML samples (Figure 2.1d). These plastic DHSs (i.e. DHSs that are present in only a fraction of cancer samples and therefore are altered over the course of cancer development) are primarily distal to genes' TSSs (Supplementary Figure #1a-b) and are specific to each cancer (Figure

2.1e). The predominance of low frequency, plastic DHSs in the accessibility landscapes of SCLC and AML is indicative of the cancers' highly altered regulatory states.

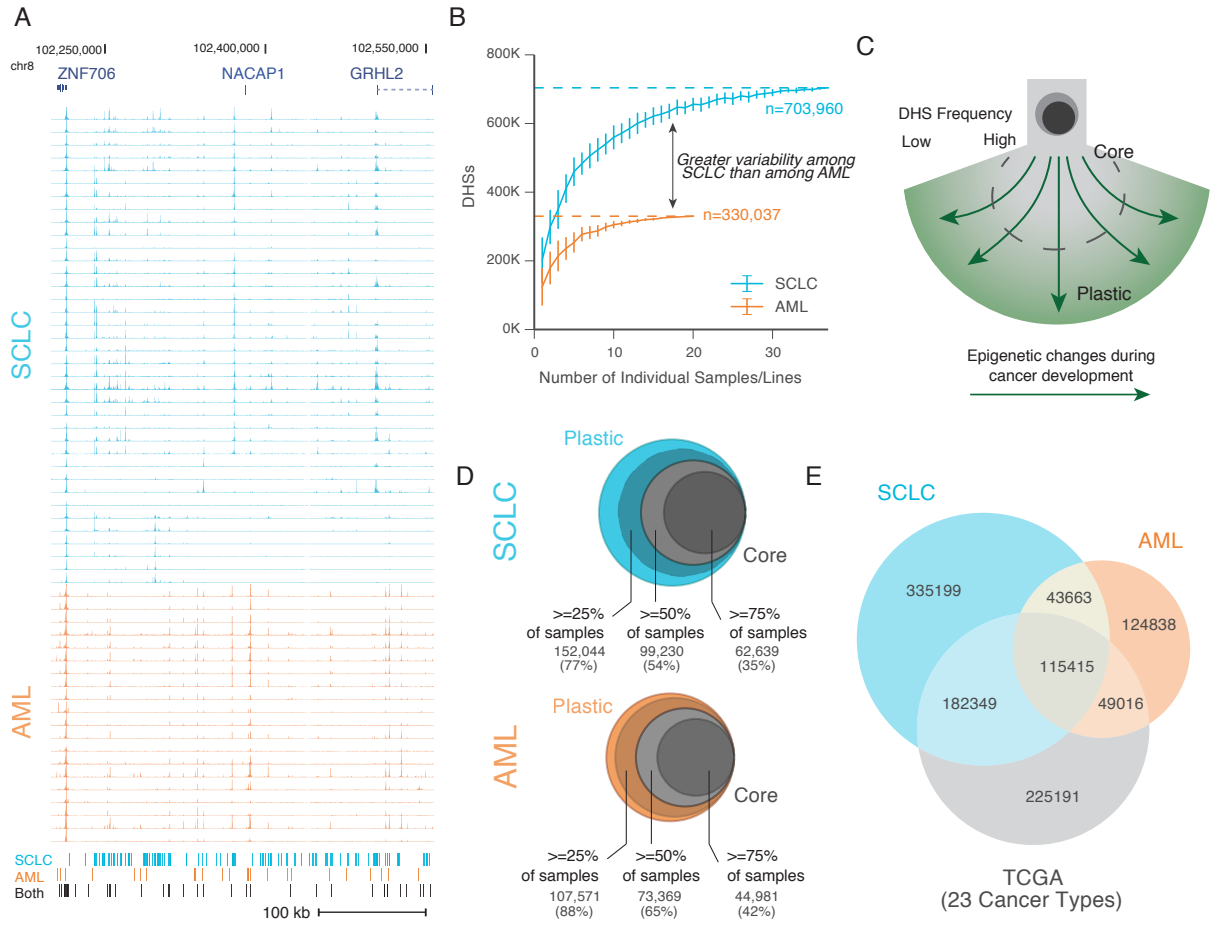


Figure 2.1: DNaseI Profiling of SCLC and AML

(a) DNaseI cleavage density across 37 SCLC and 20 AML cell lines from a representative genomic region.

(b) Saturation analysis showing the number of master list elements covered versus the number of cell lines assayed. Plotted is the mean +/- standard deviation of the number of elements observed from randomly sampled subsets of cells.

(c) Model of regulatory changes during cancer progression. Different epigenetic alterations during cancer development will be reflected in variation between the assayed end-state of the samples.

(d) Venn diagram showing the size of core (DHSs found in >=50% of samples) and plastic (DHSs found in <50% of samples) regulatory compartments in SCLC and AML. Values are the average over the 37 SCLC and 20 AML lines. Fraction of DHSs found in >=25% and 75% of samples are also included to show dependence on the frequency threshold.

(e) Venn diagram showing the comparison of DHSs identified in SCLC and AML samples to accessible sites identified in TCGA samples from 23 diverse cancer types.

Regulome plasticity connects oncogenic and developmental programs

To study the relationship between the observed chromatin variation and regulatory elements active during normal development, we next compared DHSs in SCLC and AML to chromatin accessibility in diverse cell types and tissues assayed by the ENCODE and Roadmap consortia (Roadmap Epigenomics Consortium et al. 2015; Thurman et al. 2012). Compared to existing DNaseI-Seq datasets, chromatin accessibility in AML cell lines is most similar to monocytes (CD14+) and hematopoietic progenitors (CD34+) while SCLC cell lines, consistent with its neuroendocrine origin and undifferentiated morphology (Park et al. 2011; Sutherland et al. 2011), are most similar to primitive neural tissues and poorly differentiated neural cancers (Supplementary Figure #2). The plastic DHSs in SCLC and AML are enriched for cell type specific DHSs from these cell types related to the cancers' cell of origins (Supplementary Figure #3a-b), and the frequency with which a DHSs appears within a cancer recapitulates the developmental persistence of the DHSs during normal development: for example, within AML DHSs restricted to the myeloid lineage are the most variable, while non-constitutive DHSs found across the hematopoietic lineage and broadly in hemangioblast derived cells show more consistent activity within AML (Supplementary Figure #3c).

The variation in chromatin accessibility correlates with differences in gene expression, and consistent with the association between cell type specific regulatory activity and cancer plasticity, the variable DHSs appear enriched near genes involved in cell fate commitment such as FOXA2 in SCLC or CD93 in AML (Figure 2.2a). To assess the relationship between variable chromatin and developmentally regulated genes more generally, we first linked DHSs to predicted target genes using the correlation between DHS activity and gene expression and then performed gene set enrichment analysis (GSEA) on genes ordered by the average variability of their linked DHSs (Methods). Genes linked to more variable DHSs are involved in developmental processes such as cell fate commitment in both AML and SCLC as well lineage specific terms such as myeloid cell differentiation in AML and epithelial cell differentiation in SCLC. In contrast, genes linked to highly stable chromatin are enriched for terms associated with constitutive cellular process such as mRNA processing and ribonuclear complex biogenesis in both SCLC and AML (Figure 2.2b).

Plasticity in cancer regulomes is concentrated at cell type specific DHSs and around key regulators of cell development. The variation in these elements reflects a diverse activation of normal developmental programs during oncogenesis.

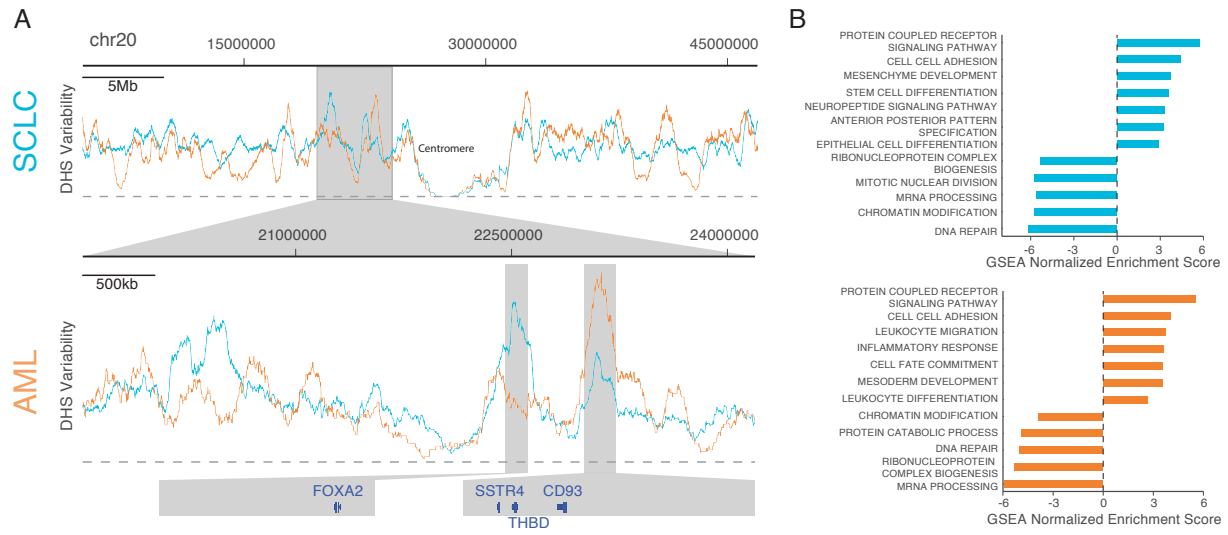


Figure 2.2: Enrichment of variable DHSs around developmental Regulators
a) Variability in DHS activity across a representative genomic region. DHS variability was quantified as the -Log(frequency) to reflect the information content in the DHS presence and averaged over 150 kb windows. To plot the average across SCLC/AML lines, each DHSs was weighted by the number of times it was observed across the samples. Regulatory variability peaks around key developmental regulators and cell type specific genes such as FOXA2 and CD93.
b)GSEA analysis of genes ranked by the average variability score of linked DHSs. Barplots show normalized enrichment score of terms significantly associated with genes in SCLC and AML linked to more variable (positive enrichment) and more stable (negative enrichment) DHSs.

Chromatin plasticity is tightly linked to expression of lineage specifying TFs

The sequences of variable DHSs are enriched for the motifs of lineage specifying TFs from the cancers' origin—such as ASCL1, NXK2-1, and NFIB in SCLC and CEBPA, SPI1, and IRF8 in AML—and depleted of motifs of TFs involved in housekeeping processes, including CTCF, SP1, and E2F1 (Supplementary Figure #3c-d). We therefore linked accessibility of variable DHSs to the activity of specific TFs by correlating the accessibility of each DHS across different SCLC and AML samples to the expression of TFs whose motifs are present in the DHS (Figure 2.3a).

The number of DHSs linked to TFs is highly unequal, with tens of thousands of DHSs linked to regulators of neuroendocrine cell state in SCLC and regulators of myelopoiesis in AML (Figure 2.3b). In SCLC and AML, a relative handful of TFs explain the majority of the linked DHSs: in SCLC 49 TFs (34 TFs in AML) account for 50% of the linked DHSs. Expression of the TFs linked to variable DHSs appear to be the main determinant of chromatin state in SCLC and AML, as hierarchical clustering of SCLC and AML samples based on genome-wide chromatin accessibility (which recapitulates known regulatory subtypes in

SCLC (Borromeo et al. 2016) and AML) groups samples with similar expression of the linked TFs (Figure 2.3c-d).

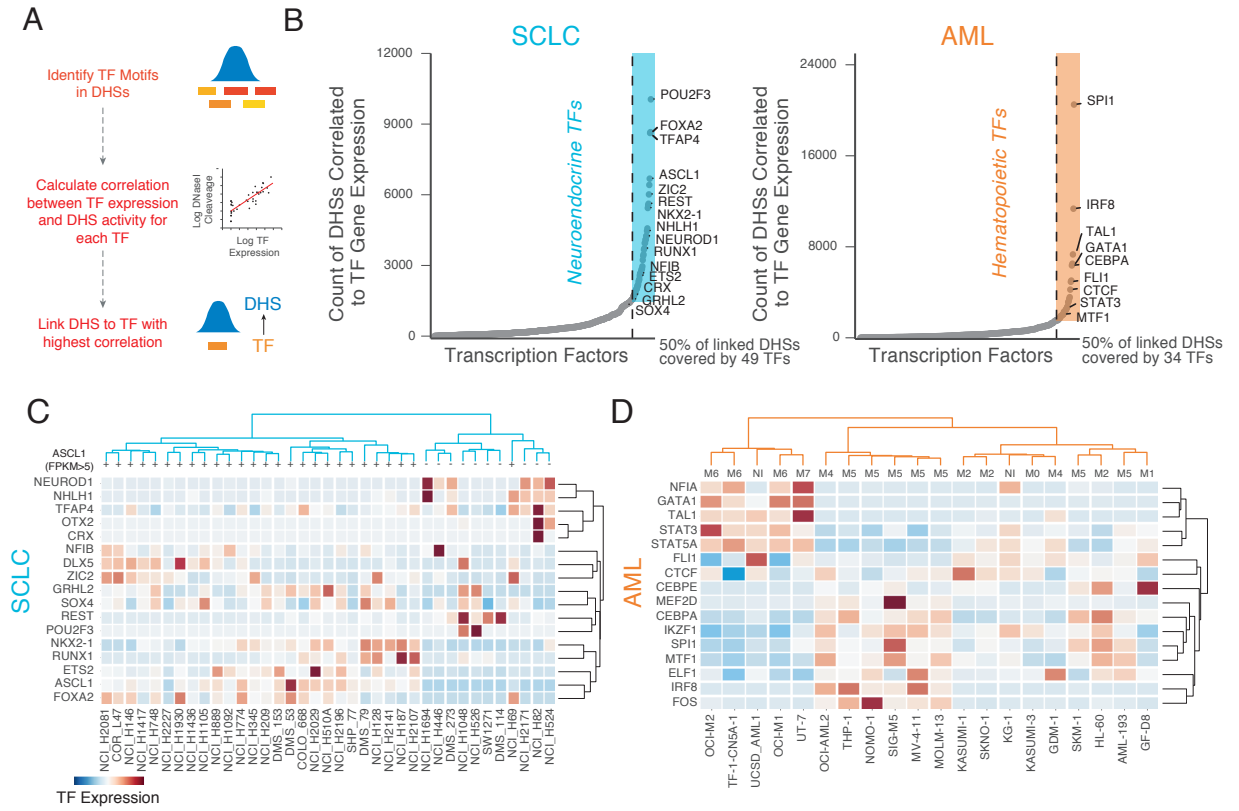


Figure 2.3: Cancer regulome plasticity is tightly linked to the expression of a small number of TFs

(a) Schematic of approach to link DHSs to TFs. For each DHS, we identified the TF whose expression was most highly correlated with the activity of the DHS (selecting from TFs whose motifs are present in the sequence of the DHS).

(b) Number of DHSs linked to TF expression across 687 TFs. In SCLC (left), TFs with the largest number of correlated DHSs included the neuroendocrine TFs ASCL1, FOXA2, and NKX2-1. In AML (right) master regulators of myeloid differentiation including SPI1(PU.1), IRF8, and GATA1 have the largest number DHSs correlated with their expression.

(c) Hierarchical clustering (Ward) of SCLC samples by DNaseI cleavage in master list elements demonstrates clear separation between ASCL1 (+) and ASCL1 (-) SCLC lines. Expression of other TFs involved in SCLC pathogenesis and predicted to drive the higher number of variable elements, such as NFIB, SOX2, and NEUROD1, appear to distinguish subgroups of chromatin accessibility.

(d) Hierarchical clustering (ward) of DNaseI cleavage separates M6/M7 cell lines from other FAB groups. Expression of TFs as predicted to drive large numbers of variable DHSs appear to underlie global clustering of chromatin accessibility landscape

We next tested whether the chromatin programs we identified might be also be affected by somatic mutations in the cancers. We leveraged exome sequencing data from CCLE (Barretina et al. 2012) to identify TFs associated with differential chromatin accessibility that also had recurrent coding mutations in the cell lines assayed. Consistent with AML's

low mutation rate, in AML we identify recurrent mutations only in CEPBA, but in SCLC we identify 15 TFs with mutations in more than one sample (Methods). In aggregate, these mutations are associated with a reduction in accessibility around the TF's motif (Figure 2.4a), and analyzing each TF separately we link coding mutations in TP73, which is mutated in 13% of SCLC patient samples (George et al. 2015) and is associated with the pathogenesis of other neural cancers including glioblastoma and neuroblastoma (Jancalek 2014), to a loss of chromatin accessibility around its motif (Figure 2.4b).

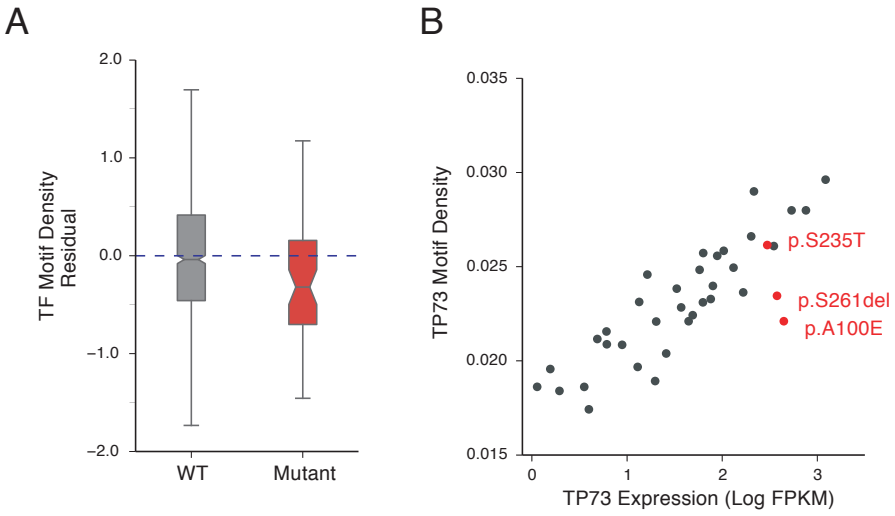


Figure 2.4: Recurrent coding mutations in TFs affect chromatin accessibility

(a) The relationship between Log TF expression and TF motif density in DHSs was fit for each TF identified with recurrent mutations. Boxplots of the residual of TF motif density from the predicted value based on TF expression in samples with WT TF sequence (grey) and mutant TF sequence (red). For TFs with a negative relationship between TF expression and chromatin accessibility around the motif, the signs of the residuals were flipped.

(b) Scatterplot of TP73 expression versus TP73 motif density in accessible chromatin across SCLC lines. Cell lines with mutations in TP73 are colored in red.

DHS plasticity is associated with a graded relationship between DHS activity and TF expression

We next analyzed the features of the individual DHSs whose activity was linked to variation in the expression of the lineage specific TFs. The activation of the linked DHSs in samples with higher TF expression occurs at both relevant cell-type and cancer specific elements (Figure 2.5). To quantify which DHSs are specific to higher levels of TF expression, we ranked each DHS linked to a TF by the average z-score normalized expression of the TF in cell lines where the DHS is active. The DHSs specific to higher levels of TF expression are

progressively more cell type and cancer specific (Figure 2.6, Supplementary Figure #4a-b). Using a single TF as an example, DHSs linked to NKX2-1 (a TF involved in lung development) in SCLC are a mixture of broadly accessibility elements as well as fetal lung and cancer specific DHSs. The DHSs specifically active at higher levels of NKX2-1 expression are progressively more enriched for fetal lung DHSs and, at the highest level of expression, cancer specific DHSs (Supplementary Figure #4a).

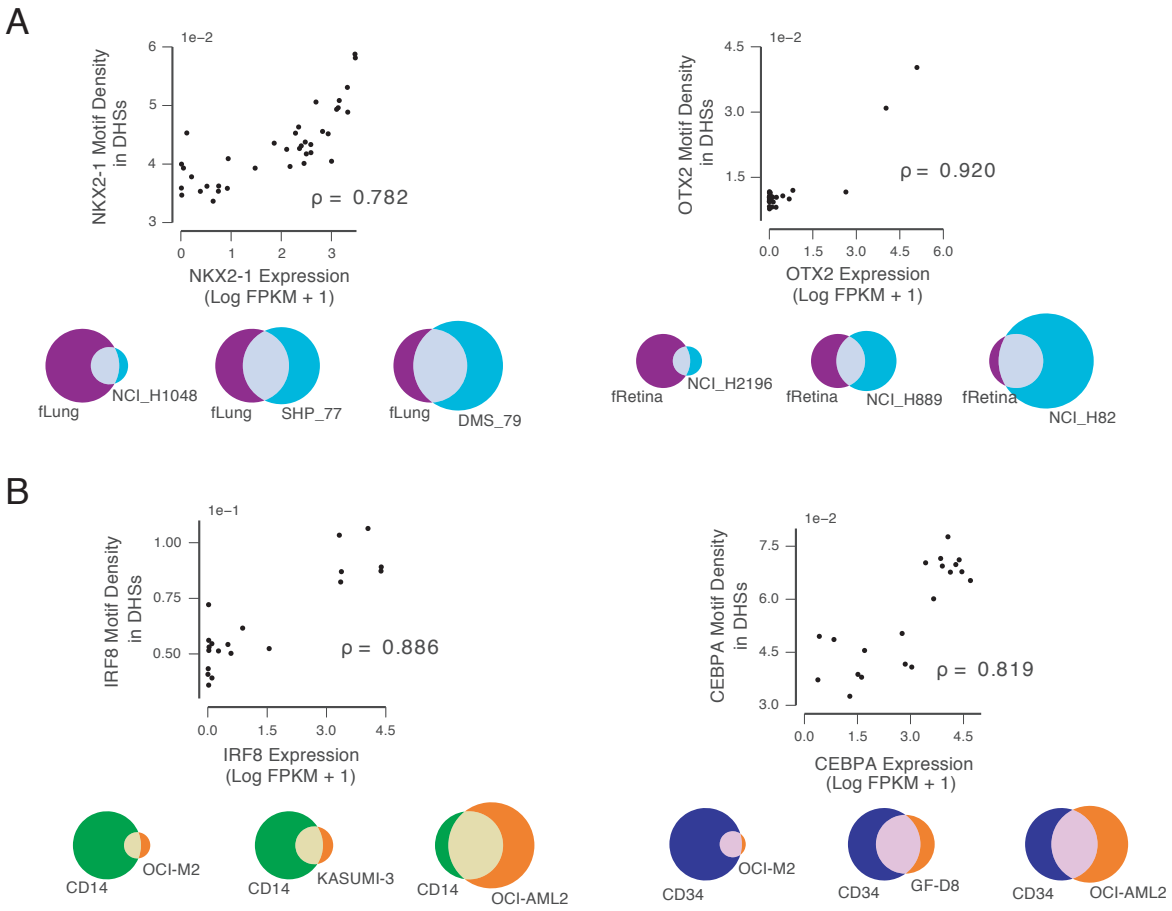


Figure 2.5: Increased expression expands TF compartment

(a) Scatterplots of the expression of NKX2-1 (left) and OTX2 (right) versus the TF's motif density in SCLC samples. Venn diagrams below scatterplots display the overlap of SCLC DHSs with DHSs from relevant cell types for each TF (fetal lung for NKX2-1 / fetal retina for OTX2) at low, medium, and high levels of the TF expression.

(b) As (a) for IRF8 (left) and CEBPA (right) in AML.

We hypothesized the DHSs specific to higher levels of expression represented less buffered regulatory elements whose activity is more sensitive to the dosage of the bound TFs. To quantify the relationship between TF expression and DHS activity, we fit a sigmoid rela-

tionship between TF expression and DHS activity and calculated the fold change between predicted DHS activity at the minimum and maximum TF expression (Methods). We observe that DHSs active at higher levels of expression show a greater sensitivity to the TF expression (Figure 2.6). These DHSs also have a more homotypic motif content (Figure 2.6), a feature of enhancers sensitive to graded TF expression (Crocker et al. 2015; Farley et al. 2015; Matthew T Maurano et al. 2015). To ensure these trends are not due to linking DHSs to TFs by correlation, we performed the same analysis on DHSs that overlap ASCL1 and NEUROD1 ChIP-Seq peaks in SCLC and SPI1 ChIP-Seq peaks in AML. We observe a comparable relationship between TF expression, motif density, and DHS variability (Supplementary Figure #5a-c), with the labile sites containing a higher fraction of low affinity motifs for the bound TF (Supplementary Figure #5d). The connection between cell type specificity and plasticity within cancer instances therefore appears mediated by the differential sensitivity of certain *cis* elements to alteration of the trans regulatory environment.

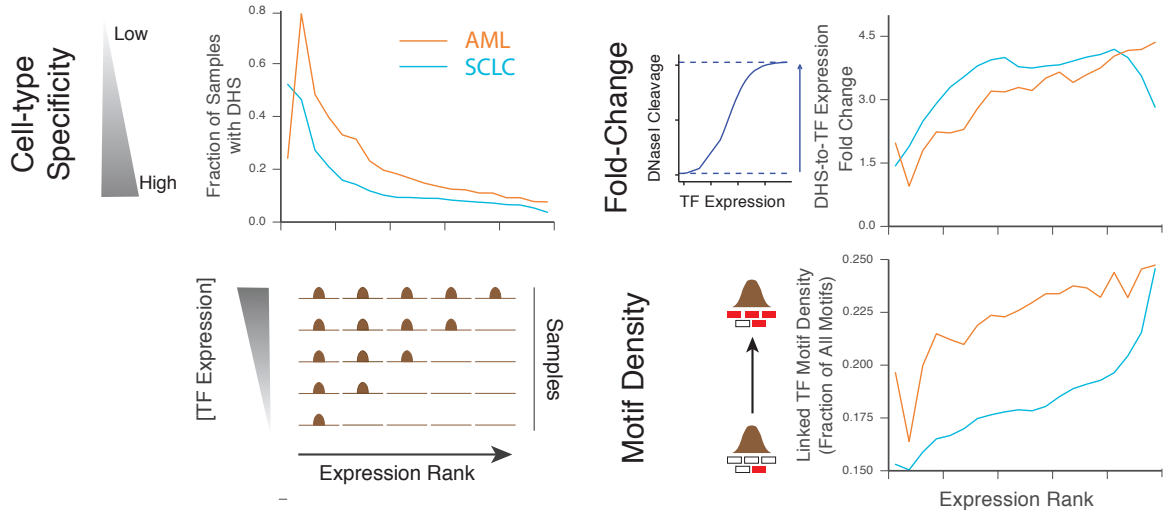


Figure 2.6: DHS plasticity is associated with a graded relationship between DHS activity and TF expression DHSs linked to TFs were ranked by the expression of the TF in samples where the DHS is present (for the minority of DHSs negatively correlated with TF expression, the sign of the quantity was flipped). Top: The average cell type specificity of the DHS, frequency of the DHS in SCLC samples, and fraction of cancer specific DHSs across the quantiles of the expression rank. Middle: Fold change of DNaseI cleavage across the range of TF expression across the quantiles of expression rank. Bottom: Fraction of the total number of base pairs by any motif that are covered by the motif of the linked TF across quantiles of the expression rank. DHSs active only at higher levels of TF expression are more cell type specific, show greater fold change in activity over the range of TF expression, and have more homotypic motif content.

Although global variation in chromatin accessibility is tightly linked to the expression of key lineage specifying TFs, genome-wide trends obscure the highly heterogeneous response of individual cis regulatory elements to differences in expression of trans regulatory factors. Cell type specific DHSs display higher sensitivity to TF expression and a distinct sequence

content associated with response to graded TF expression. The instability of developmentally regulated DHSs to changes in the trans regulatory environment suggests a link between developmental and oncogenic programs due to an intrinsic susceptibility of developmentally regulated elements to global epigenetic alterations during cancer progression.

Frequent expansion of chromatin accessibility around motifs of lineage specific TFs

In both SCLC and AML, we observed activation of novel, cancer specific DHSs at the motifs of lineage specifying TFs in samples with high expression of the TF. Lineage specifying TFs are frequently associated with cancer development, and cancers are often dependent on non-oncogene lineage specific factors (Garraway and Sellers 2006). We hypothesized the activation of novel regulatory elements would be one mechanism for lineage specifying TFs to act in cancer progression. For example, comparing chromatin accessibility in the prostate cancer cell line, LNCaP, to chromatin accessibility in prostate epithelium, we observe a dramatic increase in the relative accessibility around the HOXB, FOXA, and AR motifs (Figure 2.7b), which is consistent with their role in prostate cancer and previous reports of rewiring of the *cis* landscapes around these TFs during prostate cancer development (Pomerantz et al. 2015).

To test if high levels of accessibility around lineage specifying TFs was a common feature of chromatin accessibility in cancer, we calculated the skew of the z-score transformed motif density in cancer cell lines compared to normal samples. We observe significantly increased skew (Figure 2.7a, p-value <0.001, Wilcoxon signed rank test) in cancer samples compared to their normal counterparts suggesting expansion of chromatin accessibility around the motifs of lineage specifying TFs is a common feature of cancer samples' regulatory environments.

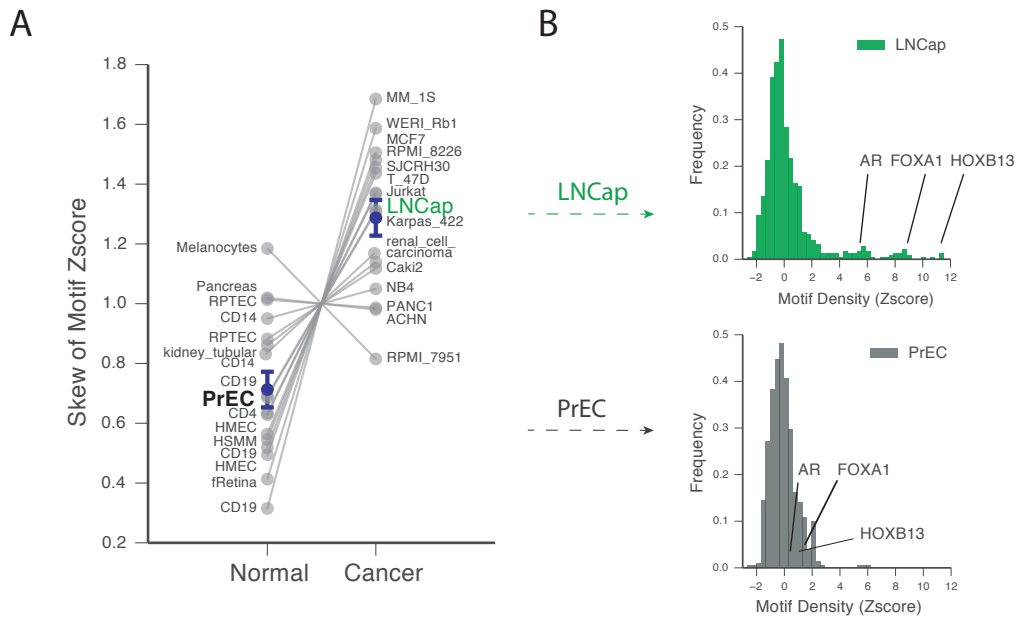


Figure 2.7: Expansion of chromatin accessibility around TF motifs in cancer

(a) Skew of motif density z-scores in matched normal and cancer chromatin accessibility landscapes. For comparison across cancers, values are normalized to the average of the normal and cancerous sample. The distribution of motif density (z-score) in DHSs profiles in the cancer samples have significantly greater skew compared to their matched normal counterparts (pvalue <0.001, Wilcoxon signed test). Mean +/- S.E.M across all samples is plotted in blue.

(b) Histogram of z-score normalized TF motif density in DHSs in and the prostate adenocarcinoma cell line LNCap (top) and normal prostate epithelial cells (bottom). Note the bias towards increased accessibility around the motifs of lineage specific TFs previously implicated in the pathogenesis of the cancer including AR, FOXA1, and HOXB13.

Regulatory plasticity is associated with increased single cell variability

Since changes in the activity of plastic regulatory elements appeared to be a common mechanism during cancer development, we wanted to investigate whether the plasticity in the regulatory programs that vary between cancer samples might also play a role in epigenetic variation within an individual tumor. We therefore performed single cell RNA-Seq on the SCLC cell line NCI-H446 (Figure 2.8a) and identified genes with larger than typical variation in expression between single cells (Methods, Figure 2.8b). These stochastic genes are preferentially linked to variable DHSs (Figure 2.8c) and include several TFs linked to variation in chromatin accessibility across SCLC samples that are expressed in NCI-H446, including NEUROD1 and SOX4(Figure 2.8a-b). This results connects variation variation among different cancer samples to variation within a single cancer sample and suggests the variation in chromatin accessibility between cancer samples may reflect an underlying regulatory plasticity of the elements that can be observed in other contexts.

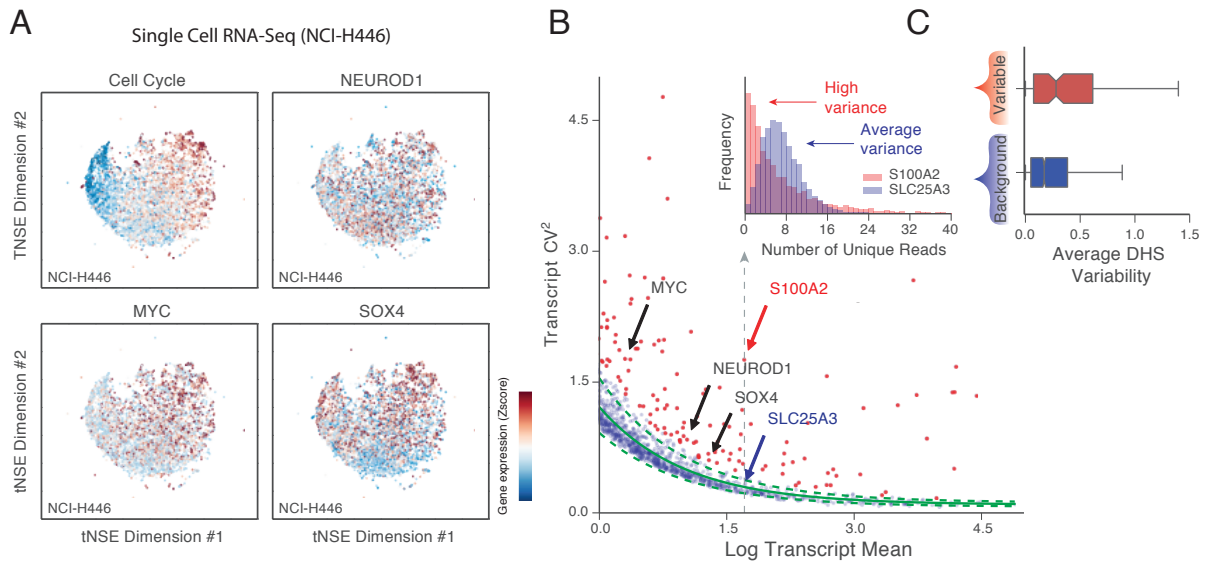


Figure 2.8: Stochastic gene expression near variable DHSs

(a) tSNE representation of single cell RNA-Seq data from the SCLC cell line NCI-H446. Cells are colored by the average expression of cell cycle associated genes (GO term GO:0007049) as well as the expression of TFs identified as playing a key role in SCLC chromatin variation.

(b) Identification of highly variable genes from scRNA-Seq in NCI-H446. The fit between a genes average expression and coefficient of variation is plotted in green. Dashed lines show the estimated 5th-95th percentiles of the distribution. Genes identified as having significantly higher variation than the global trend are colored in red. Inset shows histogram of gene counts for a highly variable and background gene with the same mean expression.

(c) Boxplot of average DHS variability (self-information) of linked DHSs for genes with variable expression at the single cell level and background genes. (pvalue <0.001, Wilcoxon rank test).

2.3 Discussion

Cancer phenotypes are frequently described by analogy to processes found during normal development. However, the correspondence of oncogenic and developmental regulatory programs remains incompletely understood. Here we present a comprehensive mapping of chromatin accessibility in two distinct cancer types to study the cancers' regulatory states and their relationship to the regulatory programs found in normal cell types.

In both SCLC and AML, we observe a central role for lineage specifying TFs from normal differentiation in defining chromatin accessibility in oncogenic states. Variation in the expression of key lineage specific TFs is assocciated with genome-wide differences in chromatin state. Many of the TFs linked to variation in chromatin accessibility are known to be involved in the pathogenesis of SCLC and AML. In SCLC, ASCL1 is a key regulator of pulmonary neuroendocrine development (Borges et al. 1997) and SCLC pathogenesis

(T. Jiang et al. 2009; Linnoila et al. 2000; Osada et al. 2005), while SOX2 (Rudin et al. 2012) and NFIB(Dooley et al. 2011) are a recurrently amplified, and TP73 is affected by frequent mutations and genomic rearrangements (George et al. 2015). In AML, CEBPA is recurrently mutated (Pabst et al. 2001) and decreased SPI1 (Pu.1) promotes leukemia formation (Rosenbauer et al. 2004).

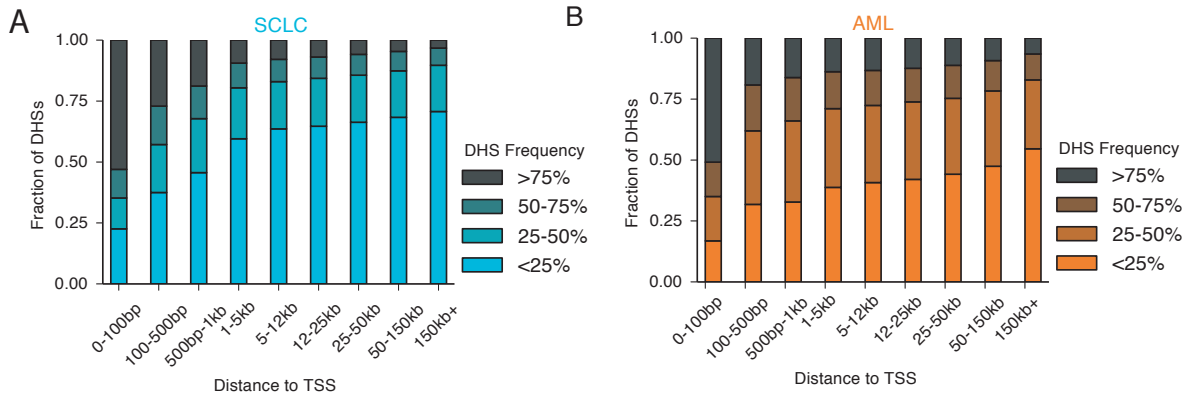
Variation in the expression of the lineage specifying TFs is most tightly linked with accessibility at cell type specific DHSs, linking the TFs to diverse activation of cell-type specific DHSs from neural/epithelial tissues (SCLC) and myeloid development (AML) between different cancer samples. The composition of TF binding motifs in the sequences of the cell specific DHSs suggests the heightened responsivity to differences in TF expression is an intrinsic feature of the elements that leads to the preferential alteration of these sites across cancer development. Key developmental enhancers are characterized by graded sensitivity to TF expression(Crocker et al. 2015; Farley et al. 2015), and our results imply the ability of developmental enhancers to respond to precise levels of TF expression will make them inherently more likely to be affected when the *trans* regulatory environment changes in cancer cells.

Lineage specifying TFs are also linked to the emergence of novel, cancer specific DHSs at the highest levels of expression as chromatin accessibility expands around the TF's motif. This expansion of certain lineage-specific TFs' regulatory programs appears to be a common feature of cancer-specific gene regulation, as we observe a bias towards increasing chromatin accessibility at specific TF motifs in multiple cancer samples compared to their normal counterparts. The emergence of these novel accessible sites provides the opportunity for TFs to regulate targets outside of normal developmental pathways and may play a role in the TFs' actions as lineage specific cancer dependencies. Additionally, since we connect more variable DHSs to variation in single cell expression, the novel sites may be more likely to stochastically vary across cancer development and thereby provide a substrate for the epigenetic evolution of the cancer. However, this plasticity of regulatory activity may lead the novel DHSs to be preferentially susceptible to alterations in the regulatory environment.

Methods

See Chapter #3

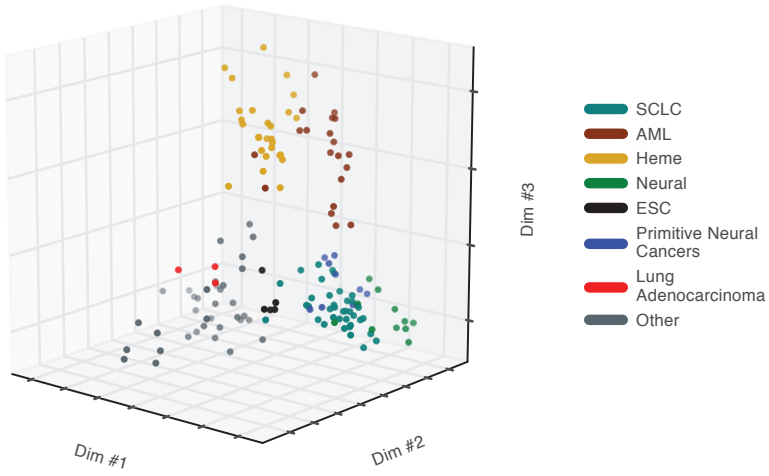
2.4 Supplement



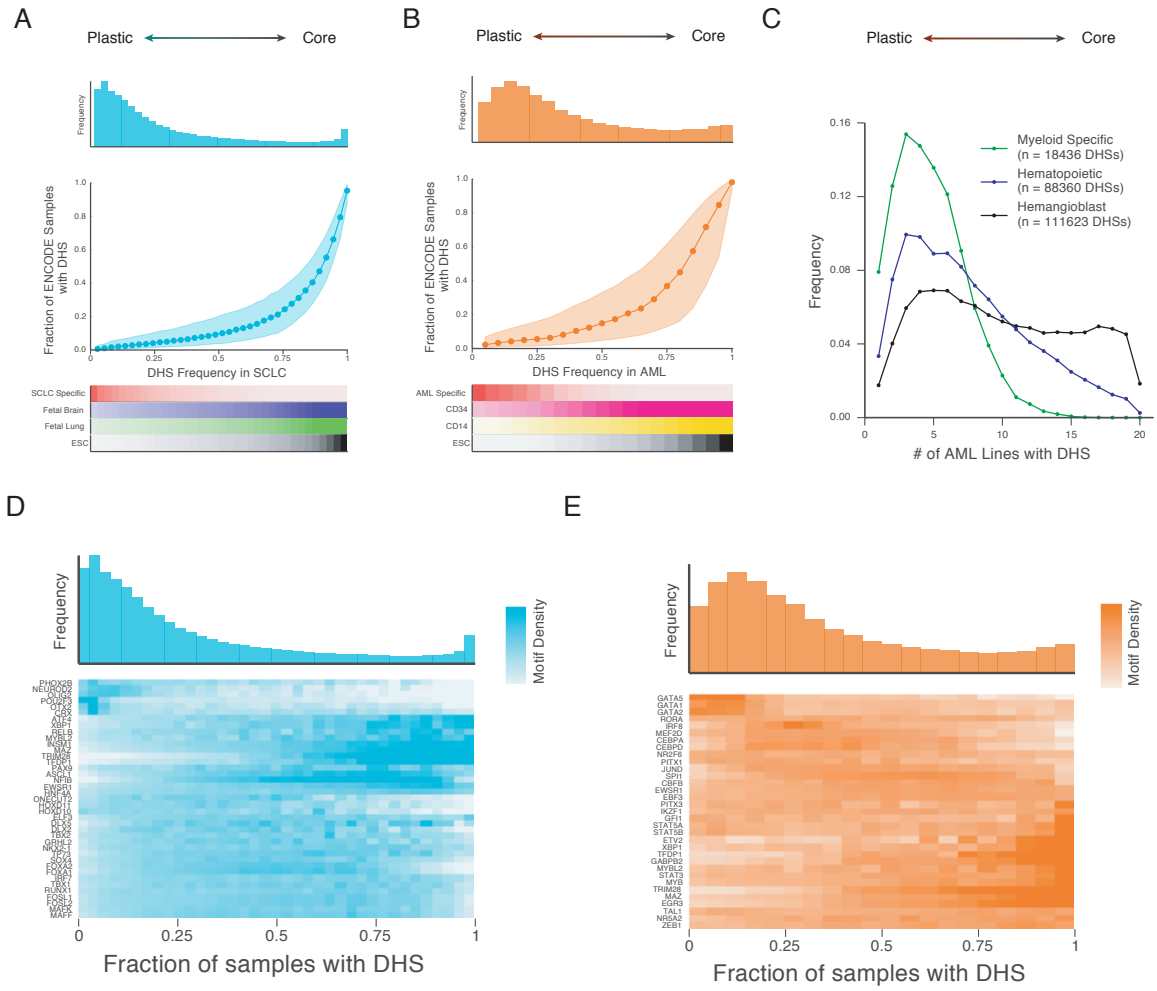
Supplement #1-Variable regulatory elements are distal to genes' TSSs

(a) Stacked barplot showing fraction of DHSs in each frequency bin across SCLC (bottom) lines as a function of distance to the closest TSS.

(b) As (a) for AML.



Supplement #2-Chromatin accessibility in SCLC and AML retains signatures of cell types of origin
Dimensionality reduction (PCA) of DNaseI cleavage in SCLC and AML datasets with various normal tissues included for reference. AML cell lines resemble normal hematopoietic cell types while SCLC samples lie between ESC and primitive neural cell types including fetal brain, fetal retinal, and fetal spinal cord. Other "small blue cell" cancers including rhabdomyosarcoma, retinoblastoma, and neuroblastoma fall close to SCLC samples in the lower dimensional space. Finally, SCLC samples group separately from the two lung adenocarcinoma cell lines (A549 and PC9).



Supplement #3-Cell type specific elements are variable over cancer development

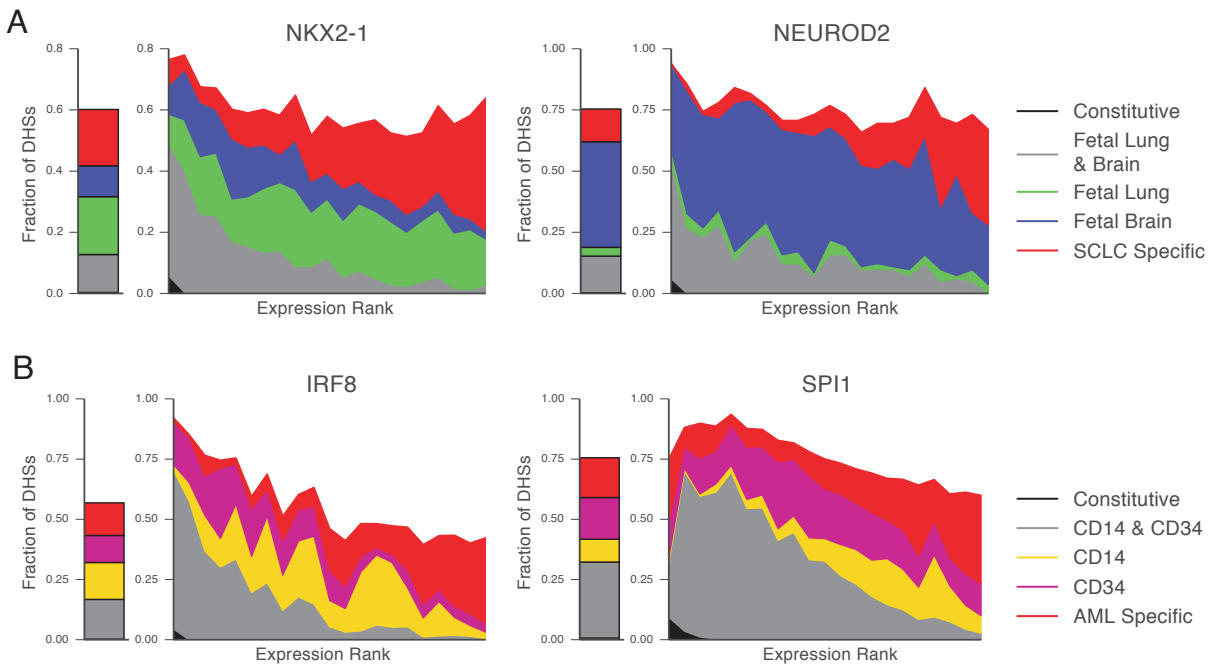
(a) Comparison of cell type specificity across 665 ENCODE samples to DHS frequency in SCLC. Top: histogram of DHSs frequency within 37 SCLC cell lines. Middle: Median frequency of DHS activity in 665 ENCODE samples across the spectrum of DHSs frequency in SCLC. Shaded area reflects 25th-75th percentiles of the distribution. Bottom: Overlap of SCLC elements with DHSs from related cell types across the spectrum of DHS frequency.

(b) As (C) for AML

(c) Comparison of DHS persistence across myeloid differentiation with DHS frequency in AML samples. DHSs present in the myeloid lineage were grouped based on whether they are myeloid specific (orange), present other hematopoietic lineages (blue), or present across the hemangioblast lineage (black). Broadly constitutive DHSs (DHSs active in ¿90% of ENCODE samples) were excluded from the analysis.

(d) Density of TF motifs in DHSs across the spectrum of DHS frequency in SCLC. Motifs of key lineage specific regulators such as ASCL1, NEUROD1, and NKX2-1 are enriched in the variable DHSs.

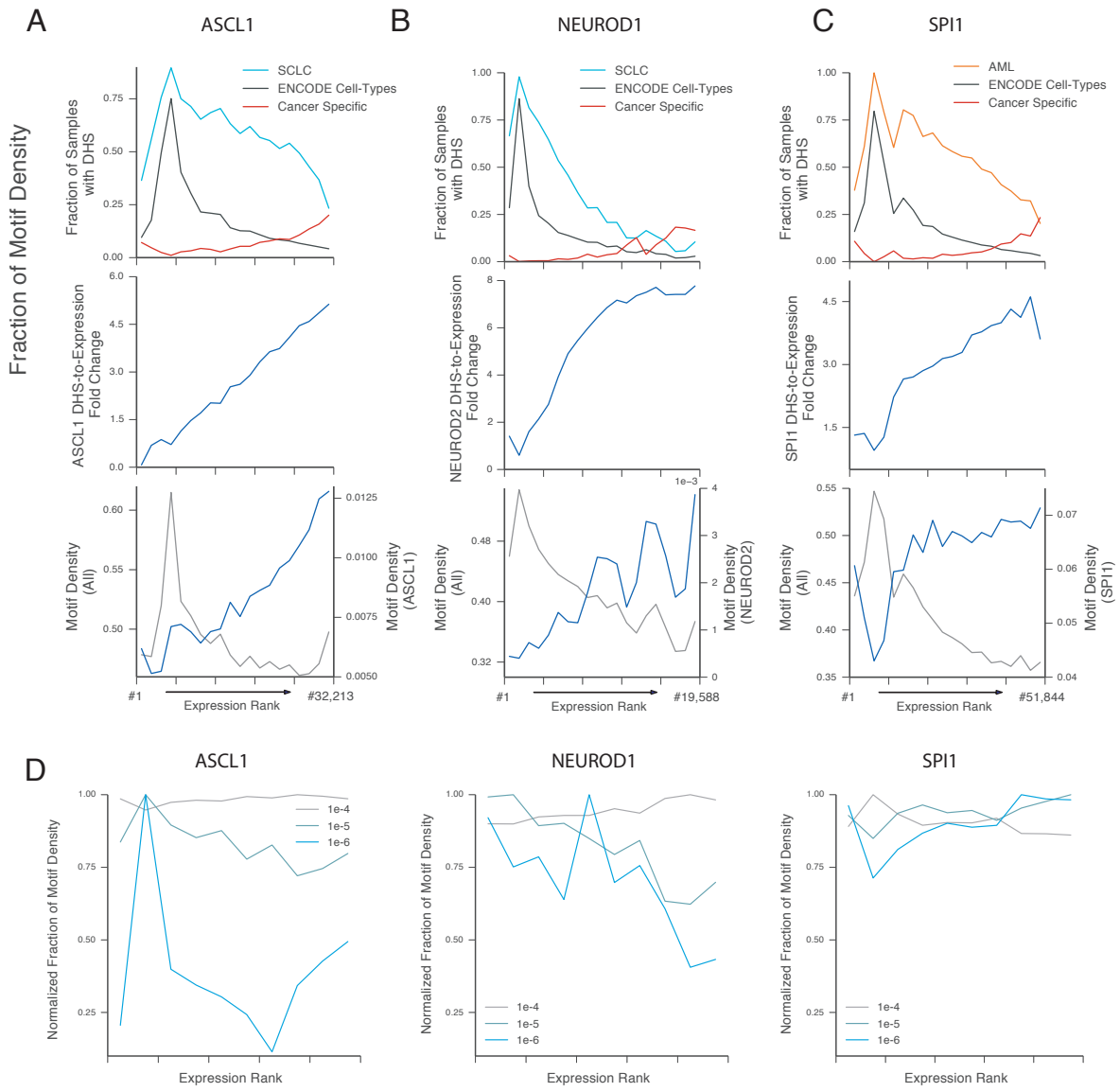
(e) As (d) for AML. Motifs of hematopoietic regulators such as SPI1 (PU.1), IRF8, and CEBPA are enriched in variable DHSs.



Supplement #4-DHSs activated at higher TF expression are more cell type specific

(a) Overlap of all (left, stacked barplot) linked DHSs and linked DHSs ranked by TF specificity (right, stacked lineplot) with related cell types for NKX2-1 and NEUROD2 in SCLC.

(b) As (a) for IRF8 and GATA1 in AML



Supplement #5-Plastic DHSs are sensitive to graded TF expression

(a) DHSs overlapping ASCL1 ChIP-Seq peaks in SCLC were ranked by the expression of ASCL1 in samples where the DHS is present (for the minority of DHSs negatively correlated with TF expression, the sign of the quantity was flipped). Top: The average cell type specificity of the DHS and frequency of the DHS in SCLC samples. Middle: Fold change of DNaseI cleavage across the range of ASCL1 expression across the quantiles of expression rank. Bottom: Fraction of the total number of base pairs by any motif that are covered by the motif of the linked TF across quantiles of the expression rank. DHSs active only at higher levels of ASCL1 expression are more cell type specific, show greater fold change in activity over the range of ASCL1 expression, and have more homotypic motif content.

(b) As (a) for NEUROD1

(c) As (a) for SPI1 (PU.1) in AML

(d) Left: The density of ASCL1 motifs at different motif pvalue thresholds across the quantiles of the expression rank. The density is plotted as a fraction of all ASCL1 motifs (any motif pvalue <=1e-4) in the DHS. To facilitate comparisons, each line was scaled to set its maximum value to one. Center: As left for NEUROD1. Right: As left for SPI1.

Chapter 3: Regulatory plasticity links epigenetic therapy to oncogenic and developmental programs

This chapter (along with Chapter #2) is adapted with minor modifications from a manuscript in preparation: JE Lazar, S Hansen, S Neph, J Hallow, H Wang, R Sandstrom, D Bates, J Nelson, A Johnson, H Mohammad, C Carpenter, R Kruger, W Meuleman, and JA Stamatoyannopoulos. *Regulatory plasticity links epigenetic therapy to oncogenic and developmental programs.*

Abstract

Alteration of chromatin machinery is a common feature in cancer development, but the connection between often non-specific regulatory factors and oncogenic phenotypes remains unclear. Here, to study the relationship between a ubiquitous chromatin remodeler and oncogenic regulatory programs, we comprehensively measure chromatin and expression changes after inhibiting the histone demethylase LSD1 in two distinct cancer types—small cell lung cancer (SCLC) and acute myeloid leukemia (AML)—that are susceptible to LSD1 inhibition. DHSs that appear sensitive to differences in transcription factor dosage are preferentially altered by LSD1 inhibition, leading to highly distinct chromatin responses despite common recruitment of LSD1 by the repressors GFI1 and GFI1b in both cancers. Although the response in each chromatin context is distinct, the bias towards plastic chromatin targets the effects of LSD1 inhibition towards the regulatory programs of diverse neuroendocrine (SCLC) or myeloid (AML) specific TFs active in a given sample. Overall, our results support a role for the inherent plasticity of developmentally variable regulatory elements in driving their variation over the course of cancer development and susceptibility to epigenetic perturbations.

3.1 Introduction

During cancer development, genetic and epigenetic alteration of factors that modify chromatin cause regulatory dysfunction and are associated with a wide range of oncogenic phenotypes(Baylin and Jones 2011). As a result, chromatin modifiers have become popular targets for drug development(Dawson and Kouzarides 2016; Helin and Dhanak 2013). However, identifying the link between the, often non-specific, mechanisms of chromatin remodelers and the cancer pathways they affect has proved challenging. This difficulty is in part due to the high context specificity of the phenotypes: chemical inhibition of ubiquitously expressed

chromatin modifiers affects different subsets of genes in different cell types; mutations in common regulatory factors often cause cancer in only a small number of cell types; and disruption of the same complex may have pro- or anti-tumorigenic effects depending on the cell type. Understanding how the cellular context modulates the effects of perturbing chromatin modifiers will therefore be important for rationally designing and applying epigenetic therapies.

The histone demethylase LSD1 (KDM1A) is associated with to malignant phenotypes in a diverse set of cancers ranging from carcinomas(Lim et al. 2010) to sarcomas(Sankar et al. 2014; Schildhaus et al. 2011) to leukemias(Harris et al. 2012). But due to LSD1's pleiotropic effects on gene expression, the link between LSD1 activity and cancer behavior can be murky. LSD1 acts to both repress and activate gene expression. As a component of the repressive CoREST(Y.-J. Shi et al. 2005) and NuRD(X. Wang et al. 2009) complexes, LSD1 demethylates H3K4me1/2, a histone mark associated with enhancer activity, and in activating complexes LSD1 removes the heterochromatin associated mark H3K9me1/2(Metzger et al. 2005; Laurent et al. 2015). At multiple stages during normal development, LSD1 is required for proper regulation of the system specific differentiation programs. In embryonic stem cells, LSD1 co-localizes with the core embryonic TFs and represses expression of developmental genes in undifferentiated stem cells(Adamo et al. 2011) and stem cell genes after induced differentiation(Warren A Whyte et al. 2012). During hematopoietic differentiation, LSD1 is essential for multiple steps with loss of LSD1 leading to failure to repress gene expression programs from earlier stages of differentiation(Kerenyi et al. 2013; Sprussel et al. 2012; Thambyrajah et al. 2016). Similarly, in cancers LSD1 activity has been linked to multiple oncogenic programs related to the cancer's cellular identity such as blocking differentiation in AML(Harris et al. 2012), maintaining stem cell properties in germ line tumors(J. Wang et al. 2011), and promoting migration and invasion in carcinomas and neural cancers(Lv et al. 2012). But while LSD1 overexpression is typically associated with aggressive cancer phenotypes, LSD1 may also play role as a tumor suppressor in certain cellular contexts: truncating mutations of LSD1 are associated with multiple myeloma and decreased LSD1 activity is associated with plasma cell proliferation and activation of myeloma associated transcriptional programs(Wei et al. 2018).

To understand how LSD1 regulates distinct pathways in varied cellular contexts, we studied the effects of LSD1 inhibition in two orthogonal cancer types that are sensitive to LSD1 inhibition: acute myeloid leukemia (AML)(Harris et al. 2012; Schenk et al. 2012) and small cell lung cancer (SCLC)(Mohammad et al. 2015). In both AML and SCLC, LSD1 is highly expressed and TFs that recruit LSD1 are important for the differentiation of the

putative cell types of origin(Karsunky et al. 2002; Kazanjian et al. 2004). Using DNaseI-Seq, we directly mapped the chromatin changes after LSD1 inhibition across a panel of SCLC and AML cell lines. The response to LSD1 inhibition is highly dependent on the initial chromatin state, resulting in a diverse set of responses between the different cancers and samples of the same cancer type. However, in each cellular context, the regulatory elements most variable among different samples of the same cancer and most sensitive to differences in the trans regulatory were preferentially affected. The diverse response at the plastic regulatory elements links the effects of LSD1 inhibition to developmental programs active in each cellular context and suggests a unifying role for the intrinsic plasticity of regulatory elements in linking development, oncogenesis, and chromatin remodeling.

3.2 Results

Targeted epigenetic drugs preferentially impact plastic elements

To understand the effect of inhibiting LSD1 in the context of the highly diverse chromatin states we observe in SCLC and AML, we treated 10 SCLC and 10 AML lines with GSK-LSD1, a specific LSD1 inhibitor with demonstrated anti-proliferative activity in AML samples and a subset of SCLC lines (Mohammad et al. 2015). For each cell line we performed DNaseI-Seq and RNA-Seq at two days and two drug doses (10nM and 500nM) plus a control. Additionally, for the 10 SCLC lines, we performed H3K4me1 and H3K4me2 ChIP-Seq at the same time points and conditions (Figure 3.1a). Since the effects of LSD1 were modest relative to the pre-existing variation in the cell lines (Supplementary Figure #1a), we identified DHSs responsive to LSD1 inhibition (LSD1i) independently in each cell line using DESeq2 (Love, Huber, and Anders 2014).

In most cell lines, we observed a clear separation between treated and untreated samples in the first two principle components for a single cell line (Supplementary Figures #1b) and identified thousands of significantly responsive DHSs (FDR <5%). The exceptions were four SCLC cell lines—which notably did not show growth arrest after LSD1 inhibition—that had few responsive DHSs. LSD1 inhibition causes primarily quantitative shifts in the pre-existing chromatin accessibility rather than de novo regulatory activity or complete repression of elements (Figure 3.1c-d). In SCLC, an average of 68% of DHSs (74% for AML) with increasing activity were called in the cell line prior to treatment with an additional 32% (24% for AML) called in a different untreated SCLC cell line (Supplementary Figure #1c). The differential DHSs occur at sites are marked with lower levels of H3K4me2 and higher levels of H3K2me1

than background, intensity matched DHSs, suggestive of weak or poised enhancers (Supplementary Figure #2a), and after LSD1 inhibition, we see a concordant increase/decrease in the strength of the H3K4me1/me2 marks at both increasing and decreasing DHSs (Supplementary #2b).

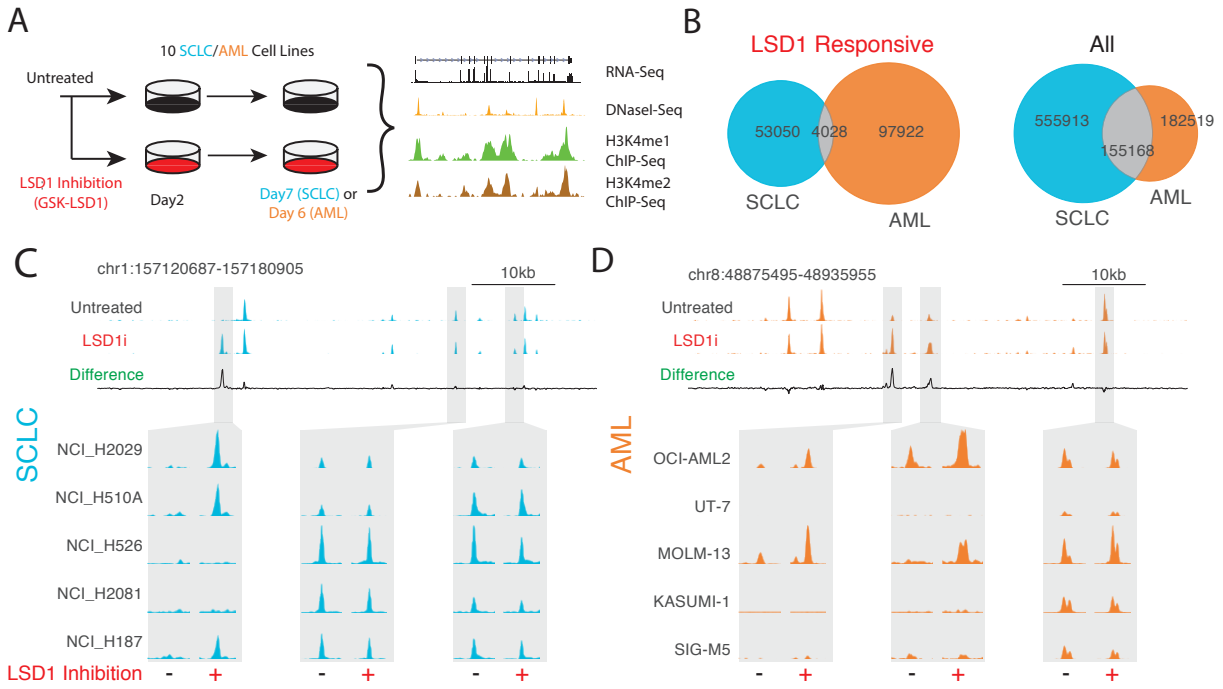


Figure 3.1: LSD1 inhibition affects pre-accessible sites

(a) Schematic of LSD1 inhibition experiments. 10 SCLC and 10 AML lines were treated with GSK-LSD1. For SCLC, DNaseI-Seq, RNA-Seq, H3K4me1 ChIP-Seq, and H3K4me2 ChIP-Seq were performed after two and seven days of treatment. For AML, DNaseI-Seq and RNA-Seq were collected after two and six days of treatment.

(b) Overlap of LSD1 responsive DHSs and all DHSs between AML and SCLC

(c) DNaseI cleavage density before and after treatment with GSK1-LSD1 around a representative LSD1 responsive DHS in SCLC. Left highlighted DHS is identified as responsive to GSK-LSD1.

(d) As (B) for AML. Left and middle DHSs are identified as responsive to GSK-LSD1.

To determine factors responsible the response of specific DHSs to LSD1 inhibition, we identified motifs enriched in the LSD1i responsive DHSs (Methods). In AML and a subset of SCLC samples, the sequences of DHSs with increasing accessibility were significantly enriched for the GFI1/GFI1b motif, and across samples the magnitude of the change in chromatin accessibility at the GFI1/1b motif correlates with the expression of GFI1 + GFI1b prior to treatment (3.2a). The transcriptional repressor GFI1 recruits LSD1 through its SNAG domain (Saleque et al. 2007) and is necessary for normal myeloid differentiation (Karsunky et al. 2002) and pulmonary neuroendocrine development (Kazanjian et al. 2004). The effects

of LSD1 in AML and SCLC have previously been linked to GFI1/GFI1b (Cusan et al. 2018; Ishikawa et al. 2017; Maiques-Diaz et al. 2018; Saleque et al. 2007; Takagi et al. 2017). Consistent with GFI1/1b mediating direct effects of LSD1 inhibition, we observe DHSs containing the GFI1 motif show greater changes after two days compared to other LSD1 responsive DHSs (Supplementary Figure #3a), and in the SCLC line NCI-H526, DHSs containing the GFI1 motif display elevated LSD1 ChIP-Seq signal (Supplementary Figure #3b).

Despite the shared role for GFI1/1b, we observe an almost non-overlapping response to LSD1 inhibition in SCLC and AML at all DHSs (Figure 3.1b) as well as DHSs containing the GFI1 motif. The divergent chromatin response between the two cancers cannot be explained only by the accessibility of the elements prior to treatment, as the LSD1i responsive DHSs are enriched for non-overlapping sites. Instead, we observe in each cancer, DHSs containing the GFI1 motif that are more tightly associated with the expression of the linked TF at the basal state are enriched in the LSD1 responsive DHSs (3.2b). However, the identity of the TFs between the two cancers is distinct: in SCLC the variable DHSs are linked to neuroendocrine TFs and the LSD1i responsive DHSs are most often correlated with the expression of neuroendocrine TFs, and analogously in AML the responsive DHSs are most often correlated with myeloid specific TFs (Figure 3.2c). Therefore, although GFI1 plays a central role in mediated to response to LSD1 inhibition in both SCLC and AML, the preferential response at DHSs whose activity is closely linked to TF expression leads to a response at divergent regulatory programs in the two cancer types.

Looking beyond GFI1, we observe the enrichment of DHSs whose accessibility is closely linked to TF expression holds across all LSD1i responsive DHSs (Supplementary Figure #4a). Notably, the enrichment holds when conditioning on DHSs that overlap with a single TF's ChIP-Seq peaks (Supplementary Figure #4b-c), suggesting the effect is not due solely to the identity of the linked TFs but reflects a quality of the DHS.

TF expression patterns pre-determine drug effects

Since expression of the TFs linked to the LSD1 responsive DHSs varies among SCLC and AML samples, we analyzed how this variation affects the response to LSD1 inhibition. Focusing first on DHSs containing the GFI1 motif, the response to LSD1 inhibition at pre-existing elements leads to a pattern of chromatin response among cell lines that mirrors the accessibility of the elements and the expression of the neuroendocrine (in SCLC) and hematopoietic (in AML) TFs linked to the activity of the DHSs prior to treatment (Figure

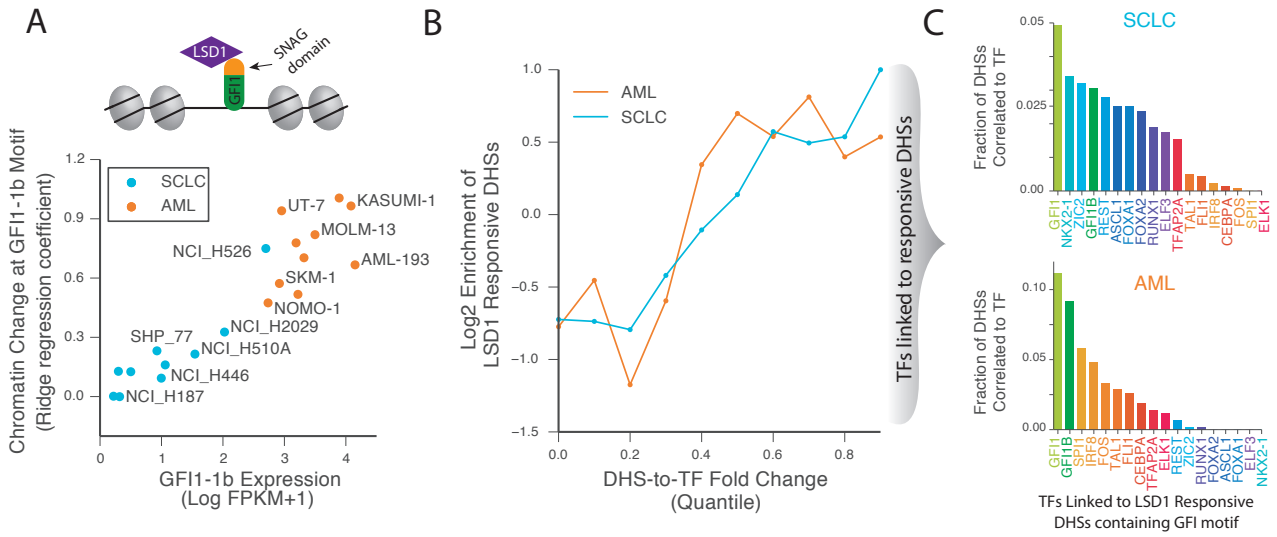


Figure 3.2: LSD1 inhibition affects labile sites

(a) The motif of the SNAG domain TF GFI1 is enriched in LSD1 responsive DHSs in all AML samples and a subset of SCLC samples. Scatterplot of the summed expression of the repressors GFI1 and GFI1B versus the chromatin change at the GFI motif (ridge regression coefficient) after LSD1 inhibition. Chromatin change at the motif (ridge regression coefficient) strongly correlates with expression of GFI1 + GFI1b prior to treatment.

(b) For DHSs containing the GFI1 motif, enrichment of LSD1 responsive DHSs compared to a peak height matched background across quantiles of the fold change of the DHS over the linked TF expression. DHSs from each cell line where the GFI1 motif is enriched in LSD1 responsive DHSs are combined.

(c) Barplot of the fraction of LSD1 responsive DHSs containing a GFI1/1b motif linked to TFs in SCLC and AML. The 10 TFs with the largest number of linked DHSs in SCLC and AML are displayed. Note the distinct set of TFs linked to LSD1 responsive DHSs between SCLC and AML.

3.3a-b). By establishing the basal chromatin state, lineage specifying factors may therefore act to modify the effects of TFs directly recruiting LSD1.

Genome-wide, variation in the chromatin and transcriptional response to LSD1 inhibition correlates with to the similarity in chromatin accessibility prior to treatment (Supplementary Figure #5a). This link between basal chromatin state and LSD1i response cannot be completely explained by the bias towards DHSs accessible prior to treatment. Instead differences in the genome-wide response at TF motifs enriched in LSD1 responsive DHSs explain additional variance in the response (Supplementary Figure #5b). As key lineage specifying TFs are linked to variation in the largest number of DHSs in SCLC and AML, the enriched motifs belong to diverse set of neuroendocrine (SCLC) and hematopoietic (AML) TFs (Figure 3.4a). The response at the regulatory programs of these TFs occurs specifically in the subset of samples where they are expressed prior to treatment leading to perturbation of each sample's main lineage specifying TFs' regulatory programs and a diverse response to

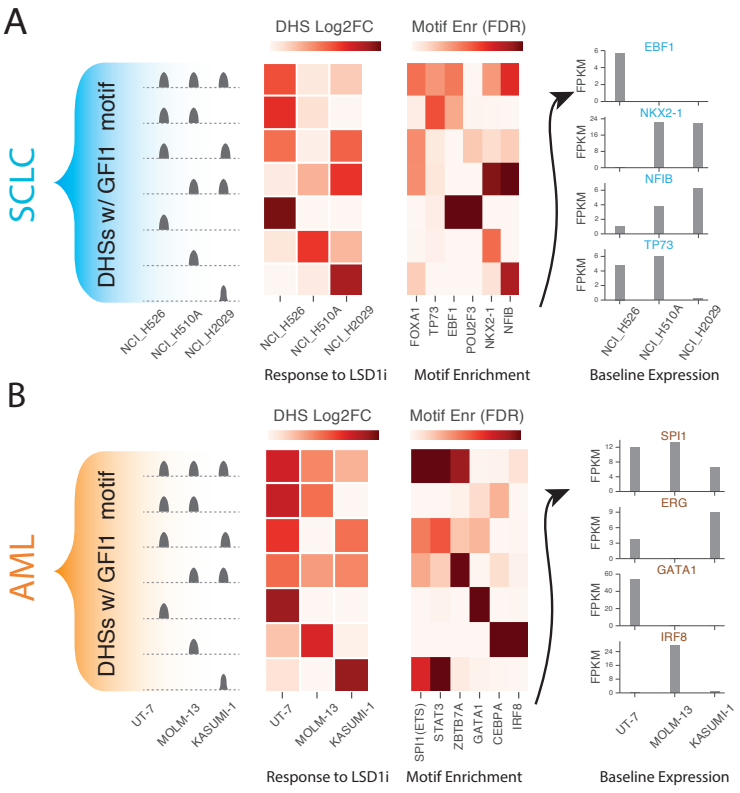


Figure 3.3: Lineage specifying TFs shape LSD1i response by establishing baseline accessibility

(a) DHSs containing the GFI1 motif were grouped based on pattern of activity in three SCLC lines prior to LSD1 inhibition. Response to LSD1 inhibition (LFC) in each of the DHS subsets shows preference for response at pre-existing accessible sites. The pre-existing variable elements are enriched for TF motifs related to the variable expression of lineage defining TFs and expression of those TFs matches the patterns of DHS activity.

(b) As (a) for AML

LSD1 inhibition both at the DHS and TF level (Figure 3.4b-c).

In SCLC only a subset of samples are sensitive to LSD1 inhibition. Given how the pre-existing chromatin landscape—and the TFs that establish it—shape the response to LSD1 inhibition, we hypothesized the TF context prior to treatment might explain differences in the response to LSD1 inhibition among SCLC lines. We screened the chromatin response to GSK-LSD1 in 37 SCLC cell lines (9 were sensitive to the treatment, 8 were indeterminate, and 20 were insensitive) with 10nM of GSK-LSD1 and a DMSO control and performed DNaseI-Seq after 4 days of treatment. DHSs with greater increase in the sensitive cell lines compared to insensitive lines are linked to genes involved in cell morphology, epithelial development and the cell cycle (Supplementary Figure #6a). To identify differentially af-

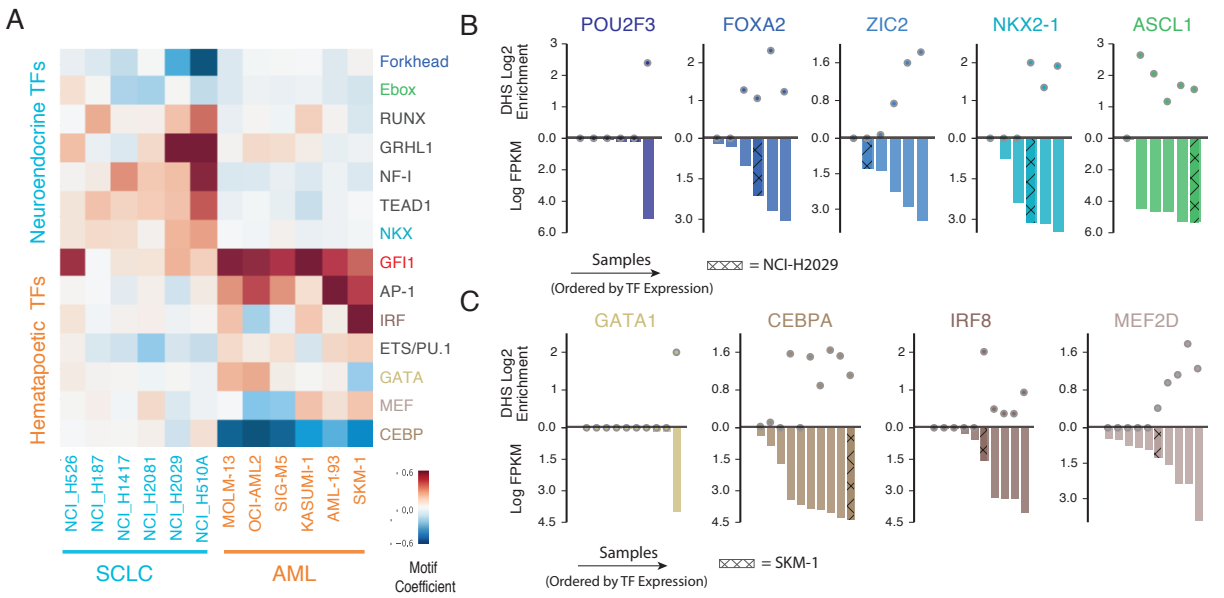


Figure 3.4: TF expression patterns pre-determine drug effects

(a) Heatmap of change in DNaseI accessibility at motifs (ridge regression coefficients) enriched in LSD1 responsive DHSs in SCLC and AML.

(b) Enrichment of LSD1 responsive DHSs in DHSs linked to TFs across SCLC lines (scatter, top) compared to the expression of the TF prior to treatment in SCLC (barplot, bottom). Cell lines were ordered by expression of the TF.

(c) As (b) for AML

fected trans factors, we compared motifs' ridge regression coefficients between the sensitive and insensitive samples. DHSs containing the motifs of SNAG domain TFs, GFI1/1b and SCRT1/1, showed consistent change in both sensitive and insensitive cell lines, suggesting disrupting LSD1 activity is not sufficient for the observed sensitivity in SCLC. However, the change in chromatin accessibility at the GRHL and NKX motifs was significantly greater in the sensitive samples than insensitive samples (permutation test, B-H FDR < 0.1 with optimal thresholding for coefficient magnitude (Love, Huber, and Anders 2014)) with the greatest difference at the GRHL motif (Figure 3.5a).

GRHL2 and NKX2-1 regulate normal lung development(Varma et al., 2012) and within SCLC expression of the two TFs is linked to the activity of fetal lung and lung epithelial specific DHSs . The differential response at these motifs is therefore associated with a greater increase in accessibility after LSD1 inhibition ($p < 0.05$, permutation test) at small airway epithelium and fetal lung specific DHSs in sensitive compared to insensitive cell lines (Figure 3.5b). The higher activity of epithelial regulatory elements may be associated with a less aggressive phenotype as we observe a trend towards longer progression free survival

(Figure 3.5e,p = 0.02, log rank test between top and bottom tertile) and overall survival (Supplementary Figure #6c, p = 0.12) in patients with higher expression of GRHL2. In line with the baseline chromatin state and underlying TFs pre-defining the response to LSD1 inhibition, the chromatin activity and expression of GRHL2 and NKX2-1 differs between sensitive and insensitive cells prior to treatment (Figure 3.5c-d, Supplementary Figure #6b), and so the TFs may be useful biomarkers for LSD1 treatment.

We observe highly sample specific responses to LSD1 inhibition that are determined by the sample's initial chromatin state. Therefore, the TFs identified as responsible for establishing a cell's chromatin context also define the response to a perturbation of a general regulatory factor and can be used to predict patterns of response.

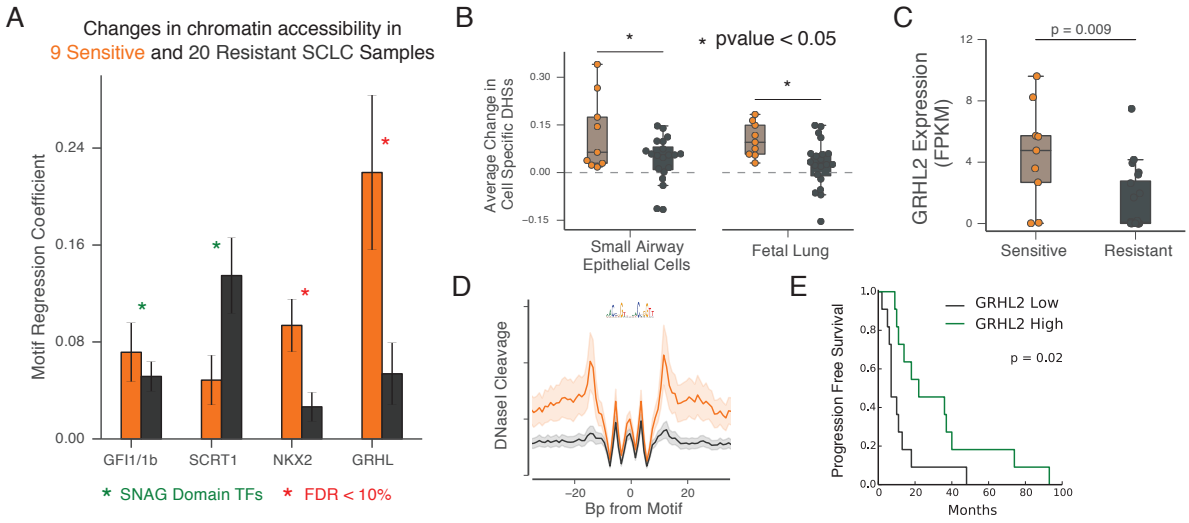


Figure 3.5: Expression of epithelial TFs correlates with LSD1 sensitivity

(a) Chromatin change around select motifs in sensitive and insensitive SCLC cell lines. Bar represents the average ridge regression coefficient for the motif, error bars equal +/- standard error of the mean. FDR from permutation test with BH correction on TFs thresholded by average absolute value of ridge regression coefficients.

(b) Boxplot of average change in small airway epithelium (SAEC) and fetal lung selective DHSs after LSD1 inhibition in sensitive and insensitive cell lines (p ¡ 0.05, permutation test)

(c) Boxplot of GRHL2 expression in sensitive and resistant cell lines. Pvalue from permutation test with Bonferroni correction for testing GRHL2 and NKX2-1.

(d) Kaplan-Meier curves for progression free survival for GHRL2 high and GRHL2 low patient samples. Pvalue from log rank test on top and bottom tertiles of GRHL2 expression.

Regulome plasticity predicts response to epigenetic perturbation

Despite the highly diverse set of responses to LSD1 inhibition both within and between SCLC and AML, there was a consistent relationship between the sets of DHSs and TFs

that varied between different instances of the cancer at baseline and the response to LSD1 inhibition. To directly test the relationship, we examined the enrichment of LSD1i responsive DHSs across the range of DHS frequency in untreated samples. In both SCLC and AML, we observe an almost monotonic relationship between the cell-to-cell variability of a regulatory element and its responsive to LSD1 inhibition (Figure 3.6a-b). The effect is consistent in both distal and proximal DHSs (+/- 500 bp of a TSS) considered independently (Supplementary Figure #7a-b).

The response at the variable DHSs links the effects of LSD1 inhibition to cell type and cancer specific DHSs. Just as cell type and cancer specific DHSs are progressively more variable between cancer instances, they are preferentially affected by LSD1 inhibition (Figure 3.6c). The changes in chromatin accessibility after LSD1 inhibition correlate with change in expression of the linked genes (Supplementary Figure #7c), and so consistent with the chromatin changes at cell type specific regulatory elements, we observe the genes that change expression after LSD1 inhibition are involved in diverse developmental processes in both AML and SCLC (Figure 3.6d).

Changes to a cell's epigenetic state during cancer development are reflected in DHSs that exist at low frequency when comparing chromatin accessibility across different samples of the same cancer. The set of low frequency regulatory elements in SCLC and AML that are altered over cancer development overlaps with the sets of DHSs that vary across normal development, that are affected by multiple distinct regulatory perturbations including variation in TF expression and LSD1 inhibition, and that vary at the single cell level within a single cancer cell line (Figure 3.6e). We propose these findings are best explained by an inherent plasticity of certain regulatory elements that is reflected by variation in response to a diverse set of perturbations.

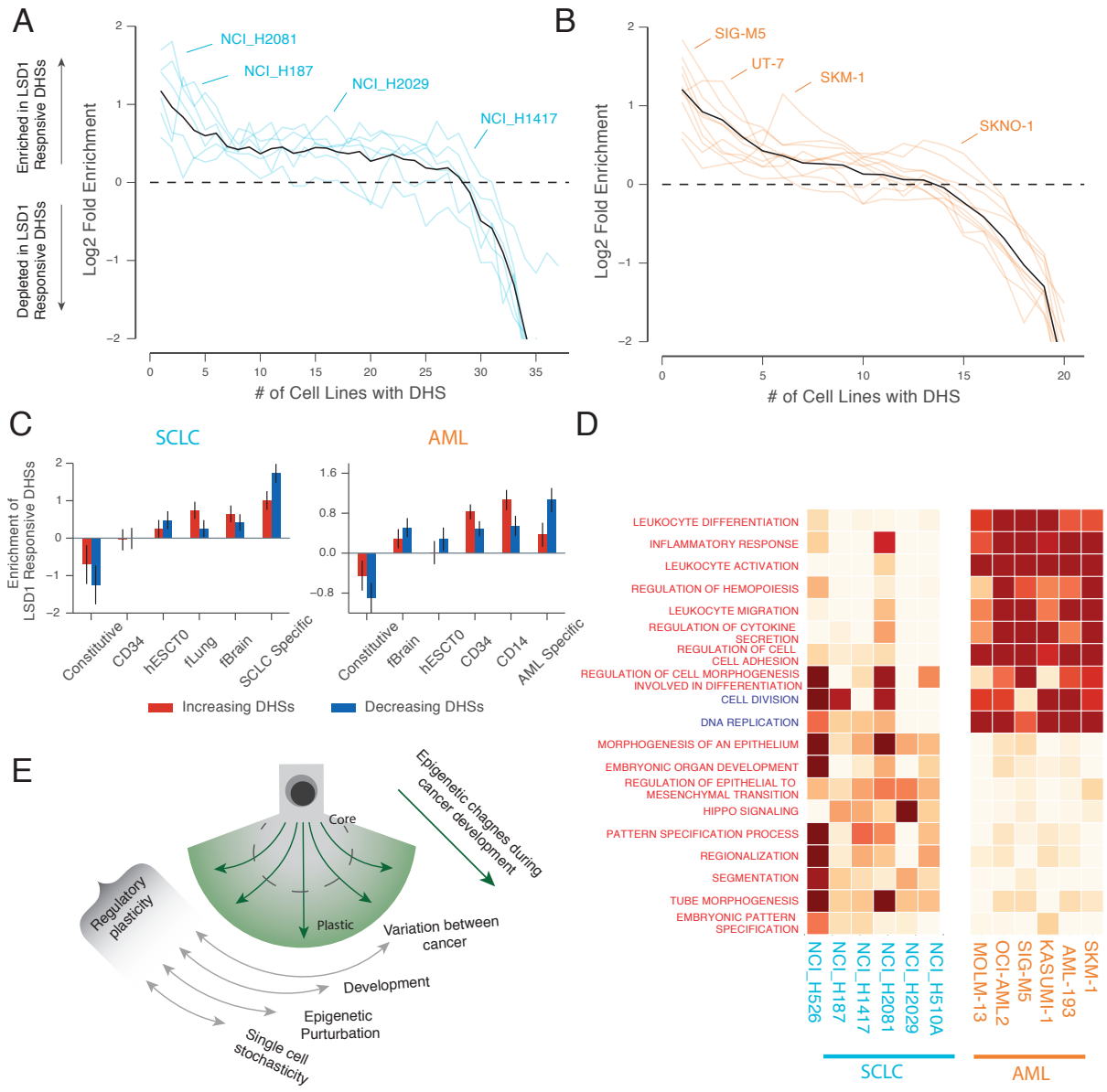


Figure 3.6: Regulome plasticity predicts response to epigenetic perturbation

(a) Enrichment of LSD1 responsive DHSs in SCLC at each DHS frequency. Teal lines show data from individual SCLC samples. Black line is the inverse variance weighted average across the lines.

(b) As (a) for AML.

(c) Enrichment of cell type and cancer specific DHSs in LSD1 responsive DHSs. Plotted are inverse variance weighted averages of all cell lines tested. Error bars +/- S.E.M

(d) GSEA analysis of RNA-Seq changes after LSD1 inhibition. Heatmap displays -log10(FDR) of selected GO terms in a subset AML and SCLC lines. Terms enriched in increasing genes are colored red and those enriched in decreasing genes are colored in blue.

(e) Model of LSD1 response. Multiple distinct regulatory perturbations, including LSD1 inhibition, affect a common set of regulatory elements. Intrinsic regulatory plasticity of certain classes of regulatory elements may explain the overlapping response observed across different types of epigenetic perturbations.

3.3 Discussion

Like many chromatin modifiers, LSD1 has been linked to a highly diverse set of oncogenic and developmental phenotypes, but the source of its target specificity remains murky. To study how LSD1 affects specific regulatory programs and how the observed heterogeneity in SCLC and AML influences the effects of a regulatory perturbation, we broadly mapped the effects of LSD1 inhibition on chromatin accessibility across multiple cancer cell lines. We find a stark contrast between the common set of TFs, most prominently the repressors GFI1 and GFI1b, that recruit LSD1 and the highly diverse set of accessible sites affected by LSD1 inhibition. The contrast between shared targeting mechanism and distinct regulatory responses can be understood considering the heterogeneous response of individual regulatory elements to changes in the trans regulatory environment. Within each sample, LSD1 inhibition affects a sample-specific set of plastic regulatory elements: the affected DHSs are more variable between different samples of the same cancer type, they show a greater response to differences in TF expression between samples, and their TF motif content is more homotypic for the bound TF's motif. We believe this observed specificity of LSD1 inhibition for plastic elements may explain several features of the response to epigenetic perturbations.

Developmental and oncogenic expression programs are often specifically affected by chromatin modifiers even without an obvious link between the chromatin modifier and the system of interest. For example, inhibition or knockdown of LSD1 affects stem cell or early developmental genes in ESCs, endothelial pathways in hemangioblasts, and genes associated with B-cell differentiation and the pathogenesis of multiple myeloma in plasma cells. We observe analogous behavior in AML and SCLC after LSD1 inhibition as cancer and lineage specific DHSs and gene expression programs are preferentially affected after LSD1 inhibition in SCLC and AML. Our results link this targeted behavior to an intrinsic instability of cell-type specific regulatory elements. Across SCLC and AML, cell-type specific DHSs from neural/epithelial and myeloid development respectively are highly variable and correlate most highly with expression levels of the cells' lineage specific TFs. LSD1 preferentially affects these variable DHSs and correspondingly the motifs of a highly diverse—and often sample specific—set of master regulators are enriched in the responsive DHSs. Although in some cases the master regulators might directly recruit LSD1, our model suggests that due to the lability of the cell-specific elements, chromatin modifiers will preferentially affect the regulatory elements of master regulators, and therefore gene expression of developmental programs, even in the absence of direct recruitment.

Inhibiting chromatin modifiers often shows increased activity in cancer cell lines despite

the targets being widely expressed in normal cell types. Cancer specific regulatory elements appear less buffered to changes in the trans regulatory environment, and in both AML and SCLC, we observe cancer specific DHSs are most likely to respond to LSD1 inhibition. The relative instability of cancer specific elements may therefore make them preferential targets for epigenetic modulation.

Finally, we observe the effects of LSD1 inhibition on chromatin are dominated by the effects due to the variation in expression of lineage specifying TFs. Rather than inducing a full reprogramming of the epigenome or causing context independent effects, LSD1 acts as an epigenetic rheostat, tuning the chromatin landscape established by a cell's master regulators. This suggests the therapeutic effects of LSD1 inhibition and other epigenetic therapies will be constrained by the epigenetic state in the cancer prior to treatment. In SCLC, where we directly compare those samples which are sensitive to LSD1 inhibition to those that are resistant, growth arrest after LSD1 inhibition correlates with the increased activity of DHSs from normal lung epithelium. However, these changes only occur in samples where those regulatory programs remain active, likely due to the continued expression of master regulators of lung epithelium, NKX2-1 and GRHL2. Response to epigenetic therapy may therefore rely on the residual activity of the normal cellular identity, and cancers with more highly dysregulated epigenomes could be less sensitive to epigenetic targeting. Therefore epigenetic therapies may be most effective if it is possible to target them in conjunction with context independent transcription factors, as is observed with the synergy between LSD1 inhibition and ATRA in inducing myeloid differentiation in AML. In an analogy with knockdown of chromatin modifying complexes improving the efficiency of cellular reprogramming, full trans-differentiation from the oncogenic state might require the combination of therapies targeting chromatin remodelers as well as therapies that can modify the activity of lineage specifying TFs.

3.4 Methods

Contact for Reagent and Resources Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, John Stamatoyannopoulos (jstam@altius.org)

Experimental Model and Subject Details

Cell Lines

SCLC and AML cell lines were purchased from ATCC, DSMZ, and Sigma Aldrich and cultured under the recommended conditions (Supplementary Table #1).

Method Details

DNaseI-Seq:

DNaseI-Seq was performed as previously described(Hesselberth et al., 2009; John et al., 2011; Thurman et al., 2012b). Briefly 5x10^5 -1x10^6 cells were lysed using 0.012%-0.05% IGEPAL with the concentration optimized for each cell line (Supplementary Table #1). Nuclei were collected by centrifugation at 500g for 5 min, and DNaseI digestion was performed for 3 min at 37C. DNaseI cleavage fragments were size selected by PEG fractionation, fragments were end repaired, and Illumina sequencing adapters were ligated. Libraries were sequenced to a typical depth of 50 million reads with a subset of samples sequenced to a higher depth ($\tilde{2}$00M) for DNaseI footprinting (Supplementary Table #2).

DNaseI-Seq Data Analysis:

Reads were aligned to GRCh38/hg38 using bwa (version 0.7.12) (H. Li et al. 2009) (Li and Durbin, 2009) and DHSs were identified using the hotspot-v2 algorithm (manuscript in preparation, https://github.com/Altius/hotspot2). DNaseI activity in master list elements was quantified by the number of cleavages falling within each element. Since we observed a non-linear relationship in peak intensities between replicates, DNaseI signal was normalized using a loess normalization scheme similar to the procedure used for microrrays (Yang et al. 2002). (Yang et al., 2002) When comparing between different cell lines, we normalized DNaseI signal using only the subset of peaks identified in all samples.

RNA-Seq:

Total RNA was isolated from 2.5 million cells using Thermo Fisher's MagMax mirVana Total RNA isolation kit, and libraries were prepared using Illumina's TruSeq Stranded Total RNA kit.

RNA-Seq Data Analysis:

Reads were aligned to hg38 using RNA-STAR (version 2.3.1) (Dobin et al. 2013), and counts per gene (Gencode v25 basic annotation) were quantified using HTSeq-count (Anders, Pyl, and Huber 2015). RNA counts were normalized for library size using DESeq's median ratio normalization and the normalized counts were converted to estimated FPKM values using the average library size.

H3K4me1/H3K4me2 ChIP-Seq:

Antibodies for H3K4me1 and H3K4me2 were purchased from Active Motif (Catalog Numbers AM39297 and AM39141). ChIP-Seq for H3K4me1/me2 was performed according to established protocols (Schmidt et al. 2009). (Schmidt et al., 2009) Briefly, 1x107 cells were cross linked in 1% formaldehyde for 10 minutes at room temperature and quenched with 0.125M glycine. Nuclei were collected, and chromatin was sheared using the Covaris S2 Focused-ultrasonicator. Chromatin immunoprecipitation was performed with 30uL of pre-cleared chromatin, 3uL antibody, and 15uL each of protein A/G agarose beads (Millipore). Libraries were constructed using Illumina's TruSeq DNA kit.

Chip-Seq Data Analysis:

Reads were aligned to GRCh38 using bwa (version 0.7.12). A similar approach as DNaseI-Seq was used to quantify H3K4me1/me2 ChIP-Seq data, except activity was quantified as the number of reads +/- 500 base pairs from the element to account for the more dispersed histone ChIP-Seq signal.

DNaseI Master List:

To create a representative list of DHSs elements across cancer samples, first DHSs were called (FDR <5%) independently in all individual samples (275 samples for SCLC and 157 samples for AML). At loci with extremely high copy number (for example around the MYC locus in a subset SCLC lines), the elevated background led to an increase in spurious peak calls. We therefore first filtered out peaks called only in samples where the average signal outside of hotspots in the region was >10x the genome wide average. All remaining peak calls were merged, and for each merged element a core peak was defined by the full width half maximum of DNaseI cleavage density in samples that contributed to the merged element. All peaks that overlapped (at least 50%) the master list elements were removed, and

the process was repeated until all DHSs were represented. Since combining many datasets may increase the fraction of spurious peaks, an additional filter was applied based on the combined z-score of each element. To choose an empirical threshold, logistic regression was performed on all pairs of replicates to identify the z-score at which 99% of the DHSs were expected to be replicated, and the average value of all sets of replicates was used as the final threshold. Master list elements were considered active in a given sample if a peaks called in the sample overlapped the elements by 25%.

Single Cell RNA-Seq:

Single cell RNA-Seq was performed on the SCLC line NCI-H446 using the 10x Genomics Chromium Platform with the Single Cell 3' v1 chemistry. The number of transcripts per cell were quantified using the associated Cell Ranger software (version 1.3.1). We identified an estimated 4,122 cells with a median of 2,692 genes detected per cell. To define genes with higher than typical cell to cell variability we first normalized the number of reads from each cell using median normalization, and then fit the relationship between gene CV and expression using robust GLM with gamma errors and inverse link (Brennecke et al. 2013) (Brennecke et al., 2013) For each gene we calculated a p-value of that the observed CV was greater than the trend by comparing to the gamma distribution estimated from the robust regression. We then used the Benjamini-Hochberg procedure to adjust for multiple comparisons and selected genes with FDR<10% as high variance. For each gene we calculated its chromatin variability as the average self- information of linked DHSs, considering only the DHSs that were active in NCI_H446 (the cell line assayed). Analysis was also performed excluding cell cycle genes (GO term GO:0007049), and the relationship between chromatin variability and single cell expression variation held (pvalue = 0.0002, Mann-Whitney U test).

LSD1 Treatment:

For time course experiments 10 SCLC cell lines (NCI-H1417, NCI-H187, NCI-H1930, NCI-H2029, NCI-H2081, NCI-H446, NCI-H510A, NCI-H524, NCI-H526, SHP-77) and 10 AML cell lines (AML-193, GF-D8, KASUMI-1, MOLM-13, NOMO-1, OCI-AML2, SIG-M5, SKM-1, SKNO-1, UT-7) were left untreated or treated with DMSO, 10nM of GSK-LSD1, or 500nM of GSK-LSD1. Cells were harvested for RNA-Seq, DNaseI-Seq, and ChIP-Seq (SCLC cell lines only) after 2 days and 6 days for AML and 2 days and 7 days for SCLC. For DNaseI-Seq, at least two independent time course experiments were performed for each cell line. For experiments screening the chromatin response in GSK-LSD1 sensitive and resistant

SCLC cell lines, the 37 SCLC lines were treated with 10nM GSK-LSD1 or a DMSO control and DNaseI-Seq was performed after 4 days.

LSD1 Time Course Analysis:

DNaseI cleavage counts in master list peaks, histone ChIP-Seq reads surrounding DHSs, and RNA-Seq gene counts were quantified as described above. For each feature (master list DHSs for DNaseI-Seq, gene for RNA-Seq) DESeq2 was used to fit the model $DHS - Time + Treatment + Time : Treatment + Batch$. Differential DHSs/genes were identified using the likelihood ratio test with a reduced model of $DHS = Time + Batch$. For quantification of the DHS/RNA fold change, the average contrast for 10nM and 500nM was used since the response at the two doses was similar.

DHS to Gene Links:

For SCLC and AML samples independently, we calculated the pearson correlation between log transformed gene expression and log transformed DHS activity for all DHSs within 75kb of each gene's TSS. All DHSs with correlation greater than 0.4 (which corresponds to 0.1 BH FDR in SCLC and 0.2 BH FDR in AML) as well as DHSs within 1kb of a gene's TSS were linked to the gene and used for downstream analysis.

DHS to TF Expression Fits:

To directly analyze the relationship between chromatin accessibility and TF expression at a given DHS, we fit the sigmoid function (Cleavage=$\frac{C_{max}}{1+e^{B_0+B_1*Expression}}$) using the expression of the TF with the highest correlation to DHS activity. All sigmoid fits were performed using scipy's least_squares function. We compared the sigmoid fits to linear fits and found across DHSs the average AIC was significantly lower for the sigmoid fits than linear fits ($p<0.01$, paired t-test). To summarize the fitted relationship between the DHS accessibility and TF expression, we calculated the fold-change between predicted accessibility at the observed minimum and maximum values of TF expression.

GO Term analysis:

To assess the relationship between regulatory plasticity and a gene's biological function, we ranked all protein coding genes by the average self-information (-1*Log(frequency)) of linked DHSs. GSEA version 2.2.3 (Mootha et al. 2003; Subramanian et al. 2005) was run on

the ranked genes using the classic scoring scheme against the biological process GO terms (msigdb version 5.2). Similarly, to analyze gene sets affected by LSD1 inhibition, we ranked protein coding genes by the standardized fold change ($\frac{LFC}{LFC\_SE}$) and performed GSEA on the ranked list.

TF Mutations: Mutation data for all SCLC cell lines except NCI-H2017 and NCI-H774 and all AML cell lines except GF-D8, UCSD-AML1, KASUMI-1, and OCI-M2 was available from the CCLE. To test if mutations in TFs affected the chromatin accessibility around that TF's motif, we compared a the full model Motif Density Log(TF Expression + 1) + Mutation + Mutation*Log(TF Expression + 1) to the reduced model Motif Density Log(TF Expression + 1), where Mutation is an indicator variable for the presence of any missense mutation in the TF. All TFs whose expression was significantly correlated with the TF motif density and contained ¿1 mutation were tested, and p-values were corrected using the Benjamini-Hochberg procedure.

Comparison to ENCODE samples: 665 high quality DNaseI-Seq samples produced by the ENCODE and NIH Roadmap Consortia were used as a set of reference samples (Supplementary Table #3). Master list elements were called as present in a sample if the overlap between an element and sample peak was >=25% of either the element or peak size. DHSs present in >95% of the 665 samples were classified as constitutive. When calculating the enrichment of sample specific DHS, we filtered out constitutive elements, and for cell types and tissues with more than one sample we used replicated peaks. For AML and SCLC, cancer specific elements were defined as elements that had an overlapping peak in <= 1 of the 665 samples.

TF ChIP-Seq Comparisons:

Peak calls for TF ChIP-Seq previously performed in the studied SCLC and AML cell lines were obtained from the original papers and GEO (Supplementary Table #7). Peaks were converted to hg38 coordinates using liftOver (Kent et al. 2002) and DHSs overlapping the TFs' ChIP-Seq peaks were used for downstream analysis.

Motif Analysis:

TF motifs were identified by scanning hg38 (FIMO (Bailey et al. 2009), pvalue < 1e-5)

with PWMs from the Jaspar (Portales-Casamar et al. 2010), Transfac(Matys et al. 2006), and Uniprobe(Newburger and Bulyk 2009) databases and a published SELEX dataset(Jolma et al., 2013). To identify motifs enriched in differential DHSs, the changing DHSs were compared (Fisher's exact test) to a background set of DHSs with LFC < 0.5 and matching initial peak intensities to account for possible confounding effects of ascertainment bias. Pvalues were corrected for multiple testing using the Benjamini-Hochberg procedure.

For analysis where a quantitative measure of change in accessibility around a motif was necessary, we performed ridge regression (sklearn version 0.18.1) on the standardize DHS LFC (LFC / LFC SE) with presence of the motif in the DHS as categorical variables. Since the motifs of TFs in the same families are highly similar, in order of have an uncorrelated set of features for the regression, we used groups of motifs that were clustered base as motifs based on the motif PWM (Matthew T Maurano et al. 2015). The ridge penalty was set by cross validation.

Quantification and Statistical Analysis

Identification of LSD1 responsive DHSs: Significance of changes in chromatin accessibility was determined by the likelihood ratio test in the R package DESeq2.

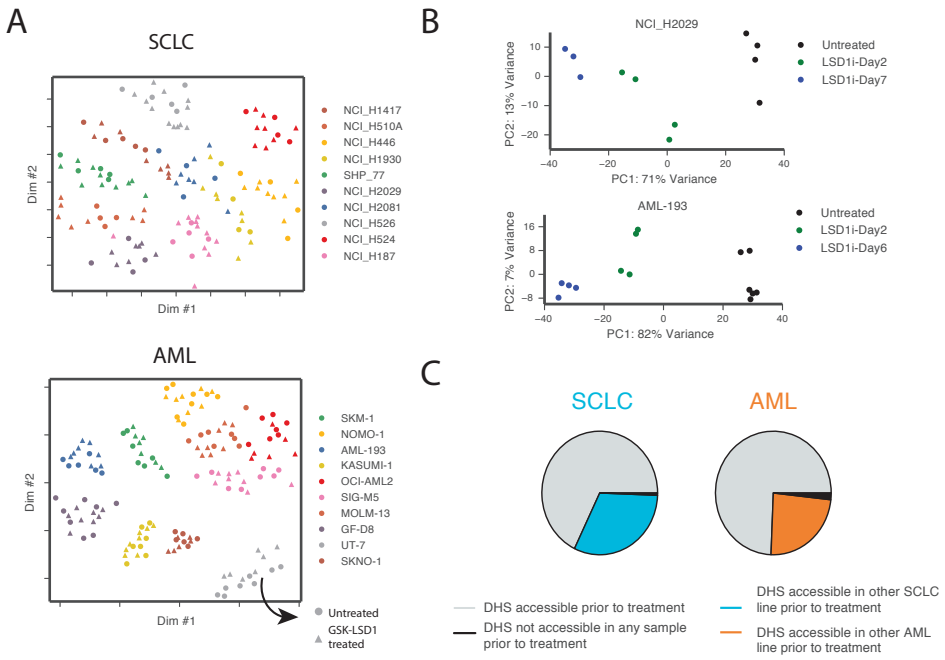
Enrichment between LSD1 responsive and background DHS sets: All pvalues for feature enrichments in LSD1 responsive DHSs were calculated by Fischer exact test (python scipy package) comparing LSD1 responsive DHSs to a peak height matched background set. Sensitive vs Resistant SCLC Comparisons: Statistical significance for comparisons between GSK-LSD1 sensitive and resistant SCLC cell lines was determined by p-values generated by permutation tests.

SCLC Survival Data: Expression and patient survival data was taken from George et al. (George et al. 2015). The top and bottom 33% of samples based on GRHL2 expression were selected as GRHL2 high/GHRL2 low sets. Kaplan-Meier curve and log rank test were calculated using the R package survival.

Data and Software Availability

Data from this study is available at NCBI Gene Expression Omnibus ($GEO; https : //www.ncbi.nlm.nih.gov/geo$), accession number GSEXXX.

3.5 Supplement

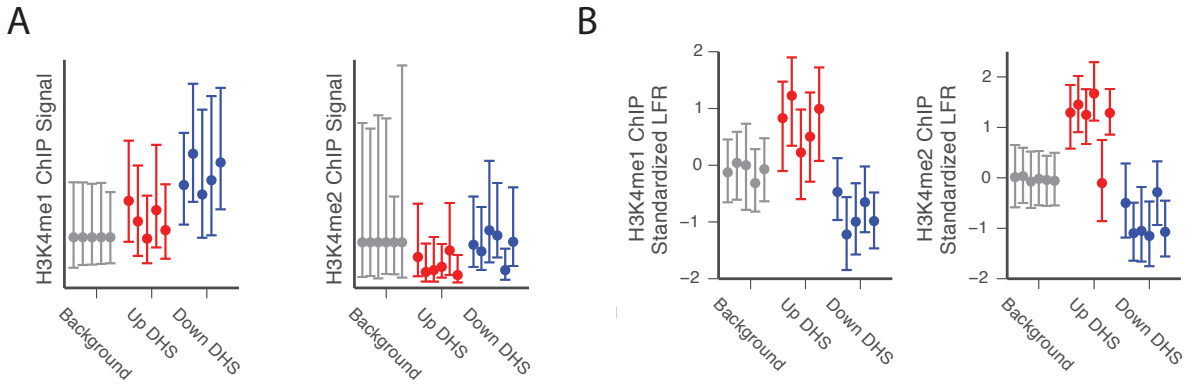


Supplement #1-LSD1 inhibition tunes pre-existing chromatin state

(a) Reduced dimensionality representation (MDS) of all DNaseI signal from 10 SCLC and 10 AML lines combining both treated and untreated samples. Untreated samples are plotted as dots and GSK1-LSD1 treated samples are marked with triangles.

(b) Principal component analysis of DNaseI cleavage intensity in the 2000 most variable DHSs in a representative SCLC (NCI-H2029) and AML (AML-193) cell lines. Note the clear separation between treated and untreated samples within a single cell line.

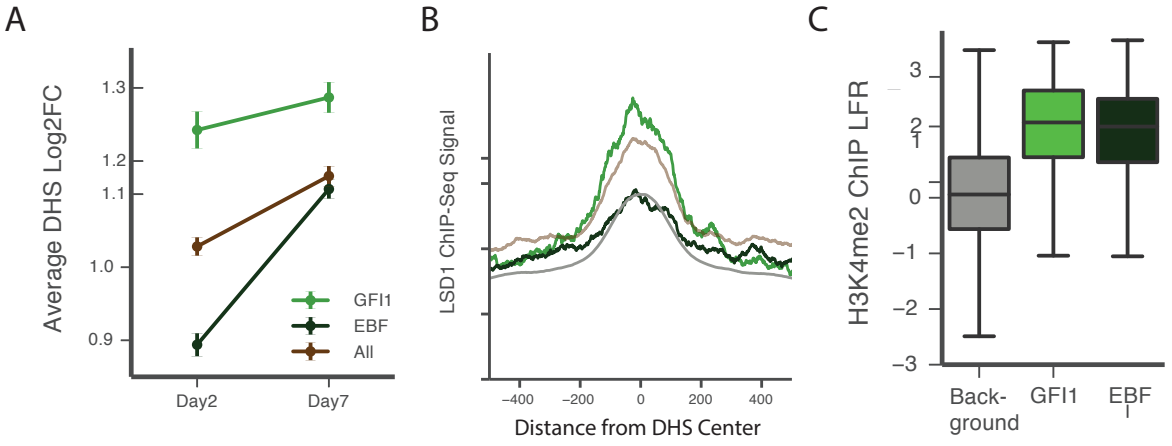
(c) Pie chart showing the fraction of increasing LSD1 responsive DHSs called in the sample prior to treatment, not called in the sample but called in another untreated sample of the same cancer, and not called in any untreated samples. Values are averaged over the 6 SCLC and 10 AML with appreciable chromatin changes after LSD1 inhibition.



Supplement #2-Histone modifications at LSD1 sensitive DHSs

(a) H3K4me1 and H3K4me2 signal prior to treatment at LSD1 responsive DHSs in the 6 SCLC cell lines with appreciable chromatin changes after treatment.

(b) Change in H3K4me1 and H3K4me2 signal at LSD1 responsive DHSs in 6 SCLC cell lines.

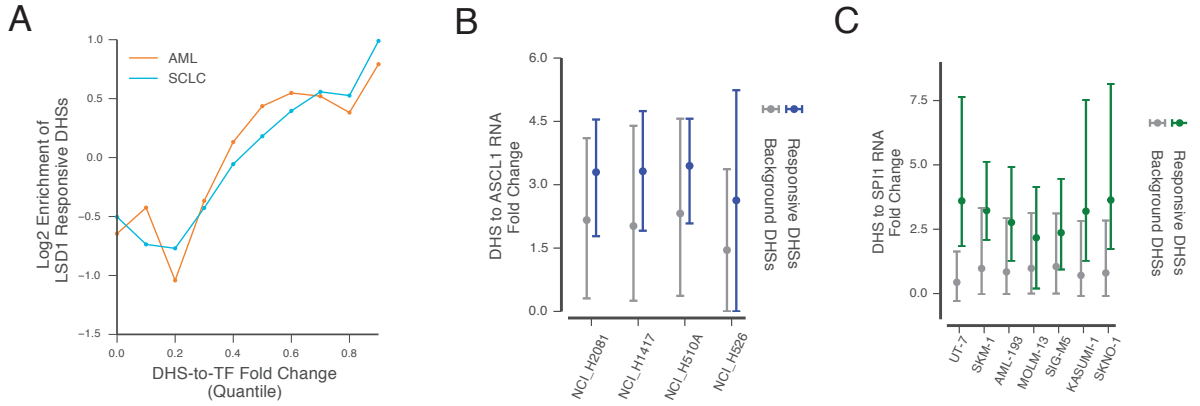


Supplement #3-DHSs with GFI motif reflect direct effects of LSD1 inhbition

(a) DHS change at day 2 vs day 7 in LSD1 responsive DHSs for the cell line NCI-H526. Consistent with direct recruitment of LSD1, DHSs with the GFI1 motif have a higher change at the early time point. DHSs containing the EBF1 motif are included as a comparison to a likely indirect effect.

(b) LSD1 ChIP-Seq signal around DHSs with the GFI1 motifs is greater than the typical responsive

(c) Change in H3K4me2 signal occurs around LSD1 responsive DHSs likely due to both direct and indirect effects of LSD1 inhibition.

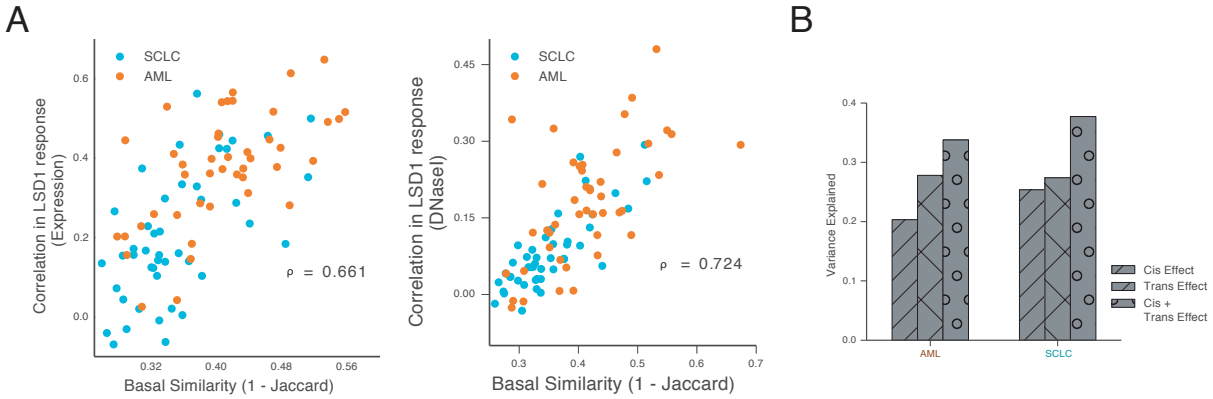


Supplement #4-DHSs sensitive to TF expression are preferentiallly affected by LSD1 inhibition

(a) Fold change of the DHS over the linked TF expression in LSD1 responsive DHSs and DHSs from a peak height matched background.

(b) Correlation between ASCL1 expression and DHS activity in LSD1 responsive and background DHSs that overlap an ASCL1 ChIP-Seq peak. Median is plotted with 25-75th percentile represented as error bars. The three cell lines shown are the SCLC lines with the ASCl1 motif enriched in LSD1 responsive DHSs.

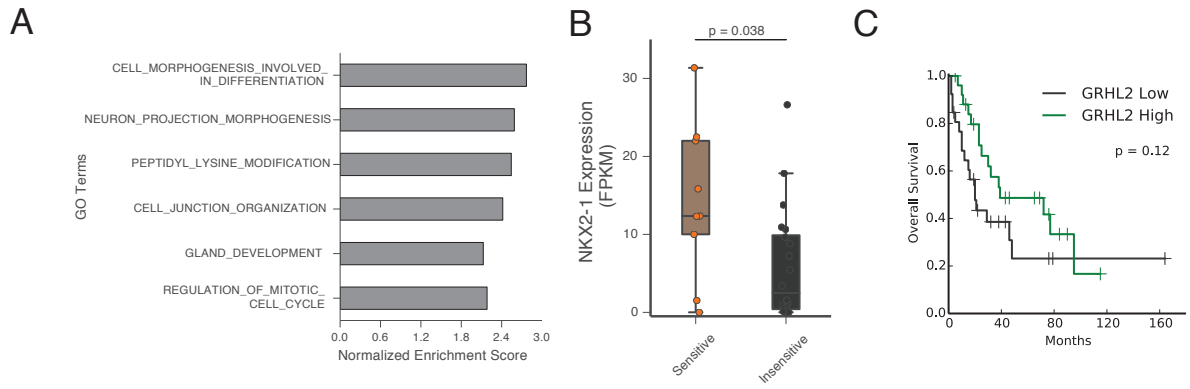
(c) As (l) for SPI1 (PU.1) in AML



Supplement #5-Pre-existing chromatin environment shapes LSD1i response

(a) Pairwise correlation in gene expression (right) and DNaseI response (left) between samples after LSD1 inhibition compared to similarity in chromatin landscape (Jaccard index) prior to treatment.

(b) Barplots of the fraction of variance in the LSD1 response between different SCLC/AML cell lines explained by three different models: 1. Accessibility of the DHS prior to treatment 2. Genome-wide changes in chromatin accessibility at motifs contained in the DHS. 3. The interaction of model #1 and model #2 (i.e. the genome wide changes in chromatin accessibility at a motif only affect pre-accessible sites in a given cell line). Plotted is the variance explained averaged over all LSD1 responsive DHSs.

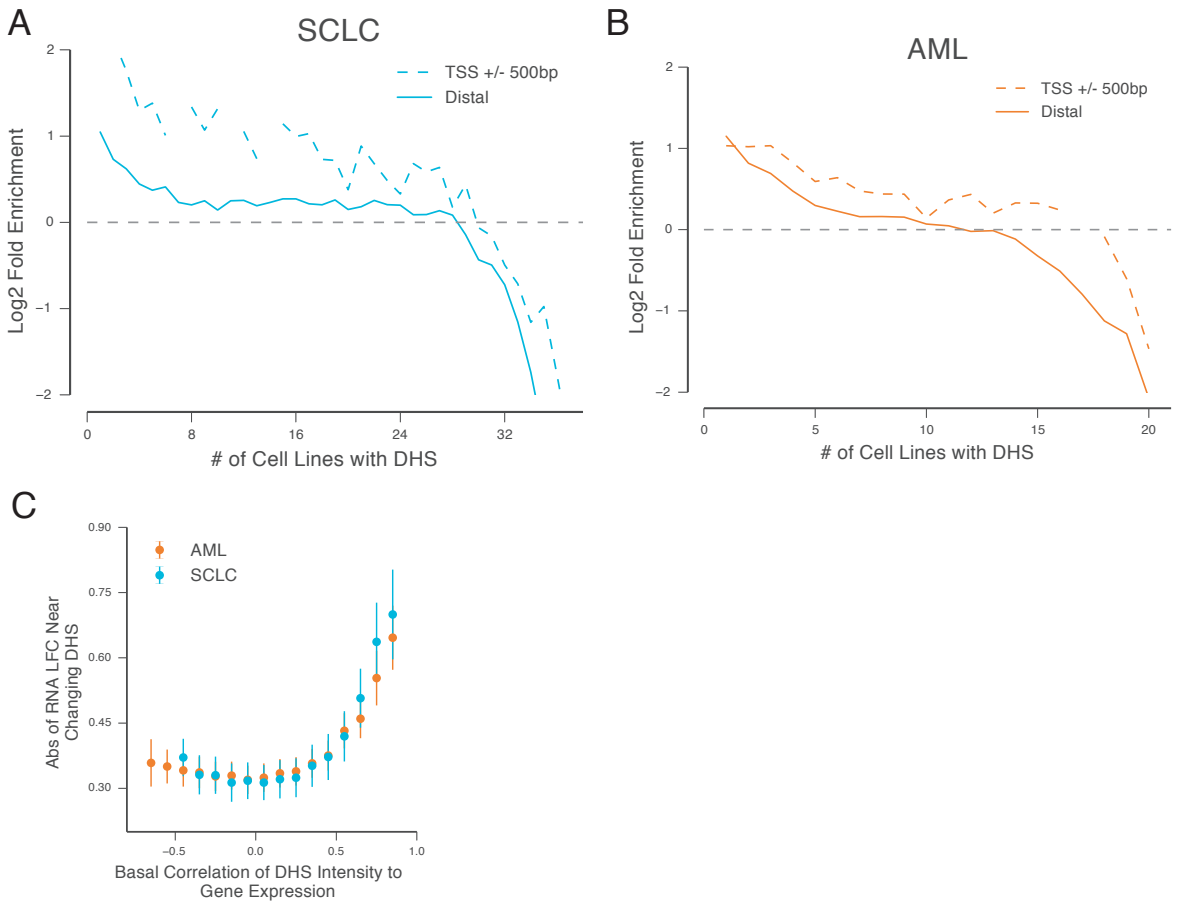


Supplement #6-Changes in epithelial programs distinguish sensitive from insensitive cell lines

(a) GSEA of biological process GO terms. Genes were ordered by the difference in the change in DNaseI cleavage between sensitive cell lines and insensitive cell lines.

(b) Boxplot of NKX2-1 expression in sensitive and insensitive cell lines. Pvalue calculated using permutation test with Bonferroni correction for testing GRHL2 and NKX2-1.

(c) Kaplan-Meier for overall survival curve for GHRL2 high and GRHL2 low patient samples. Pvalue from log rank test on top and bottom tertiles of GRHL2 expression.



Supplement #7-Changes in epithelial programs distinguish sensitive from insensitive cell lines

(a) Enrichment of LSD1 responsive DHSs in SCLC at each DHS frequency separated by distal and proximal DHSs.

(b) As (a) for AML

(c) Gene expression response near significantly changing DHSs. Genes were binned by the correlation between DHS accessibility and gene expression across the basal AML and SCLC lines. Error bars (+/-) S.E.M across cell lines.

Chapter 4: Locus specific effects of global unmasking of low-occupancy regulatory elements by SMARCA4

This chapter is adapted with minor modifications from a manuscript in preparation: JE Lazar, S Sterling-Sun, V Nandakumar, H Wang, F Urnov, A Funnell, and J Stamatoyannopoulos. *Promoter state and compartmentalized chromatin remodeling govern specific transcriptional responses to SMARCA4.*

Abstract

SMARCA4 is a central ATPase subunit of the human BAF/PBAF chromatin remodeling complexes. Mutations affecting SMARCA4, and associated BAF/PBAF subunits, occur frequently in a diverse range of cancers. In many cases, the oncogenic effects of these mutations have been linked to the de-regulation of specific target genes. However, there is an incomplete mechanistic understanding as to how aberrant, genome-wide chromatin remodeling results in specific transcriptional perturbations. To address this question, we used genome engineering to correct a homozygous mutation in *SMARCA4* in the well-characterized lung adenocarcinoma A549 cell line and profiled changes in chromatin accessibility, histone marks and transcription. Restoration of SMARCA4 causes a dramatic increase in chromatin accessibility and creates 23,861 novel DNaseI hypersensitive sites (DHSs), a 28% increase throughout the genome. The emergent DHS arise at sites that are pre-marked by transcription factors known to recruit BAF/PBAF and enhancer-associated histone modifications. Despite the widespread increase in chromatin accessibility, we observe comparatively attenuated changes in gene expression. Although there is a marked correlation between the number of local activated DHSs and the transcriptional responsivity of a gene, the influence of distal DHSs appears gated by a gene's promoter state. Specifically, promoters harboring bivalent histone marks (H3K4me3/H3K27me3) and low levels of DNA methylation are most prone to transcriptional upregulation. The largest changes in expression occur for genes in domains containing clusters of SMARCA4 binding sites that undergo a region-wide increase in accessibility. The regional changes align with topologically associated domains and occur at a subset of bivalent domains with matched active and repressive histone marks. Our results indicate that SMARCA4 elicits transcriptional changes through domain level increases in accessibility and reveal that a complex relationship between genome architecture, promoter context and cell-type specific enhancers shapes the effects of BAF/PBAF complex perturbation.

4.1 Introduction

Eukaryotic gene regulation requires the concerted action of sequence specific transcription factors (TFs) and chromatin modifying complexes to reorganize nucleosome bound DNA to permit the basal transcriptional machinery to access DNA (Kadonaga 1998). The SWI/SNF family is a set of closely related chromatin remodeling complexes that use ATP to remodel nucleosomes along DNA (Narlikar, Sundaramoorthy, and Owen-Hughes 2013) and thereby activate promoter and enhancer elements in diverse regulatory contexts (Bao et al. 2015; Bossen et al. 2015; H. C. Hodges et al. 2018; John, Sabo, Johnson, et al. 2008).

Human orthologues of SWI/SNF—which include the closely related BAF (BRG-/BRM-associated factor) and PBAF (Polybromo-associated BAF) complexes (C. Hodges, Kirkland, and Gerald R Crabtree 2016)—are frequently mutated in cancers, implicating these complexes in oncogenesis (Kadoch and Gerald R. Crabtree 2013). However, understanding the role these mutations play in cancer development is challenging. BAF/PBAF regulates diverse sets of target genes across different developmental stages and cancer types, and as a result, modifying BAF activity can have highly heterogeneous effects (C. Hodges, Kirkland, and Gerald R Crabtree 2016). In cancers, BAF subunits commonly act as tumor suppressors, yet these complexes also activate oncogenic regulatory programs in certain circumstances. In fact, the same BAF/PBAF subunit may act as a tumor suppressor or oncogene at different stages of cancer progression (Glaros et al. 2008; Roy et al. 2015; Sun et al. 2017). In these scenarios, such heterogeneous behavior has been linked to distinct sets of target genes in different cellular contexts. Therefore, understanding how BAF/BPAF's target genes are specified is necessary to interpret its diverse behavior.

Context specific BAF activity is caused in part by selective recruitment of the complex to regulatory elements either through cell type specific BAF/PBAF subunits (Ho and Gerald R Crabtree 2010; Kadoch and Gerald R. Crabtree 2013) or by cell type specific TFs (Boulay et al. 2017; Kadam et al. 2000; Vierbuchen et al. 2017). However, in eukaryotes, the relationship between recruitment of regulatory factors and target gene expression is not straightforward. In humans, genome-wide assays identify thousands of binding sites for most TFs and chromatin remodelers (Consortium et al. 2012; Gerstein et al. 2012); yet these factors affect the expression of more restricted sets of target genes (Lin et al. 2009; Ramagopalan et al. 2010; Reddy et al. 2009). For the BAF/PBAF complexes, in most systems only a subset of genes are dependent on their activity despite widespread binding (Hu et al. 2011; McBride et al. 2018; Raab, Resnick, and Magnuson 2016).

Various models have been proposed to explain the discrepancy between factor recruitment

and effects on expression including small effects for low-affinity binding (Fisher et al. 2012; Tanay 2006), synergistic relationships between clustered binding sites (Courey et al. 1989; Q. Li et al. 2002; Parker et al. 2013; Warren A. Whyte et al. 2013), and enhancer-promoter specificity (Butler and Kadonaga 2001; X. Li and Noll 1994; Zabidi et al. 2014). Experiments using inducible TFs, such as the glucocorticoid and hormone receptors, suggest binding site strength (Vockley et al. 2016), clustering codependent TF binding sites (Hakim et al. 2011; Reddy et al. 2009; Vockley et al. 2016), and activation of chromatin domains (Le Dily, Baù, et al. 2014; Le Dily, Vidal, et al. 2019) all play a role in defining the link between TF binding sites and the gene regulation.

To investigate the relationship between BAF/PBAF recruitment, resulting chromatin remodeling, and specific regulation of oncogenic programs, we reactivated SMARCA4 (BRG1), in A549s, a well-characterized, SMARCA4-null lung adenocarcinoma cell line. SMARCA4 is one of two mutually exclusive ATPase subunits of BAF and the sole ATPase subunit of PBAF. *SMARCA4* is frequently mutated across cancers, including ~5-10% of lung adenocarcinoma samples (Medina, Romero, et al. 2008; Network, Collisson, et al. 2014; Rodriguez-Nieto et al. 2011), and its loss leads to changes in cellular morphology and increased tumorigenicity in lung adenocarcinoma models (Medina, Carretero, et al. 2005; Orvis et al. 2014). SMARCA4 rescue causes a global increase in chromatin accessibility, but highly specific alterations to gene expression. The association between distal element activation and expression is influenced by regional chromatin architecture and the promoter state of the gene, demonstrating how a gene's genomic context can lead to heterogeneous interpretations of widespread chromatin reorganization.

4.2 Results

Efficient repair of the SMARCA4 locus

SMARCA4 is recurrently mutated in non-small-cell lung cancers (Medina, Romero, et al. 2008; Network, Collisson, et al. 2014; Rodriguez-Nieto et al. 2011), and thus we sought to investigate its function using a physiologically pertinent model system. The lung adenocarcinoma cell line A549 is homozygous for an exonic 23 bp deletion in *SMARCA4* that introduces a premature stop codon (Supplementary Figure #1). Since developmental roles of SMARCA4 are sensitive to gene dosage (Bultman et al. 2000),(Bultman et al., 2000) non-physiological levels of expression—whether introduced through transient transfection or stable insertion at a non-native locus—may not fully recapitulate key regulatory interactions.

We therefore opted to repair the 23 bp deletion to express SMARCA4 in its native regulatory context. We generated and screened a series of Transcription Activator-Like Effector Nucleases (TALENs) targeted near the mutation to identify a pair that caused efficient homology directed repair (HDR) when co-transfected with a single-stranded donor DNA harboring the wild-type sequence (Figure 4.1, Supplementary Figures #1-2). We used the lead TALENs to induce HDR, sorted single cells, and screened the resulting clonal lines by genotyping. We derived multiple clones with scarless correction of both *SMARCA4* alleles (SMARCA4$^{+/+}$) and arbitrarily selected three for subsequent phenotypic and regulatory profiling. For isogenic, matched controls, we selected two lines that were subjected to the same transfection and sorting process, but which remained unedited at *SMARCA4* (SMARCA4$^{-/-}$). Genotypes of the selected clones were confirmed by Sanger sequencing (Supplementary Figure #3).

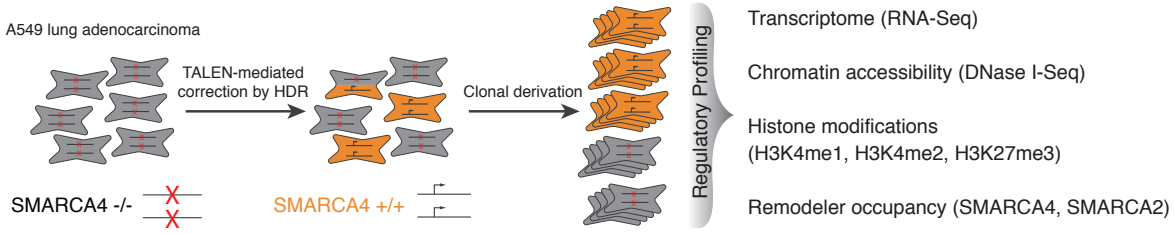


Figure 4.1: Efficient Repair of the SMARCA4 Locus
Schematic of SMARCA4 rescue experiments in A549 cells. In A549, a 23 bp deletion introduces an early stop codon at Q729. TALENs flanking the 23 bp deletion were designed to restore the deletion using a single strand oligonucleotide with homology arms flanking the deletion site. Cells were edited, sorted into single cells, and clones with homozygous WT SMARCA4 were identified.

Having established that the three SMARCA4$^{+/+}$ clones were genetically corrected, we next sought to confirm that SMARCA4 expression had been restored. In all three clones, the repair resulted in marked up-regulation of full length *SMARCA4* mRNA expression to levels comparable to normal lung tissue (Figure 4.2). We also confirmed that SMARCA4 protein is successfully expressed in these clones by Western blot (Figure 4.3a) and exhibits correct nuclear localization by immunofluorescence (Figure 4.3b-c). Phenotypically, SMARCA4$^{+/+}$ clones have a decreased growth rate (Figure 4.4) in line with SMARCA4's putative role as a tumor suppressor in lung adenocarcinoma and the involvement of the BAF/PBAF complexes in controlling the cell cycle (Nagl et al. 2005; Zhang et al. 2000).

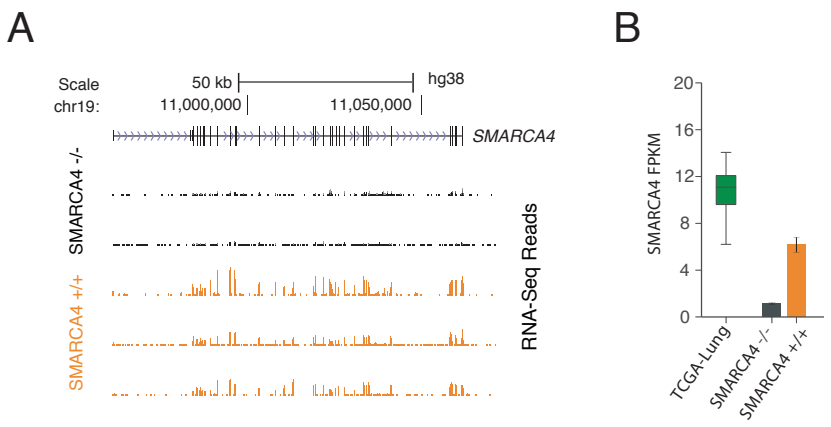


Figure 4.2: Reactivation of SMARCA4 expression

(a) Genome browser profile showing restoration of *SMARCA4* RNA expression in SMARCA4$^{+/+}$ clones.

(b) Comparison of restored SMARCA4 expression to normal lung tissues. Barplot (error bars +/- S.E.M) of SMARCA4 expression in SMARCA4$^{+/+}$ and SMARCA4$^{-/-}$ cells compared to SMARCA4 expression in normal lung tissue from TCGA (boxplot).

SMARCA4 reactivation causes a widespread increase in chromatin accessibility at pre-marked sites

To profile the effects of SMARCA4 reactivation on chromatin, we assayed chromatin accessibility by DNaseI-Seq and measured the histone marks H3K4me1, H3K4me2, and H3K27me3 by CUT&RUN (Skene and S. Henikoff 2017) in the SMARCA4$^{+/+}$ and SMARCA4$^{-/-}$ clones. In addition, we conducted CUT&RUN to compare the genomic occupancy of SMARCA4 and SMARCA2, the two interchangeable ATPase subunits of the BAF complex (Figure 4.1, 4.5a). Data from different single cell clones are highly reproducible, and we observe a clear separation in all data types between SMARCA4$^{-/-}$ and SMARCA4$^{+/+}$ samples (Supplementary Figure #4). Since chromatin accessibility provides a focal measurement of multiple classes of regulatory activity (Gross and Garrard 1988), we centered our analysis of regulatory changes around DNaseI peaks (DHSs). Consistent with BAF's effects on chromatin in other systems (Bao et al. 2015; Bossen et al. 2015; H. C. Hodges et al. 2018; Kelso et al. 2017), we observe a widespread increase in chromatin accessibility in SMARCA4$^{+/+}$ samples. We identify 32,689 sites with quantitatively increasing accessibility, hereafter termed 'remodeled DHSs', compared to only 1,573 sites with decreasing accessibility (Figure 4.5b). Strikingly 23,861 (73%) of remodeled DHSs are novel DHSs not reproducibly detected in the SMARCA4$^{-/-}$ samples. The remodeled sites are primarily distal regulatory elements (particularly within gene bodies) and are enriched for epithelial and fibroblast specific DHSs (Thurman et al. 2012) (Supplementary Figure #5a-b) analogous to the dependence of cell

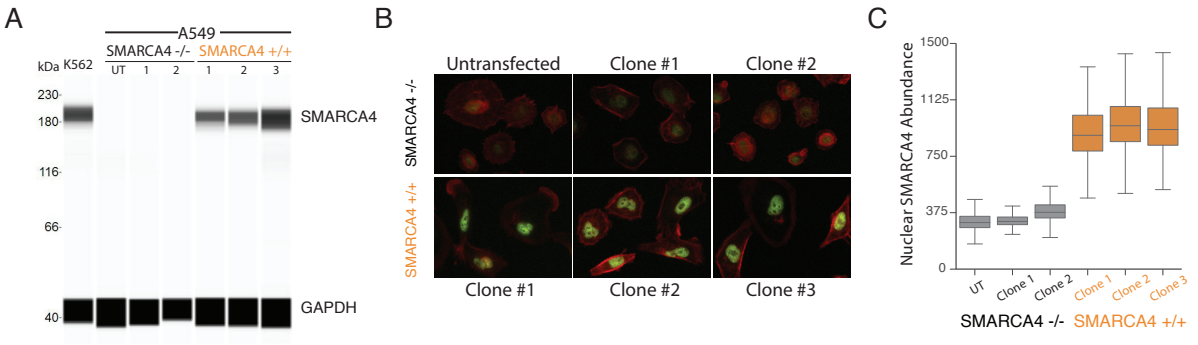


Figure 4.3: Reactivation of SMARCA4 protein expression

(a) Detection of restored SMARCA4 protein expression by Western-Blot in untransfected (UT), two unedited and three pan rescue A549 clones. K562 cells are included as a SMARCA4 positive control. GAPDH served as loading control.

(b) SMARCA4 (green) and actin (red) immunofluorescence imaging of A549 mutant and rescue clones.

(c) Image-based quantitation of SMARCA4 nuclear protein abundance. Nuclei of rescued clones contain significantly higher amounts of SMARCA4 relative to mutant cells.

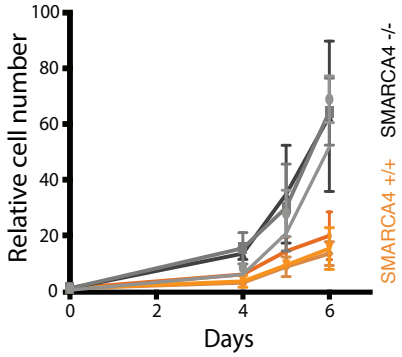


Figure 4.4: Growth curves for SMARCA4$^{+/+}$ and SMARCA4$^{-/-}$ clones.

type specific enhancers on SMARCA4 observed in other cell types (Alver et al. 2017; Attanasio et al. 2014).

Of the remodeled DHSs, 60% overlap a SMARCA4 CUT&RUN peak call in the SMARCA4 $^{+/+}$ clones, and 87% have SMARCA4 CUT&RUN signal greater than the median unaffected DHS (Figure 4.5c, Supplementary Figure #6a). Since mammalian SWI/SNF complexes contain either SMARCA4 or SMARCA2, we analyzed the extent to which the sensitivity of the remodeled DHSs to SMARCA4 activity reflects specific recruitment of SMARCA4 containing complexes over SMARCA2 complexes. SMARCA2 and SMARCA4 occupancy is highly correlated (Supplementary Figure #6b) and most of the remodeled DHSs are bound by both ATPases in SMARCA4$^{+/+}$ samples (Figure 4.5c). The SMARCA4:SMARCA2 ratio at re-

modeled DHSs that overlap a SMARCA4 peak is only modestly higher than an SMARCA4 bound, unaffected DHSs (Supplementary Figure #6c) suggesting relative SMARCA4 recruitment is a contributing rather than definitive factor in DHS sensitivity to SMARCA4 loss. Instead, the SMARCA4 bound, remodeled sites are better characterized by the levels of SMARCA2 occupancy in SMARCA4$^{-/-}$ samples: only 40% of SMARCA4 bound, remodeled sites overlap a SMARCA2 peak in SMARCA4$^{-/-}$ cells compared to 74% for SMARCA4 bound, unaffected sites. However, after SMARCA4 rescue, SMARCA2 occupancy increases specifically at the remodeled sites to levels similar to the unaffected DHSs (Supplementary Figure #6c) suggesting that both affected and unaffected sites have the potential to recruit BAF at similar levels. Taken together, these results indicate that while the occupancy profiles of SMARCA4 and SMARCA2 at accessible sites are comparable genome-wide, there is also a large subset of dependent elements that only become bound by SMARCA2 once they have been remodeled by SMARCA4.

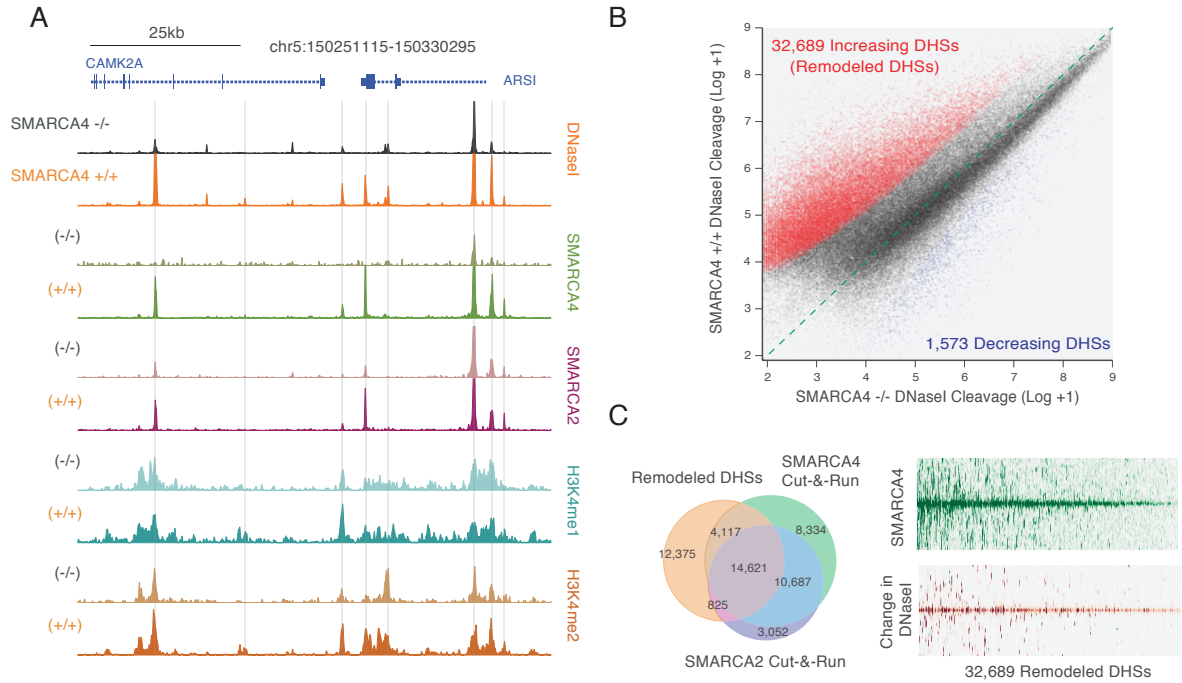


Figure 4.5: SMARCA4 reactivation causes global increase in accessibility

(a) DNaseI cleavage, SMARCA2 and SMARCA4 CUT&RUN, and histone modification CUT&RUN read density in A549 SMARCA4$^{-/-}$ and A549 SMARCA4$^{+/+}$ cells shown for a representative genomic region. Remodeled DHSs are highlighted with grey stripes.

(b) Scatterplot of DNaseI cleavages in master list elements in SMARCA4$^{-/-}$ versus SMARCA4$^{+/+}$ samples for quantitative comparisons. Note the asymmetric increase in accessibility in the SMARCA4$^{+/+}$ samples.

(c) Relationship between SMARCA2/4 CUT&RUN in SMARCA4$^{+/+}$ samples and remodeled DHSs. Right: Venn diagram of remodeled DHSs and SMARCA2 and SMARCA4 CUT&RUN peaks in SMARCA4$^{+/+}$ samples. Left: Heatmap of SMARCA4 CUT&RUN signal in SMARCA4$^{+/+}$ samples (top) and change in DNaseI signal (bottom) at the 32,689 DHSs with significant (FDR<0.05, FC>1.5) increases in accessibility. Signal is averaged across experiments from three independent clones and DHSs are ordered by SMARCA4 CUT&RUN signal.

To identify transcription factors (TFs) that might recruit BAF/PBAF complexes to target sites in A549s, we performed *de novo* motif discovery on the remodeled DHSs and identified AP-1 (JUN/FOS), RUNX, and TEAD motifs as highly enriched in the activated DHSs (Supplementary Figure #7a). AP-1 and TEAD factors recruit BAF directly or through coactivators (Ito et al. 2001; Skibinski et al. 2014), and supporting the motif enrichment, ChIP-Seq peaks of AP-1 members JUNB and FOSL2 have the highest overlap with the remodeled DHSs for the 51 TFs assayed by ENCODE in A549 (Figure 4.6a) (Consortium et al. 2012). For sites overlapping both JUNB and SMARCA4 peaks, chromatin remodeling occurs preferentially at the subset sites with low chromatin accessibility in SMARCA4$^{-/-}$ cells. The weakly accessible sites contain low occupancy AP-1 binding sites as measured by DNaseI footprinting, a direct readout of factor occupancy (Figure 4.6b) (Vierstra and John A Stamatoyannopoulos 2016). The sequences of low occupancy sites have, on average, weaker motif matches, suggesting the low occupancy is due in part to lower TF binding affinity at the sites. Upon restoration of SMARCA4, there is an increase in occupancy over the AP-1 motif at these sites to levels observed for the higher affinity motifs (Figure 4.6b). The phenomenon is not specific to AP-1 factors, as we observe similar results for sites that are bound by both SMARCA4 and SP1 (Supplementary Figure #7b). Taken together, these data suggest that SMARCA4 can potentiate TF occupancy at co-bound sites and can markedly increase binding to weaker motifs through remodeling.

Since SMARCA4 rescue activates low occupancy regulatory sites, we hypothesized the novel DHSs that appear after SMARCA4 rescue would be pre-marked for regulatory activity in SMARCA4$^{-/-}$ cells. Compared to a background of DHSs active in other cell types, we observe a higher fraction of novel DHSs overlap TF ChIP-Seq peaks for the 51 assayed TFs (Supplementary Figure #7c). However ~70% of the novel DHSs are not bound by any of the 51 TFs assayed in A549s, demonstrating a high degree of co-dependence between TF binding and BAF/PBAF activity. To extend the analysis more generally, we analyzed histone marks surrounding all remodeled DHSs. In SMARCA4$^{-/-}$ samples, the remodeled DHSs are marked by similar levels of H3K4me1 to A549 DHSs despite markedly lower chromatin accessibility (Figure 4.6c). After SMARCA4 rescue, H3K4me1 increases at the remodeled DHSs to levels higher than unaffected DHSs (Figure 4.6c), although the ratio of H3K4me1 to chromatin accessibility decreases towards the genomic average at DHSs (Figure 4.6d). We observe similar patterns for H3K4me2, but relative to H3K4me1, the H3Kme2 signal tracks closer to the levels expected from a region's chromatin accessibility (Figure 4.6c-d). Therefore, at remodeled sites, SMARCA4 rescue increases chromatin accessibility towards levels predicted by other measures of regulatory activity, and H3K4me1 specifically pre-marks the remodeled DHSs. This finding is consistent with a recent report the BAF complex preferentially binds

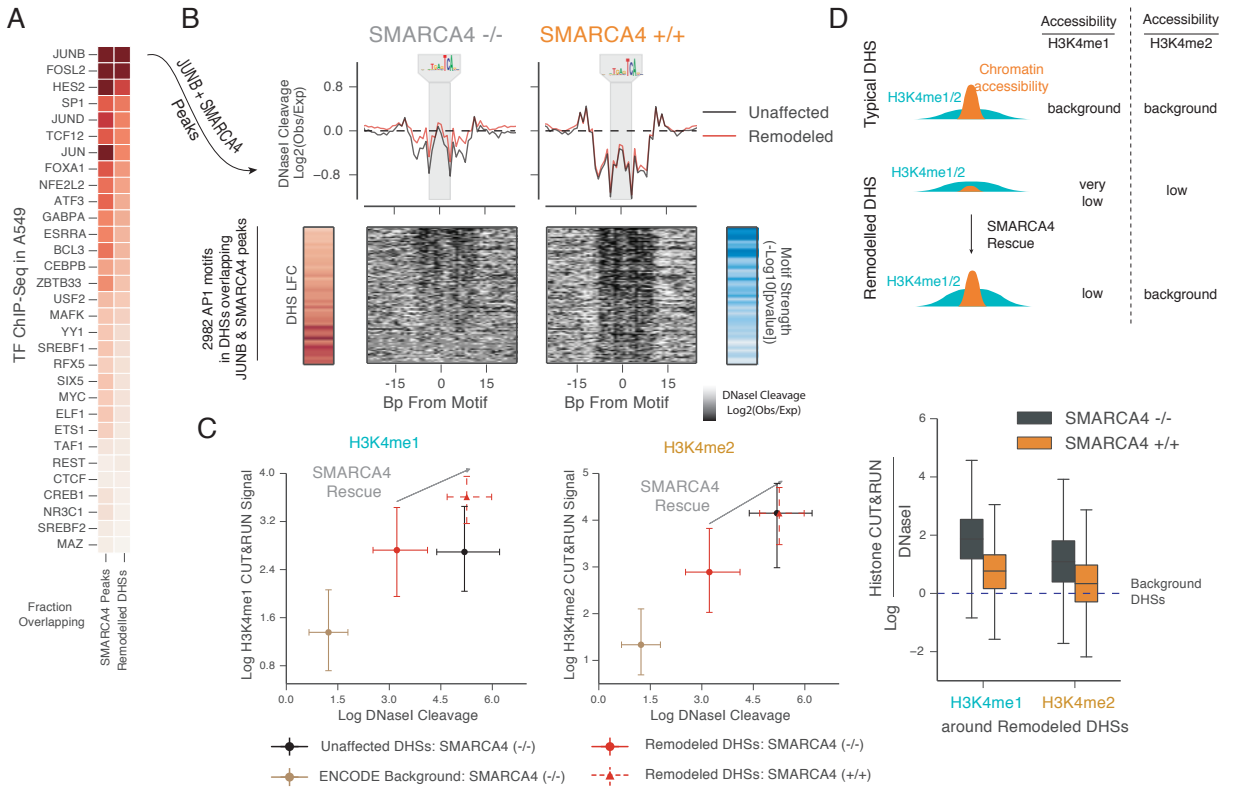


Figure 4.6: SMARCA4 preferentially remodels low affinity sites

(a) Overlap of SMARCA4 CUT&RUN peaks and remodeled DHSs with TF ChIP-Seq peaks for TFs assayed by ENCODE in A549s. All assayed TFs with at least 500 called peaks are included.

(b) Top: Aggregate cleavage profile (log2 observed/expected) around AP1 motifs in remodeled and unchanging DHSs in SMARCA4$^{-/-}$ and SMARCA4$^{+/+}$ samples. Motifs were selected from DHSs that overlap a SMARCA4 CUT&RUN peak and JUNB ChIP-Seq peak and are footprinted in either SMARCA4$^{-/-}$ and SMARCA4$^{+/+}$ cells. Bottom: DHSs were ordered by ordered by motif footprint occupancy in SMARCA4$^{-/-}$ samples. From left to right heatmaps display the average log fold change in accessibility after SMARCA4 reactivation, DNaseI cleavages around AP1 motif in SMARCA4$^{-/-}$ cells, DNaseI cleavages around AP1 motif in SMARCA4$^{+/+}$ cells, and AP1 motif strength. Ordering of DHSs (rows) is consistent across heatmaps. To highlight the trend with footprint occupancy, for heatmaps displaying DHS log fold change and motif pvalues, DHSs were separated into 25 bins and average values for the bin are shown.

(c) DNaseI cleavage versus H3K4me1 (left) and H3K4me2 (middle) CUT&RUN signal is plotted for remodeled DHSs (red), unaffected DHSs (black), and a background of DHSs accessible in other cell types (tan). Scatterplots show the median signal for each class of elements with error bars showing the 25th-75th percentile. Right: Boxplots show the ratio of H3K4me1 and H3K4me2 signal to DNaseI signal at remodeled DHSs. Values are normalized to the average histone CUT&RUN:DNaseI cleavage ratio in unaffected DHSs.

(d) Schematic of relationship between remodeled DHSs and enhancer associated histone marks.

to and remodels H3K4me1 marked nucleosomes (Local et al. 2018).

Overall, our results demonstrate the chromatin accessibility landscape of A549 is highly

dependent on SMARCA4 activity: after SMARCA4 rescue, the accessible compartment of the genome compartment expands by 21%. The remodeling occurs preferentially at latent regulatory sites with low TF occupancy but which are pre-marked by H3K4me1. Reintroduction of SMARCA4 causes increased TF occupancy at these sites and a corresponding increase in the enhancer associated marks H3K4me1/me2.

Promoter state modifies the association between distal element activation and gene expression

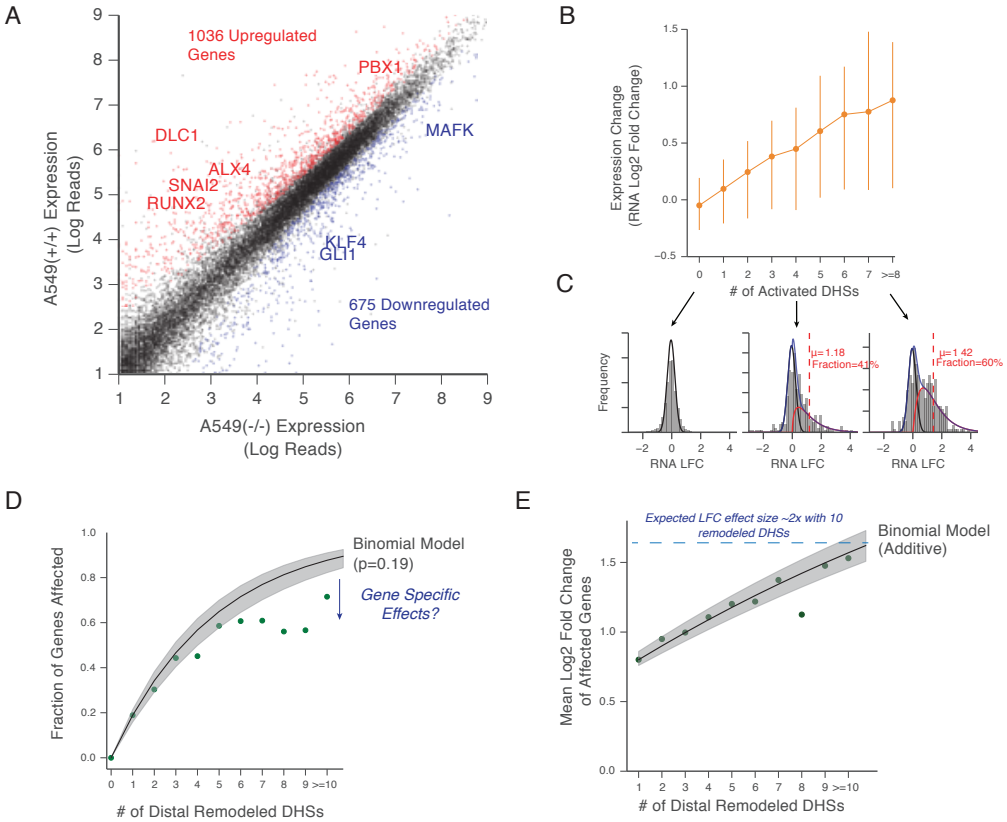


Figure 4.7: Attenuated transcriptional response to SMARCA4 reactivation

(a) Scatterplot of expression in SMARCA4$^{-/-}$ versus SMARCA4$^{+/+}$ clones. Select up and downregulated genes are highlighted. Note the attenuated change in gene expression compared to chromatin accessibility (Compare to Figure **??**b).

(b) Mean change in expression at genes grouped by number of SMARCA4 responsive DHSs (errobars show 25th-75th percentile)

(c) Distribution of changes in gene expression for genes with n = 0, n = 4, and n>=7 responsive DHSs. For each distribution, a mixture model was fit to deconvolve unaffected genes (black, distribution of expression for genes with n = 0 responsive DHSs) and affected genes (red). Mean of the affected distribution and the fraction of genes assigned to the affected distribution are reported.

We next compared the chromatin changes to the transcriptional response to understand

the relationship between the widespread increase in accessibility and SMARCA4's effects on gene expression. The genes closest to remodeled DHSs are, on average, upregulated (Supplementary Figure #8a). Additionally, the link between SMARCA4 binding sites and the transcriptional response following SMARCA4 rescue is conditional on chromatin remodeling at the site: genes near SMARCA4 binding sites that do not overlap a remodeled DHS do not have increased expression, even for CUT&RUN peaks within 1kb of a gene's TSS (Supplementary Figure #8b). However, despite the global association between chromatin and expression, the changes in expression following SMARCA4 restoration are highly attenuated compared to the increase in chromatin accessibility (Figure 4.7a, Supplementary Figure #8c).

We sought to explore the discrepancy between the magnitude of chromatin and transcriptional changes by analyzing how local changes in chromatin accessibility associate with gene expression. We linked all DHSs to their nearest transcription start site (TSS) and found a dose-dependent relationship between the number of activated DHSs neighboring a gene and the average change in expression (Figure 4.7b). For instance, genes linked to 8 or more SMARCA4-activated DHS display on average a 1.84-fold increase in expression compared to an average 1.07-fold change for genes linked to a single activated DHS.

The higher average increase in expression of genes neighboring larger numbers of remodeled DHSs could be caused by either: (a) dispersed control of gene expression, with many DHSs contributing relatively modest effects, and/or; (b) by focal control, wherein a single DHS affects the gene's expression, and the higher the number of neighboring remodeled DHSs, the higher the probability of a regulatory impact. Under the focal model, the increase in average expression should be primarily driven by a change in the fraction of genes that are upregulated, whereas in a dispersed model, the average change should be driven more by an increase in the magnitude of upregulation. To distinguish between the two possibilities, we fit a mixture model to distinguish between the fraction of upregulated genes and the magnitude of expression change for upregulated genes. We find that the increased average upregulation for genes neighboring greater number of DHSs appears primarily due to an increase in the fraction of genes that are upregulated (Figure 4.7c). This is consistent with a focal model in which the attenuated changes in gene expression arise due to only a subset of remodeled DHSs (approximately 20%) driving changes in gene expression (Figure 4.7d-e). However, we also observe the fraction of genes upregulated at higher numbers of remodeled DHSs is lower than expected based on a simple model of a constant fraction of DHSs affecting gene expression (Figure 4.7d). This saturation in the proportion of genes changing expression suggests locus specific effects outside of individual DHSs might play a

role in whether a gene's expression responds to distal element activation.

We hypothesized gene specific effects might arise, in part, due to promoter states modifying the effects of distal DHSs. To test the relationship between promoter state and the association between distal element remodeling and gene expression, we first fit a function to weight each distal element to best explain the changes in expression (Methods). We then performed multiple linear regression, regressing genes' change in expression on the weighted change in chromatin accessibility, promoter state, and interaction terms between the two (Methods). We characterized the promoter state for each gene by CpG island status, CpG methylation, overlap with H3K27me3 peaks, overlap with H3K4me3 peaks, and bivalency (both H3K4me3 and H3K27me3 peaks) in SMARCA4$^{-/-}$ A549s.

By themselves, promoter states had mostly non-significant relationships with changes in gene expression, but we observe strong interactions between the change in distal chromatin accessibility and promoter state (Supplementary Table #5). In particular, promoter bivalency had positive interaction terms with the change in distal DHSs (Figure 4.8a), providing further rationale for the observed association between BAF/PBAF activity and bivalent genes (Nakayama et al. 2017). In contrast, CpG methylation and H3K4me3 status had significant negative interactions with DHS change (Figure 4.8a-b), with genes whose TSSs were marked solely with H3K4me3 in particular showing limited response to activation of nearby distal DHSs (Figure 4.8a). Therefore, while SMARCA4 rescue activates a specific class of genes with repressed promoters marked by H3K27me3, changes at distal DHSs due to SMARCA4 rescue appear to have moderated effects on genes with promoters that are already maximally active (only H3K4me3 signal) or stably repressed (high CpG methylation). For example, although the genes *GALNT15*, *DLC1*, and *LAMC1* all show dramatic increases in distal DHSs near their TSS and in the gene body, they show distinct transcriptional responses to SMARCA4 that are associated with their specific promoter states prior to reactivation (Figure 4.8c).

In contrast to the widespread changes in chromatin accessibility after SMARCA4 reactivation, the effects on expression are more modest. While there is a strong association between chromatin remodeling at distal DHS and changes in gene expression, the effect appears modified by a gene's promoter state and therefore might be influenced by epigenetic changes to a gene's promoter.

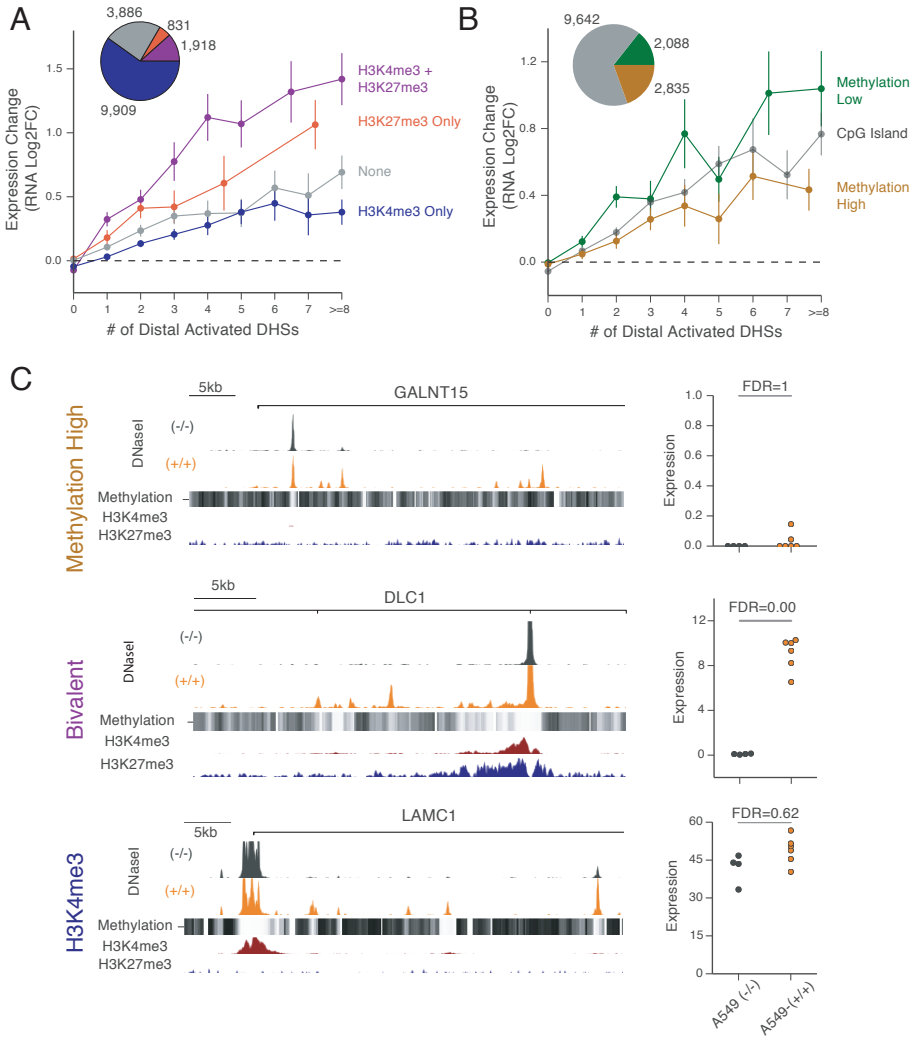


Figure 4.8: Promoter state gates response to changes in distal accessibility

(a) Genes with bivalent promoters show greater changes in RNA-Seq compared to genes with similar numbers of changing distal DHSs. Plotted is the mean change in expression vs number of SMARCA4 responsive DHSs with gene sets grouped by ChIP-Seq peaks at the gene's promoter. Error bars +/- S.E.M.

(b) Genes with high levels of promoter methylation show smaller changes in RNA-Seq compared to genes with similar numbers of changing distal DHSs. Plotted is the mean change in expression vs number of SMARCA4 responsive DHSs for methylation high and methylation low gene sets. Error bars +/- S.E.M.

(c) Examples of changes in chromatin accessibility (DNaseI cleavage profile, left) and gene expression (swarmplot, right) for three genes with similar changes in distal chromatin accessibility but different promoter states.

SMARCA4 rescue causes regional chromatin remodeling

Since genes located near multiple activated DHSs showed the largest response to SMARCA4 reactivation, we next investigated whether the grouping of SMARCA4 sensitive DHSs re-

flected a random distribution or spatial clustering. Compared to a block bootstrapped background, SMARCA4 sensitive DHSs show significant spatial clustering (Supplementary Figure #9a). We therefore fit a Hidden Markov Model (HMM) to identify regions of SMARCA4 sensitivity (Figure 4.9,4.10, Methods) and discovered 202 remodeled regions with an average size of 842 Kb. These regions contain 7.5% of identified DHSs and 19.5% of the remodeled DHSs.

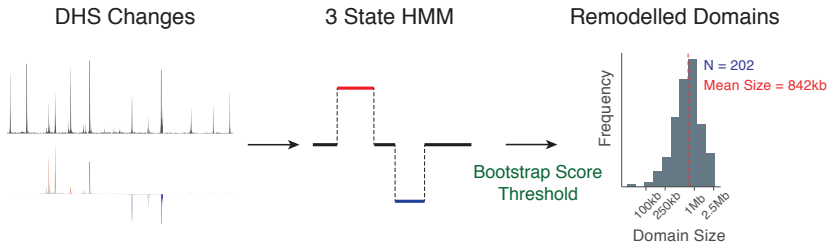


Figure 4.9: Scheme for identifying regional changes in accessibility
Changes in chromatin accessibility were segmented using a 3 state HMM. States of increasing and decreasing accessibility were compared to a block bootstrapped background to identify regions with greater changes in chromatin accessibility than expected by local clustering. 202 regions were identified with increasing accessibility with a mean size of 842 kb

In addition to changes at focal sites of accessibility (i.e. remodeled DHSs) in the remodeled regions, we observe an increase in the background accessibility to DNaseI cleavage outside of DHSs (Figure 4.11). This finding demonstrates that the chromatin changes caused by SMARCA4 reactivation are not restricted to individual elements, but instead extend over the entire remodeled region. Consistent with a model in which SMARCA4 rescue activates the entire chromatin territory, levels of H3K27me3 decline and the enhancer associated marks H3K4me1/H3K3me2 increase specifically across the remodeled regions (Figure 4.11, Supplementary Figure #9b).

Previous work has proposed that extended regions of high regulatory activity, termed super enhancers, are important in for regulating cell type specific genes and are preferentially sensitive to disruption of regulatory components (Warren A. Whyte et al. 2013). We therefore compared the SMARCA4 remodeled regions to a variation on super enhancers defined by SMARCA4 binding. The remodeled regions do not correspond to SMARCA4 super enhancers as identified by the ROSE algorithm (Lovén et al. 2013; Warren A. Whyte et al. 2013), but reflect larger regions of chromatin activation (Supplementary Figure #10a). However, the majority (74%) of these domains contain at least one SMARCA4 super enhancer (Supplementary Figure #10b), suggesting a subset of SMARCA4 binding clusters may act as bona fide locus control regions enabling chromatin activation across a region. This domain

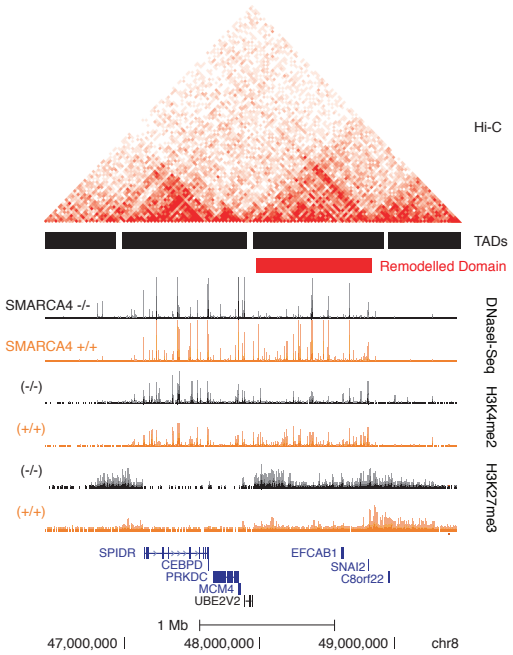


Figure 4.10: Example genome profile of locus identified as a regional change in chromatin accessibility. Genome profile showing chromatin accessibility and histone marks in SMARCA4$^{-/-}$ and SMARCA4$^{+/+}$ samples. Hi-C and TAD data from unedited A549s are included to highlight the relationship between the remodeled region and other chromatin domains.

scale chromatin activation may be a key functional consequence of clusters of SMARCA4 binding, as genes closest to the SMARCA4 super enhancers falling inside remodeled regions show large changes in expression, while genes closest to the super enhancers falling outside of remodeled regions show significantly lower change in expression (Supplementary Figure #10c-d).

Activated regions align with boundaries of topological domains

To understand the features demarcating the activated domains, we next analyzed the elements at the borders of the activated regions. Binding sites of CTCF and members of the cohesin complex, SMC3 and RAD21, are enriched at the boundary elements (Figure 4.12a). CTCF and cohesin interact to establish three-dimensional chromatin organization including long range chromatin interactions and topologically associating domains (TADs) (Ong and Corces 2014). We therefore compared the regions of SMARCA4 sensitivity to contact domains, or TADs, in A549 cells and observe a striking correspondence between activated SMARCA4 regions and TADs (Supplementary Figure #11,12a). TADs frequently occur bounded by convergent CTCF sites in close proximity leading to a model of isolated

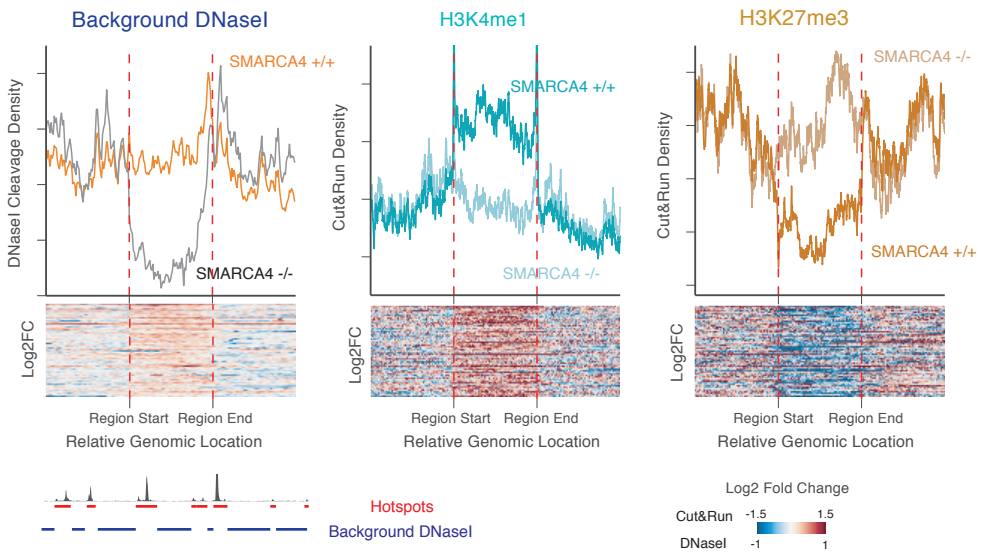


Figure 4.11: Regional chromatin activation
DNaseI background (left), H3K4me1 CUT]&RUN (middle) and H3K27me3 CUT&RUN (right) signal around identified remodeled regions in SMARCA4$^{-/-}$ and SMARCA4$^{+/+}$ clones. Top: Lineplots of the aggregate (trimmed mean) score over all regions. Bottom: heatmaps of log fold change values for individual regions.

chromatin domain formation by formation of a DNA loops (S. S. Rao et al. 2014). Comparing the SMARCA4 sensitive regions directly to Hi-C defined DNA loops in A549, we observe the boundaries of the activated regions reside in close spatial proximity as defined by Hi-C loops (Figure 4.12b, Supplementary Figure #12b). Combined, these results suggest that at a subset of loop domains SMARCA4 reactivation enables chromatin activation across the entire domain.

To investigate possible chromatin states that might specify TAD sensitivity to SMARCA4 activation, we compared the genomic features of TADs overlapping regional changes in accessibility to unaltered TADs. We do not observe a direct correspondence between H3K27me3 marked domains and SMARCA4 sensitivity. Instead, sensitive regions overlap with bivalent TADs that contain matched levels of H3K27me3 and marks associated with active chromatin, such as H3K4me3, in the SMARCA4$^{-/-}$ state (Supplementary Figure #13).

Despite identifying regional chromatin changes in an unsupervised manner agnostic to 3D chromatin conformation, we find SMARCA4 reactivation causes increases in chromatin accessibility not only at the nucleosome (or DHS) level, but also regional activation of chromatin at a subset of predefined loop domains.

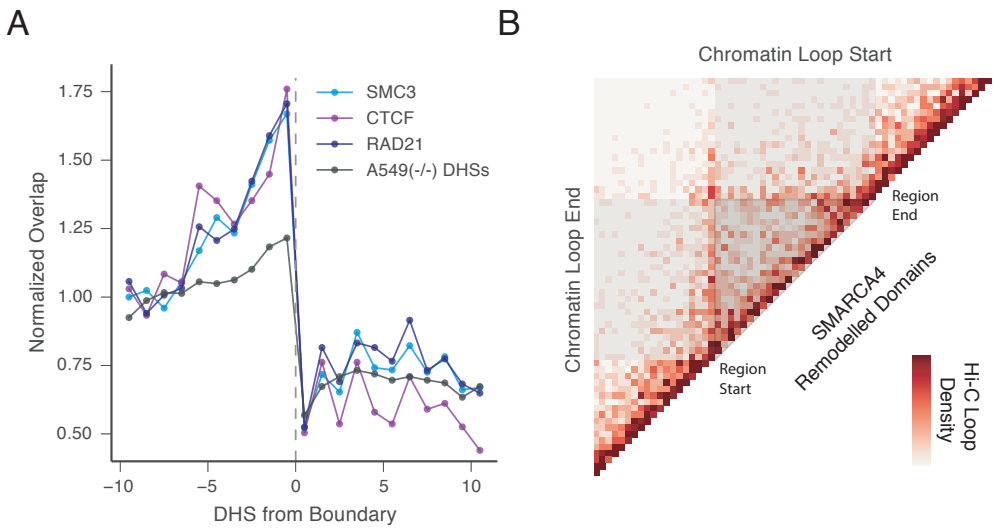


Figure 4.12: Remodeled regions align with chromatin loop domains

(a) Enrichment of CTCF and members of the cohesin complex at DHSs marking the boundaries of regional changes.

(b) Heatmap of number of observed chromatin loops neighboring the regional changes in chromatin accessibility.

Regional chromatin remodeling is associated with induction of developmental regulators

The regions of increased accessibility are enriched for the genes with the greatest response to SMARCA4 reactivation. Of the 188 protein coding genes with log2 fold change ¿ 2 in the SMARCA4 reactivated set, 34% are located in the remodeled domains (compared to 14% of all significantly upregulated genes). These highly upregulated genes include both genes with changes in promoter accessibility as well as genes with changes only at distal DHSs. Highlighting the large effect of regional chromatin remodeling on gene expression, genes found in the activated regions but without any change in promoter accessibility have similar changes in expression to genes outside these regions but with increased promoter accessibility (Figure 4.13). Further, genes that lie within the activated regions and additionally exhibit increased promoter accessibility upon SMARCA4 rescue displayed marked upregulation (Figure 4.13).

Developmentally regulated genes are proposed to exist in insulated regulatory neighborhoods (Dowen et al. 2014). (Dowen et al., 2014) Consistent with this hypothesis, we observe several lineage specifying TFs and other key developmental regulators are located in the SMARCA4 sensitive domains. The genes include the RUNX transcription factors, the EMT regulator SNAI2, and the homeodomain TF PBX1 (Figure 4.10, Supplementary Figure #11). Other genes important for epithelial morphogenesis or cell-cell interaction, such as

DLC1 and *EDIL3* are found also in the activated domains. Supporting the link between domain level activation and developmental processes more generally, genes annotated with GO terms such as *tissue development*, *cell differentiation*, and *extracellular matrix organization*, are enriched in the SMARCA4 sensitive domains (Supplementary Figure #14a). The enrichment of developmental regulators in epigenetically sensitive domains therefore links the largest effects of SMARCA4 on gene expression towards developmental regulatory programs despite relatively non-specific genome-wide increases in chromatin accessibility.

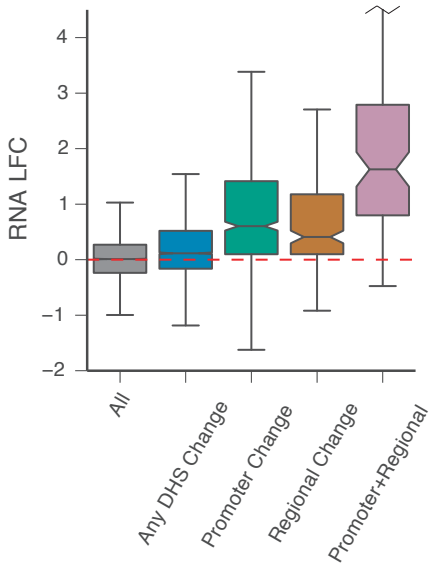


Figure 4.13: Highly upregulated genes lie in remodeled regions
Boxplots of change in expression of all genes, genes near any changing DHSs, genes with an increasing promoter DHS, genes located in region of increasing accessibility without a change in promoter accessibility, and genes located in region of increasing accessibility with a change in promoter accessibility.

Chromatin context of expression programs altered in SMARCA4-null patients

We next considered how the observed relationships between baseline chromatin state, increases in chromatin accessibility, and changes in gene expression might affect SMARCA4's targets in lung adenocarcinoma. We first identified a set of likely relevant SMARCA4 targets by selecting genes with lower expression in TCGA lung adenocarcinoma SMARCA4-null samples. In A549s, expression of the TCGA gene set matches the expression in SMARCA-null patient samples (Supplementary Figure #15a). Rescue of SMARCA4 partially upregulates the repressed gene set (Figure 4.14a-b), consistent with previous work showing partial overlap between genes affected by SMARCA4 knockdown in lung adenocarcinoma cell lines

and genes with decreased expression in SMARCA4 null patient samples (Orvis et al. 2014). Compared to all upregulated genes, the reactivated genes (i.e. genes with decreased expression in SMARCA4-null patient samples and upregulated after SMARCA4 rescue) are enriched in SMARCA4 remodeled domains, with 28% of the reactivated genes lying within the remodeled regions.

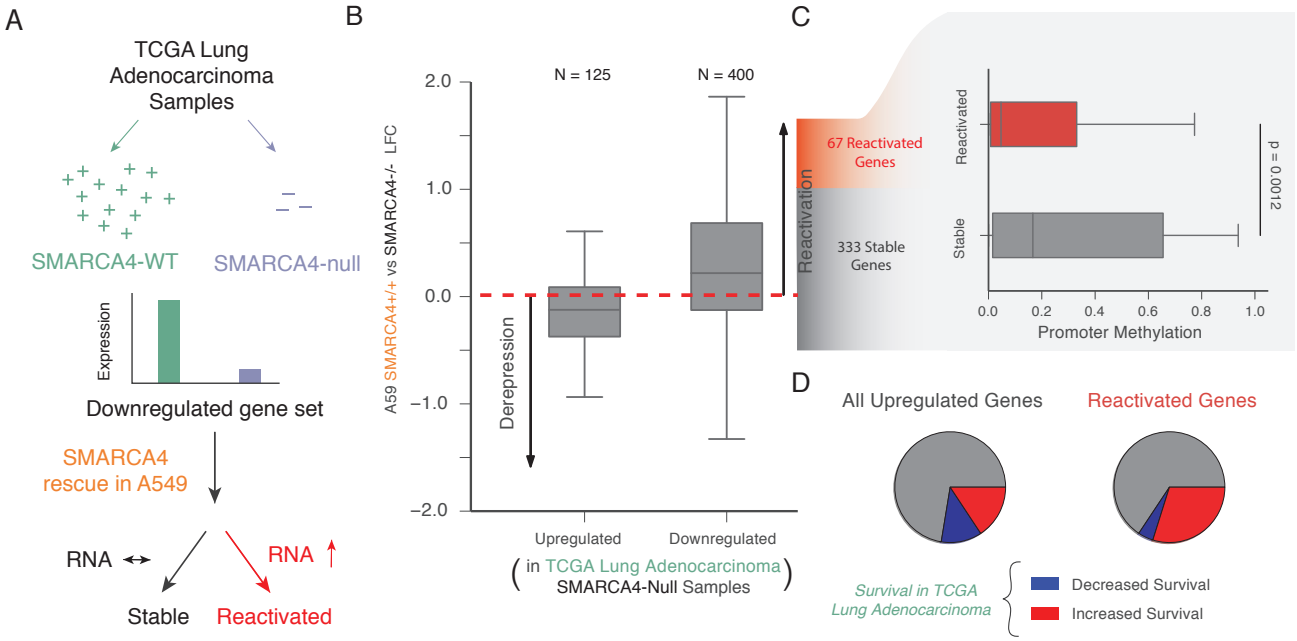


Figure 4.14: Reactivation of TCGA SMARCA4 signature

(a) Schematic of strategy to identify genes downregulated in SMARCA-null patent samples and identify the set of downregulated genes that are reactivated after SMARCA4 re-expression.

(b) Partial upregulation of SMARCA4-null associated genes after SMARCA4 rescue in A549 cells. Reactivated genes are defined as genes repressed in A549 (significantly lower in SMARCA-null tumor samples and >1.5 fold difference between A549 expression and the average in SMARCA4 WT tumor samples) with a significant (FDR<0.05) increase in expression.

(c) Reactivated genes show lower levels of promoter methylation in A549 cells compared to the stable genes (pvalue < 0.01, Wilcoxon rank-sum test).

(d) Reactivated genes show a bias towards positive associations with survival compared to all upregulated genes. Pie charts show the fraction of reactivated (right) and all upregulated genes (left) with positive, negative, and no association with survival (uncorrected pvalue < 0.05).

Since we observe that promoter state modifies the effect of distal accessibility changes, we considered whether it might affect also SMARCA4's ability to reactivate a gene's expression following the various epigenetic changes that occur during the cancer's evolution. The reactivated genes have lower promoter methylation than stable genes (Figure 4.14b), but similar levels of other repressive marks (H3K27me3 and H3K9me3) that were not linked to a diminished association between gene expression and distal DHS activation (Supplementary

Figure #15b). And while A549s' bivalent genes were more likely to be differentially expressed between SMARCA4-null and SMARCA4-WT patient samples, they were not significantly more likely to be reactivated after SMARCA4 rescue (pvalue=0.57). Promoters of tumor suppressors are frequently methylated in cancers, so the diminished reactivation of genes with highly methylated promoters suggests a mechanism for how accumulated epigenetic changes in cancer could alter SMARCA4 targets at different stages of cancer development.

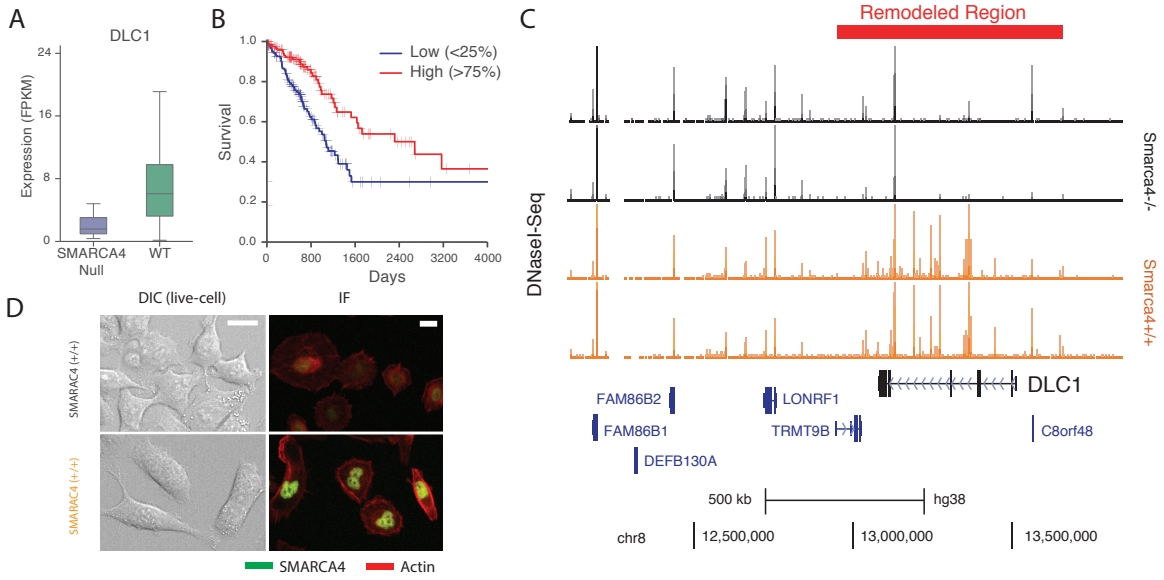


Figure 4.15: Example of reactivated gene

(a) Example of a reactivated gene: Boxplots of DLC1 expression in SMARCA4-null and SMARCA4-WT lung adenocarcinoma samples from TCGA shows decreased expression the SMARCA4-null tumor samples.

(b) Kaplan-Meier curve showing increased survival in DLC1 high (>75th percentile) patients compared to DLC1 low (<25th percentile) patients. Only SMARCA4 WT samples were included.

(c) SMARCA4$^{+/+}$ cells show changes in actin organization and cellular shape reminiscent of changes previously observed due to expression of DLC1 in cell lines with silenced DLC1.

(d) Example chromatin change around a reactivated gene located in a SMARCA4 sensitive chromatin domain. Shown are the DNaseI cleavage profiles at the DLC1 locus in SMARCA4$^{+/+}$ and SMARCA4$^{-/-}$ samples. The identified regional chromatin change is annotated.

The reactivated genes are strong candidates for mediating SMARCA4's role in lung adenocarcinoma and include genes in pathways involved in lung adenocarcinoma pathogenesis such as WNT (DAAM2, WLS, and TNIK) (Stewart 2013) and receptor tyrosine kinase signaling (ROS1, PTPRE) (Network, Collisson, et al. 2014). Compared to all genes upregulated in the SMARCA4$^{+/+}$ clones, the 90 reactivated genes are biased for genes associated with increased patient survival in wild type SMARCA4 samples ($\tilde{\chi}^2$ test, pvalue < 0.01, Figure

4.14c). The reactivated genes include genes associated with decreased cancer cell proliferation and invasiveness (Qian et al. 2007; Saintigny et al. 2012; Torrino et al. 2019), most notably DLC1, a known tumor suppressor (Yuan, Jefferson, Baldwin, et al. 2003). DLC1 regulates the actin organization (Sekimata et al. 1999), and restoration of its expression in cancer cells can induce apoptosis (Xiaoling Zhou, Thorgeirsson, and Popescu 2003), decrease migration and invasion (Goodison et al. 2005), and reduce cell growth (Yuan, Jefferson, Baldwin, et al. 2003). *DLC1* resides in a SMARCA4 sensitive chromatin domain (Figure 4.15c) and is the most highly upregulated gene after SMARCA4 rescue (Figure 4.7a,4.8c). It has lower expression in SMARCA4-null patient samples (Figure 4.15a), and its expression is positively associated with patient survival (Figure 4.15b). Interestingly, we observe changes in cell morphology after SMARCA4 reactivation (Figure 4.15d)—in line with previous literature linking mutations in SMARCA4 to changes in cell morphology and cytoskeletal disorganization (Wong et al. 2000) (Wong et al., 2000)—that match the description of cytoskeletal changes after ectopic expression of DLC1 in lung adenocarcinoma cells (Yuan, Jefferson, Millecchia, et al. 2007).

Across multiple scales—from individual TF binding to chromatin domain organization—we observe a tight relationship between genes' chromatin states and the transcriptional effects of SMARCA4 binding. This interaction between SMARCA4 and chromatin state adds additional transcriptional specificity to SMARCA4 perturbation beyond recruitment (Figure 4.16) and likely plays a role defining the developmental and oncogenic expression programs regulated by SMARCA4 in lung adenocarcinomas and other systems.

4.3 Discussion

Mutations in chromatin remodelers such as SMARCA4 are increasingly implicated in cancer development as well as a range of developmental diseases. A primary challenge in understanding these complexes' roles in human disease is understanding their targets in the relevant cell types and the mechanism by which they affect target gene expression. Here we leveraged a well profiled model system to disentangle the relationship between SMARCA4's chromatin remodeling activity and its effects on gene expression.

SMARCA4 causes extensive increases in chromatin accessibility and enhancer associated marks at cell type specific regulatory elements (Bao et al. 2015; Bossen et al. 2015; H. C. Hodges et al. 2018; Kelso et al. 2017), highlighting the central role of nucleosome competition in eukaryotic gene regulation. Recruitment of BAF/PBAF by sequence-specific TFs, including AP-1 and TEAD transcription factors, links the effects of SMARCA4 reactivation

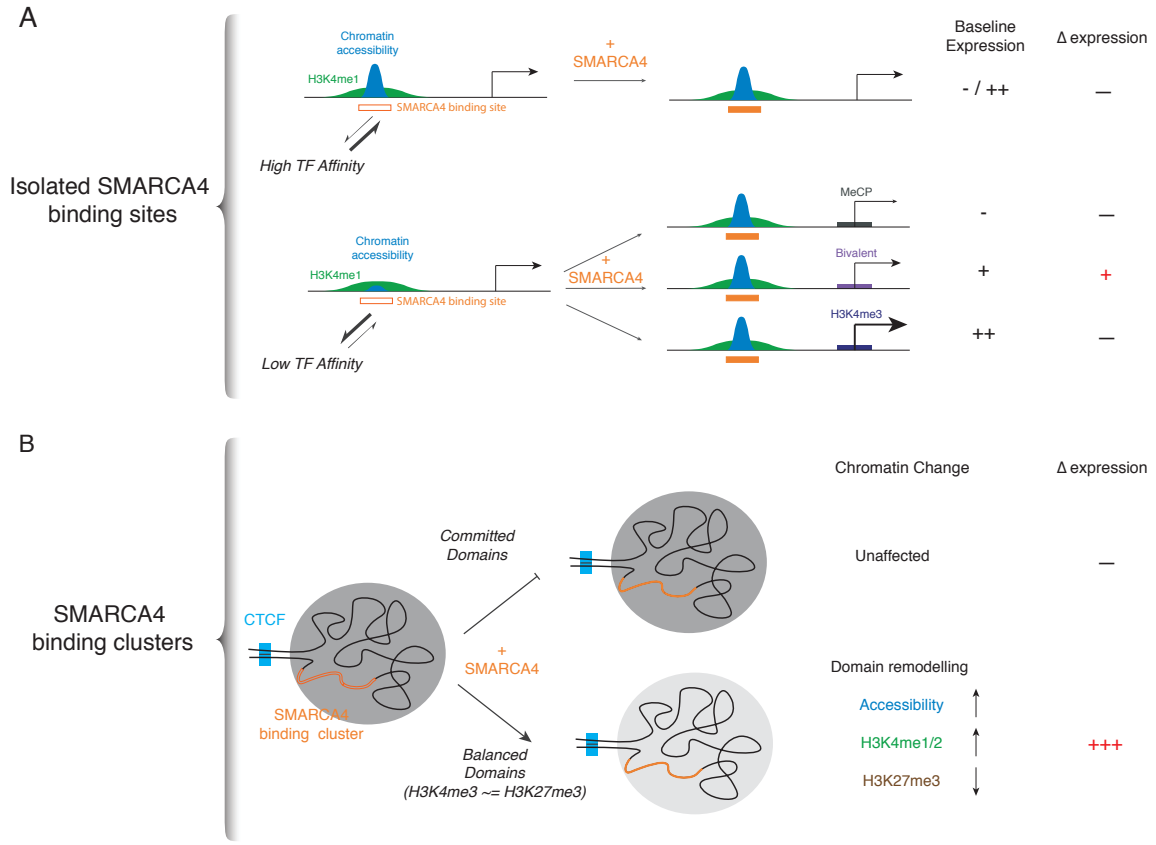


Figure 4.16: Model of relationship between SMARCA4 binding and the transcriptional effects of SMARCA4 reactivation.
The transcriptional response associated with SMARCA4 binding sites depends on chromatin remodeling. Remodeling occurs for the subset of SMARCA4 bound sites containing low affinity TF binding sites. The transcriptional effects of the remodeled sites are gated by genes' promoter states. For genes near dense clusters of SMARCA4 binding, the extensive remodeling of in distal elements leads to promoter independent changes in expression.

to specific elements. However, recruitment explains only a portion of SMARCA4's target specificity. SMARCA4 binding sites are associated with changes in expression only in the context of chromatin remodeling, which occurs preferentially at a subset of SMARCA4 binding sites with low affinity TF binding. Additionally, even among genes with similar increases in neighboring chromatin accessibility, we observe heterogeneous transcriptional responses to SMARCA4 activity: while 21% of protein coding genes neighbor at least one activated element (closest gene or within 5kb of the TSS), only 5% have increased expression. The attenuation in SMARCA4's effects on transcription compared to chromatin accessibility can partially be explained by the modification of the association between distal element activation and gene expression by a promoter's epigenetic state.

This interaction can notably be observed at genes with promoters marked solely by

H3K4me3, which show minimal change in expression even when neighboring multiple activated DHSs. We interpret the distinct behavior of H3K4me3 marked promoters compared to bivalent or H3K27me3 marked promoters to suggest that promoters have an intrinsic maximum level of expression, and additional recruitment of regulatory factors may have minimal effects on expression once that level is reached. For 51% of protein coding genes in A549, the most accessible TSS is marked solely by H3K4me3. Further increases in expression of these genes might require activation of an alternative promoter, and consistent with this hypothesis, we observe chromatin accessibility at alternative TSSs is preferentially affected and genes with annotated alternative promoters are more likely to be upregulated after SMARCA4 reactivation (Supplementary Figure #16). While alternative TSSs introduce isoform diversity, they may also have a function in allowing for graded levels of gene expression through sequential recruitment.

Mutations in BAF/PBAF subunits, and other chromatin remodelers, are frequently linked to dysregulation of developmental programs. Here we observe increased expression in developmentally regulated genes annotated with GO terms such as *Epithelial Cell Differentiation*, *Regulation of Cell Morphogenesis*, *Positive Regulation of Cell Development* after SMARCA4 reactivation (Supplementary Figure #14b). Analysis of SMARCA4-null lung adenocarcinoma samples from TCGA highlights a subset of these upregulated genes related to cell morphology, adhesion, and motility as likely important for the oncogenic phenotypes of SMARCA4. Genes with the largest increase in expression are located in isolated chromatin domains and are surrounded by regional changes in chromatin accessibility. We defined regional chromatin changes in an unsupervised manner and find they closely align with loop domains defined by Hi-C. In these domains, we observe chromatin state shows a switch like behavior with activation of chromatin across the entire domain. Many of these regions contain clusters of SMARCA4 binding sites that might act as locus control regions. We propose the activation of locus control regions explains the switch like behavior by transitioning the domain to a more active chromatin context and thereby allowing for additional regulatory activity across the entire domain.

Perturbations to a diverse set of general regulatory factors cause highly specific transcriptional responses—often preferentially affecting developmental and oncogenic programs—despite limited targeting specificity. Our results, in which a widespread chromatin change is associated with a more specific transcriptional response, demonstrate how increased regulatory activity can be interpreted in highly heterogeneous ways. Chromatin features including promoter context and domain level organization of the genome shape the selection of a regulatory factor's target genes and provide a layer of transcriptional specificity to perturbations

of the basal regulatory machinery.

4.4 Methods

Cell Culture

A549 (ATCC CCL-185) cells were maintained in F-12K Medium (ATCC, Cat #30-2004) supplemented with 10% HyClone^{TM} FBS (GE Healthcare Life Sciences, Cat #SH30071.03) and 1% Penicillin-Streptomycin (Corning, Cat #30-002-CI). Cells were passaged every 4-5 days and detached using Accutase (Innovative Cell Technology, Cat #AT-104).

Synthesis of TALEN constructs

TALEN monomers were cloned using adaptations of previously described methods (Cermak et al. 2011; Sakuma et al. 2013). To target the Q729fs*4 mutation site in SMARCA4, 5 TALEN monomers were designed to target each of the sense and antisense strands. Sense-targeting TALENs recognize the following sequences: L1, 5'-TGAATATGGCGTGTCCCAGGCC-3'; L2, 5'-TGGCGTGTCCCAGGCCCTTGC-3'; L3, 5'-TGGCGTGTCCCAGGCCCTTGCA-3'; L4, 5'-TGTCCCAGGCCCTTGCACGTG-3'; L5, 5'-TCCCAGGCCCTTGCACGTGGCC-3'. Antisense-targeting TALENs recognize the following: R1, 5'-TGTCCACTCTCTCAGTGACAGC-3'; R2, 5'-TGCTTGTCCACTCTCTCAGTG-3'; R3, 5'-TGCTTGTCCACTCTCTCAGT-3'; R4, 5'-TGACTGCTTGTCCACTCTCT-3'; R5, 5'-TAAGCGCTGACTGCTTGTCCAC-3';

Transfections

For all transfections, a BTX ECM830 device (BTX Harvard Apparatus) with a 2 mm gap cuvette was used. TALEN mRNAs were prepared using a mMessageMachine T7 Ultra Kit (#AM1345, Ambion). Per transfection, 2 x 10^5 cells were collected and washed twice with PBS. Cell pellets were resuspended in 100 μL BTXpress Electroporation Solution (BTX Harvard Apparatus, Cat#45-0805) together with 2μg mRNA per TALEN Monomer and 2 μL 100 nM 90mer ssODN containing the 23 nt corrective insertion (underlined) and 33-34 nt homology arms: 5'-ATATGGCGTGTCCCAGGCCCTTGCACGTGGCCT GCAGTCCTACTATGCCGTGGCCC
ATGCTGTCACTGAGAGAGTGGACAAGCAGTCAGC 3'. The cell/mRNA mixture was

transferred to the transfection cuvette and immediately electroporated with one pulse of 250 V for 5 ms. Following electroporation, cells were transferred to 12 well plates containing pre-warmed F-12K Medium and incubated at 37°C.

Assessment of HDR and indel rates

Cell pellets (~50,000 cells per sample) were washed with PBS and genomic DNA was harvested and amplified by PCR 48 h post-transfection as described previously(Morrow et al., 2018). Gene-specific portions of primers used are as follows: Forward, 5'-GCACATTGTCACAGATAGGAATGTGTG-3'; Reverse, 5'-CCGAGGCAGCTACGTGGC-3'. Additionally, primers contained sequences to incorporate Illumina TruSeq® adapters. Successful amplification was confirmed by electrophoresis using E-gel™48 and 96 cassettes (ThermoFisher Scientific). PCR products were subsequently diluted 1:200 and 1 μL was used as a template for a second round amplification to incorporate barcodes and adapters. Amplification was achieved using AccuPrime™Taq DNA Polymerase (ThermoFisher Scientific) as per the manufacturer's recommendations and using the following cycle conditions: 94°C 2 min; 12 cycles of 94°C 15 s, 60°C 30 s, 68°C 30 s; 68°C 10 min. PCR products were cleaned up using a QIAquick PCR Purification Kit (QIAGEN), quantified using a Qubit (ThermoFisher Scientific) and loaded on a MiniSeq using Mid or High Output 300-cycle Reagent Kits (Illumina). Paired end reads were merged and filtered to remove duplicate tags. Filtered reads were aligned to the SMARCA4$^{-/-}$ reference sequence and evaluated for insertions and deletions (indels). Editing efficiency (indel rate) was determined by tabulating only those indels that partially or completely overlap the recognition sequence of the TALEN dimer (and the intervening spacer). Scarless HDR was quantified as the proportion of reads that showed perfect complementarity to the wildtype (SMARCA4$^{+/+}$) reference sequence.

Clonal isolation

7 days post-transfection, edited A549 pools were sorted as single cells into a 96 well plate using a MoFlo Astrios EQ™Cell Sorter (Beckman Coulter). Cells were expanded for 17 days and subsequently split 1:10 for maintenance and 9:10 for genomic DNA, which was harvested 4 days later. Genotyping of clones was performed as described above and was confirmed by Sanger sequencing (GENEWIZ) using outer primers (5'-CTCGGCTCTCTGCAAGCT-3' and 5'-GGAGTTGTACTGGTTGTCTTGT-3'; IDT) that each lie ~700 bp away from the edit site to check for larger deletions.

Western blotting

A549 cells were lysed in RIPA buffer (ThermoFisher Scientific, Catalog #89900,) containing Complete, EDTA-free protease inhibitor cocktail tablets (Roche #11873580001, Roche). The total protein concentration was measured by Bradford assay (Bio-Rad). Samples were run on the Simple Western System (Protein Simple) with the 12-220kDa kit, loading 2-4μg lysate and antibodies for BRG1-(G-7) (Santa Cruz, #sc17796) and GAPDH (Santa Cruz, #sc47724) at 1:400 and1:2000 dilutions, respectively.

Immunofluorescence imaging and analysis

A549 cells were detached by mild trypsinization, washed twice with PBS, and then seeded onto sterile cover glasses (Fisher Scientific, 1.5, 18x18mm) placed in sterile 6-well tissue culture plates for about 15-20 minutes, following which the wells were gently filled with 2 ml of pre-warmed media (ATCC F-12K, 10%FBS, 1% Pen/Strep). Seeded cells were incubated for about 4-6 hours at 37°C to recover and adhere to the coverglass, following which they were washed 1x with PBS and then fixed with 4% PFA (Polysciences Inc, #18814-10) in PBS for 10 minutes at room temperature. Fixed cells were washed 3 times with 1x PBS, permeabilized with 0.25% Triton X-100 in PBS for 10 minutes at room temperature, blocked for 1 hour with 2% BSA (Jackson Immunoresearch, #001-000-161), and then incubated for 2 hours at room temperature with primary antibodies against SMARCA4 (SC-17796, mouse, 1:500 dilution) in 2% BSA/1x PBS. Subsequently, cells were washed 3x with 0.05% Tween-20 (Bio-rad, #161-0781) in PBS, and then incubated for 1 hr with donkey anti-mouse Cy3 (1:500 dilution, #711-166-152, Jackson Labs) secondary antibody and AlexaFluor647 Phalloidin (ThermoFisher Scientific, A22287,1:100) in 2% BSA/1x PBS. Lastly, cells were counterstained with DAPI (100ng/mL in 1x PBS) for 10 minutes at room temperature and washed 3 times with 0.05% Tween in PBS prior to mounting on glass slides using Prolong Gold (Molecular Probes P36930). Each wash step lasted 3 minutes at room temperature, and was performed on a shaker (Stovall Inc). Specimen coverslips were imaged using an inverted Nikon Eclipse Ti widefield microscope equipped with an Andor Zyla 4.2CL10 CMOS camera with a 4.2-megapixel sensor and 6.5μm pixel size (18.8mm diagonal FOV). Focused 2D cell images were acquired using a a 40x Nikon Plan Apo 0.9 NA air objective. Acquired images were subject to 3 rounds of iterative blind deconvolution using Autoquant software (version X3.3, Media Cybernetics, NY) to minimize the effect of out-of-focus blurring that is inherent to widefield microscopy optics. Deconvolved images were processed using in-house Matlab (version 2017B, Mathworks, Natick, MA) scripts to numerically estimate the SMARCA4

protein content in every cell nucleus, and for downstream statistical analysis.

DNaseI-Seq

DNaseI-Seq was performed as previously described (Hesselberth et al. 2009; John, Sabo, Thurman, et al. 2011; Thurman et al. 2012) (Hesselberth et al., 2009; John et al., 2011; Thurman et al., 2012b). Briefly, 5x10^6 cells were lysed using 0.01% IGEPAL. Nuclei were collected by centrifugation at 500g for 5 min, and DNaseI digestion was performed for 3 min at 37°C. DNaseI cleavage fragments were size selected by PEG fractionation, fragments were end repaired, and Illumina sequencing libraries were prepared using the ThruPLEX DNA-seq kit. Libraries were sequenced to a typical depth of 50M reads (Supplementary Table #1).

DNaseI-Seq Data Analysis

Read mapping and peak calling

Reads were processed and peaks of DNaseI cleavages (DHSs) were identified using the default ENCODE DCC DNase-DHS pipeline, version 2, paired-end (ENCPL202DNS) aligning to GRCh38/hg38. Consensus A549 peak calls: Peaks calls across the different DNaseI experiments were combined by first merging all peak calls from individual experiments. For each merged element, a core element was identified by the full width half maximum of DNaseI density in samples that had a peak overlapping the element. All peaks that overlapped (at least 50%) the core element were removed, and the process was repeated on the remaining peaks until no more elements were added. Since combining many datasets may increase the fraction of spurious peaks, an additional filter was applied based on the combined z-score $\frac{\sum Z-scores}{\sqrt{Peak\ Number}}$ of each element. To choose an empirical threshold, logistic regression was performed on all pairs of technical replicates to identify the z-score at which 95% of the DHSs were expected to be replicated, and the average value of all sets of replicates was used as the final threshold.

Identification of SMARCA4 sensitive DHSs

DNaseI cleavages were counted at consensus peaks using bedops tools (version 2.4.35) (Neph et al. 2012). Reads from technical replicates (i.e. different DNaseI experiments on the same clone) were added. Data were normalized using DESeq2's normalization strategy

restricted to DHSs identified in all samples. Using the R package DESeq2, SMARCA4 sensitive DHSs were defined as DHSs with a log2 fold change significantly (FDR<0.05) greater than log2(1.5).

Meme

Motif enrichment was performed using MEME (parameters: -dna -mod zoops -nmotifs 10 -minw 5 -maxw 12 -revcomp)on the top 1000 DHSs with the most significant increase in accessibility. To input sequences of identical length, the midpoint of each DHS was identified and padded by 100 base pairs. Motifs were matched to known motifs using TOMTOM.

RNA-Seq

RNA was extracted from 500K cells using the Qiagen RNeasy kit. RNA-Seq libraries were prepared using the TruSeq Stranded Total RNA kit (Illumina).

RNA-Seq Data Analysis

Reads were aligned to GRCh38/hg38 using RNA-STAR (version 2.3.1)(Dobin et al. 2013), and counts per gene (Gencode v25 basic annotation) were quantified using featureCounts (Smyth, W. Shi, and Liao 2013). To convert raw counts to FPKM, RNA counts were first normalized for library size using DESeq's median ratio normalization and the normalized counts were converted to estimated FPKM values using the average library size. Differential genes were identified using DESeq2. All genes with zero counts across all conditions were excluded from further analysis.

CUT&RUN

CUT&RUN was performed as described (Skene, J. G. Henikoff, and S. Henikoff 2018). Antibodies were obtained from following suppliers: H3K4me1 (Active Motif 39297), H3K4me2 (Active Motif 39141), H3K27me3 (Cell Signaling 9733), BRG1 (Cell Signaling 49360), BRM (Cell Signaling 11966).

CUT&RUN Data Analysis

Paired end reads were mapped to GRCh38/hg38 using bwa (version 0.7.12). Read counts in genomic windows around DHSs were quantified using a custom script. For SMARCA2/SMARCA4 CUT&RUN, peaks were identified using MACS2 (version 2.2.1) with a control experiment against SMARCA4 in A549 ($SMARCA4^{-/-}$) cells used as the background.

Comparison of DNaseI and RNA-Seq data

Aggregate gene level score of change in accessibility: To quantify the relationship between changes in accessibility and change in gene expression, we first created an aggregate score to incorporate information about the DHSs distance from a gene's TSS (Gencode v25 basic) to weight the potential contribution of each DHSs to a gene's change in expression. Only DHSs greater than 1kb from a gene's TSS were included to isolate the effects of distal element remodeling on gene expression.

$$DHS\_Score_{Gene_i} = \sum_{d_j > 1kb}^{d_j < 1Mb} \frac{1}{d_j^{\lambda_{other} * I_{other,ij} + \lambda_{closest} * I_{closest,ij}}} * LFC_j$$

where

d_j = distance of DHS_j to the TSS of $Gene_i$

LFC_j = Log2 Fold Change of DHS_j

$I_{closest,ij}$ = Indicator if the TSS of $gene_i$ is the closest TSS to DHS_i

$I_{other,ij}$ = Indicator if the TSS of $gene_i$ is not the closest TSS to DHS_i

$\lambda_{closest}$ = decay rate if the TSS of $gene_i$ is the closest TSS to DHS_i

λ_{other} = decay rate if the TSS of $gene_i$ is not the closest TSS to DHS_i

We used different decay rates for DHSs closest to the gene's TSS and all other genes and selected the two decay rates to maximize the correlation (minimize the squared error) between the aggregate gene level score and genes change in expression (Log2FC). Regression on promoter state: To analyze the relationship between promoter state and changes in gene

expression, we fit the model

$$
\begin{aligned}
RNA\_Log2FC &= DHS\_Score + Promoter\_State + DHS\_Score : Promoter\_State \\
Promoter\_State &= I_{H3K4me3} + I_{H3K27me3} + I_{Bivalent} + \\
&\quad I_{CpG\ Island} + C_{methylation} + I_{CpG\ Island} * C_{methylation} \\
I_{Feature} &= 1 \text{ if TSS +/- 1kb overlaps a peak of the feature, 0 otherwise} \\
C_{methylation} &= \begin{cases} -1, & \text{if methylation} < \text{33rd percentile for non CpG island genes} \\ 0, & \text{if 33rd percentile} < \text{methylation} < \text{66th percentile ''} \\ 1, & \text{if methylation>66th percentile ''} \end{cases}
\end{aligned}
$$

The promoter state included indicator variables for overlap with histone ChIP-Seq peaks and discretized the average fraction of CpG methylation at the promoter into low, medium, and high values based on the methylation tertiles in non-CpG promoters (33% and 74% methylated). For genes with multiple TSSs, the TSS with the largest number of differential DHSs +/- 5kb was selected. Ties were broken by selecting the TSS with the greatest accessibility in SMARCA4$^{-/-}$ A549s.

Regional Changes in Chromatin Accessibility

Ripley K

Ripley K at multiple scales d was calculated using the edge corrected formula $K = \lambda^{-1} \sum_{DHS_i} \sum_{DHS_j \neq i} w_i * I(distance_{ij} < d)/N$ where λ is the DHS density on the chromosome, wi is the inverse of the fraction of the window +/- d basepairs from DHSi that lies within the chromosome, I(distance$_{ij}$<d) is an indicator variable that DHSj is within d basepairs of DHSi, and N is the number of DHSs on the chromosome (Ripley 1976). To compare clustering of SMARCA4 responsive DHSs to all DHSs we subtracted the Ripley K calculated from all DHSs from Ripley K calculated from SMARCA4 responsive DHSs with increasing accessibility. DHSs. We then estimated a background distribution for this quantity by block bootstrapping with blocks of 5 consecutive DHSs.

HMM

To identify chromatin domains sensitive to SMARCA4 rescue, we fit a 3 state HMM the standardized DHS log fold changes (LFC/LFC_SE). Emissions from each state were modeled by normal distributions. Transitions between neighboring DHS_x in state i and DHS_y in state j were modeled by an exponentially decaying function $T_{xy} = \pi_j + A_{i,j} * e^{-\lambda_i * d_{xy}}$ where π_j is the starting probability of state j, A_ij is a constant, λ is a decay rate constant, and d_{xy} is the distance between the DHSs. Parameters were fit using the Baum-Welch (EM) algorithm. For the transition parameters A and λ, since no closed form solution for the maximum likelihood estimate exists, at each iteration the MLE of the parameters was found by optimization using coordinate descent. After fitting the parameters, the most probable state for each DHS was determined using the Viterbi algorithm and neighboring DHSs with the same state were merged into regions. For each region, a total score was calculated as the combined z-score $\frac{\sum Z-scores}{\sqrt{Peak\ Number}}$ across the region. We then performed block bootstrapping with blocks of 5 neighboring DHSs followed by identifying and scoring regions from the randomized sequences to estimate a null score distribution. A regional score threshold was then selected to obtain regions with significantly higher (5% FDR) scores than the background distribution.

Hi-C comparisons

Hi-C data from two A549 experiments was obtained from the ENCODE portal (http://www.encodeproject.org/). TAD file: ENCFF716CFF.
Chromatin Loop File: ENCFF803ZOW. A549 Hi-C heatmaps were visualized using the 3D Genome Browser (Y. Wang et al. 2018). Chromatin loops were visualized using the WashU epigenome browser(Xin Zhou et al. 2011). Super Enhancer Comparisons: Clusters of SMARCA4 peaks (SMARCA4 super enhancers) were identified using the ROSE algorithm(Lovén et al. 2013; Warren A. Whyte et al. 2013). SMARCA4 super enhancers were linked to the closest TSS for comparison of gene expression changes.

GO Terms

GO analysis of genes whose promoters (TSS +/- 1kb) fall in the SMARCA4 sensitive regions was performed using the Panther web tool with default parameters and GO biological process complete as the annotation set(Muruganujan et al. 2016).

ENCODE Datasets

TF ChIP-Seq: All available ChIP-Seq data in A549 aligned to GRCh38 was downloaded from the ENCODE portal (http://www.encodeproject.org/). To analyze the fraction of novel SMARCA4 dependent DHSs overlapping pre-existing TF ChIP-Seq peaks, the list was manually curated to remove non transcription factor targets and a single experiment was chosen for each TF (Supplementary Table #4). For detailed analysis of specific TFs, replicate concordant peaks were used: JUNB (ENCFF565QYS and ENCFF683JTQ), CTCF (ENCFF465EGH, ENCFF531OAI, and ENCFF751UOX), RAD21(ENCFF345LNM, ENCFF512QEA, and ENCFF178CSM), and SMC3(ENCFF046RJH, ENCFF321GIF, and ENCFF922EYU). Methylation: A549 bisulfite sequencing datasets were used to analyze promoter methylation (ENCFF005TID, ENCFF003JVR). Histone ChIP: H3K4me3 (ENCFF428UWO, ENCFF643FMK, ENCFF973TUQ) and H3K9me3 (ENCFF593SUF, ENCFF601UWZ, ENCFF801BLX) histone ChIP-Seq data were downloaded from the ENCODE portal.

TCGA Data

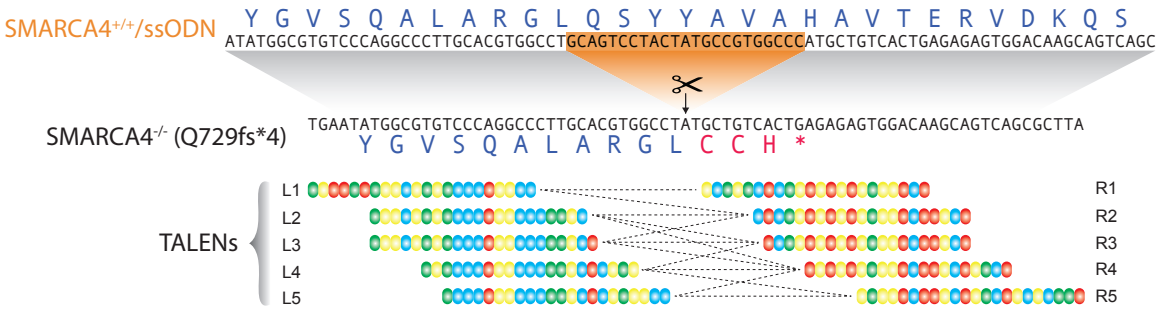
Identification of SMARCA4 gene signature: Mutation, expression, and patient data for TCGA lung adenocarcinoma samples were obtained from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/). Samples with an annotated Nonsense_Mutation or Frame_Shift_Del mutation in SMARCA4 were classified as SMARCA4 null samples. Samples with other mutations (annotated as Missense_Mutation or Splice_Site) in SMARCA4 or expression level of SMARCA4 below the average expression in SMARCA4 null samples were not included in the analysis. Expression data was log transformed (with 1 added to the FPKM value) and compared between SMARCA4 null and wild type (WT) samples. Protein coding genes significantly downregulated in SMARCA4 samples were identified by Welch's t-test (BH FDR < 0.01 and —log fold change—>log2[1.5]). Genes with low expression (mean expression < 1 FPKM in both SMARCA4 null and WT samples) were filtered from the gene signature. Stable vs Reactivated genes: Genes in the TCGA SMARCA4 signature that were significantly upregulated after SMARCA4 rescue were classified as reactivated while those that were not upregulated were classified as stable. Genes with zero mapped reads in either SMARCA4$^{-/-}$ or SMARCA4$^{+/+}$ A549s were not included in the stable gene set since the genes were often members of highly duplicated gene families and likely not detected for technical reasons (i.e. mappability). Survival analysis: Expression data from the SMARCA WT samples was log transformed and a Cox regression was fit between expression and patient

survival for each protein coding gene. Patient age and gender were included as additional variables in the regression. For comparisons between reactivated genes and all upregulated genes, only upregulated genes expressed highly enough in TCGA samples to be included in the SMARCA4 gene signature were included.

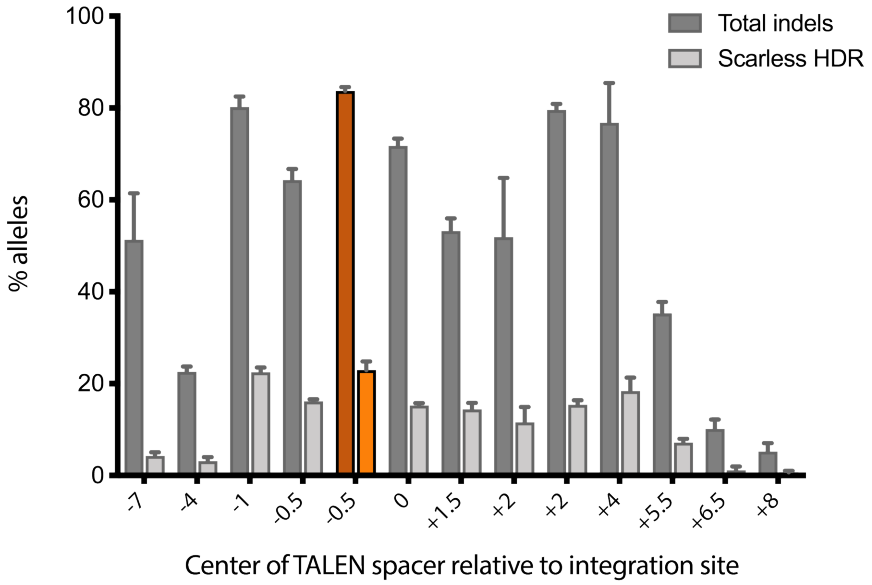
Data Access

Data from this study is available at NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm .nih.gov/geo/), accession number GSEXXXX.

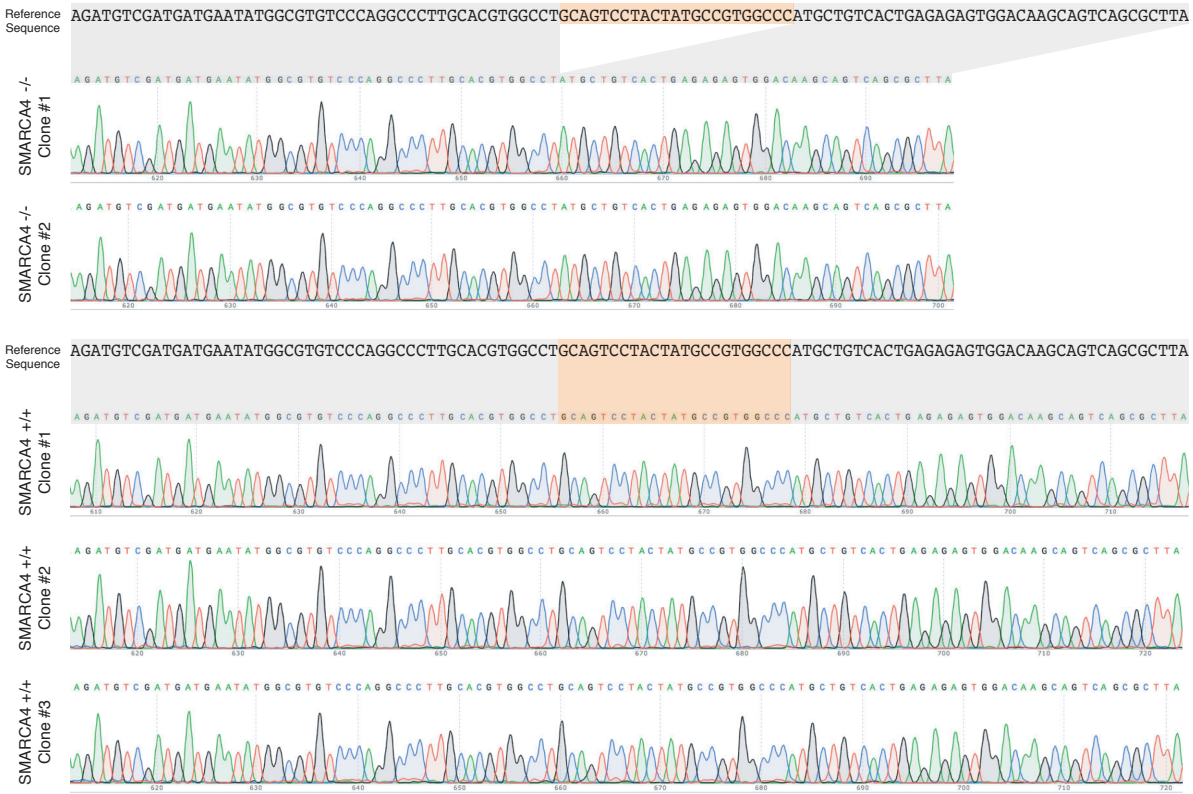
4.5 Supplement



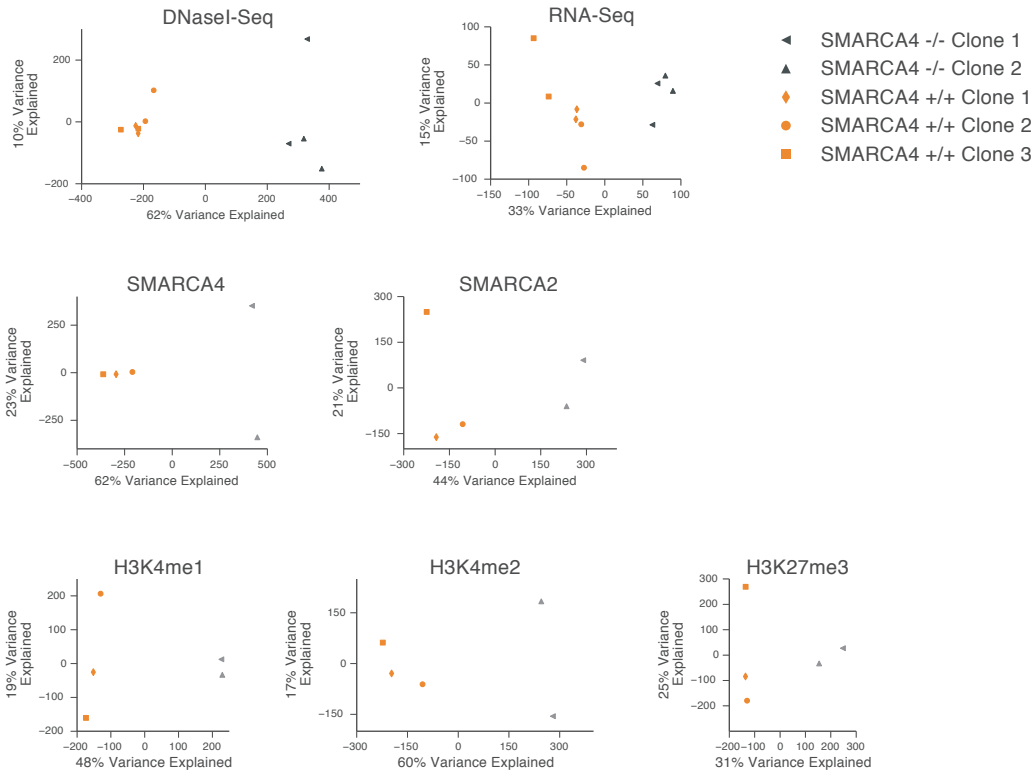
Supplement #1-Design of TALENS targeting the SMARCA4 locus



Supplement #2-HDR rates of TALENs targeting the SMARCA4 locus

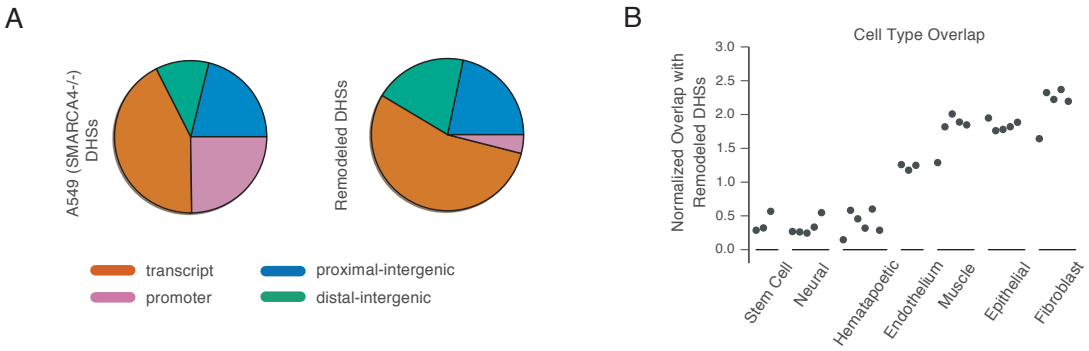


Supplement #3-Confirmation of genotypes by Sanger Sequencing



Supplement #4-Principal component analysis of genomic datasets from SMARCA4$^{+/+}$ and SMARCA4$^{-/-}$ A549 clones.

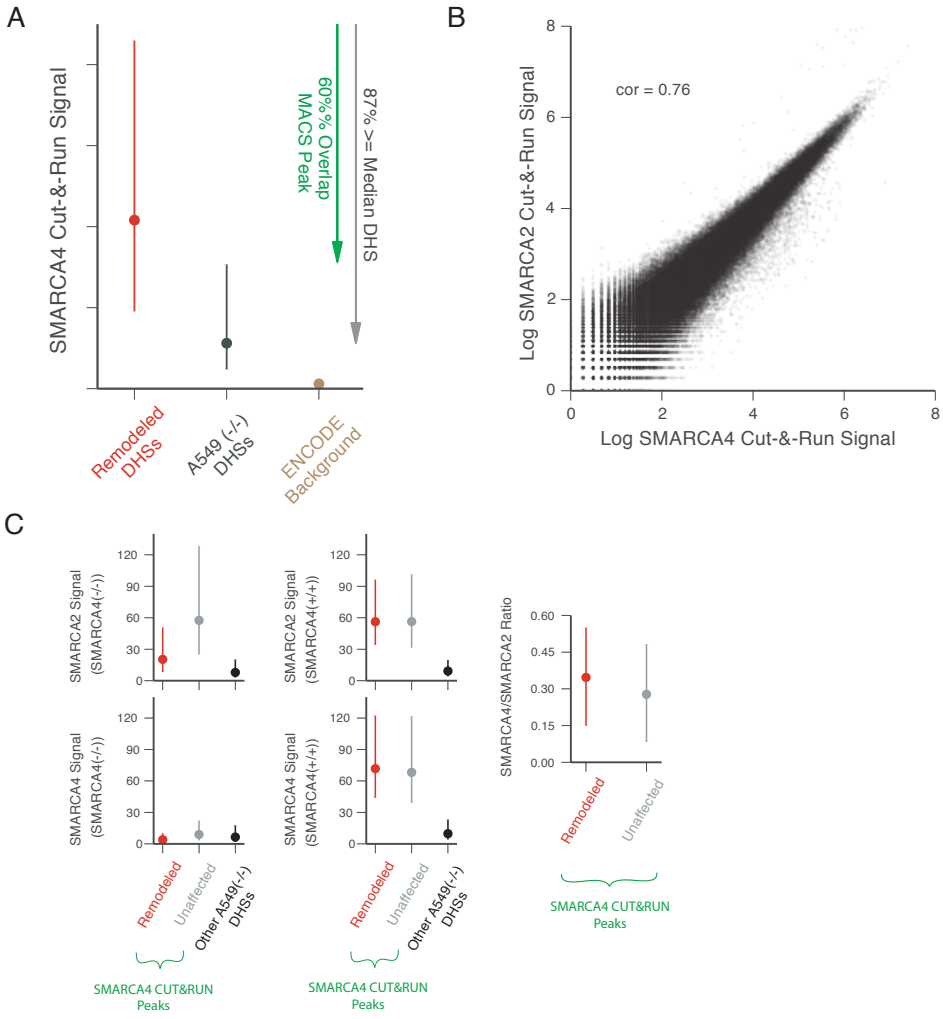
Scatterplots display reduced dimensionality representations (PCA) of data from genome-wide assays performed on SMARCA4$^{+/+}$ and SMARCA4$^{-/-}$ clones. For RNA-Seq, PCA is based on normalized count for all genes. For all other assays, PCA is based on normalized counts around A549 DHSs. All genomic assays show clear separation between SMARCA4$^{+/+}$ samples and SMARCA4$^{-/-}$ samples in the first principle component. For RNA-Seq/DNaseI-Seq, multiple data points for the same clone represent technical (independent cultures) replicates.



Supplement #5-Annotation of SMARCA4 responsive DHS

(a) Pie charts showing the fraction of DHSs overlapping GENCODE transcript annotations for all A549 DHSs (right) and sites of increased accessibility in SMARCA4$^{-/-}$ samples (left).

(b) Overlap of remodeled DHSs with cell type specific DHSs from relevant cell types. Plotted values are normalized by the fraction of SMARCA4$^{-/-}$ DHSs overlapping the cells' DHSs. Cell type specific DHSs are defined as DHSs active in a sample and present in <90% of all ENCODE samples.

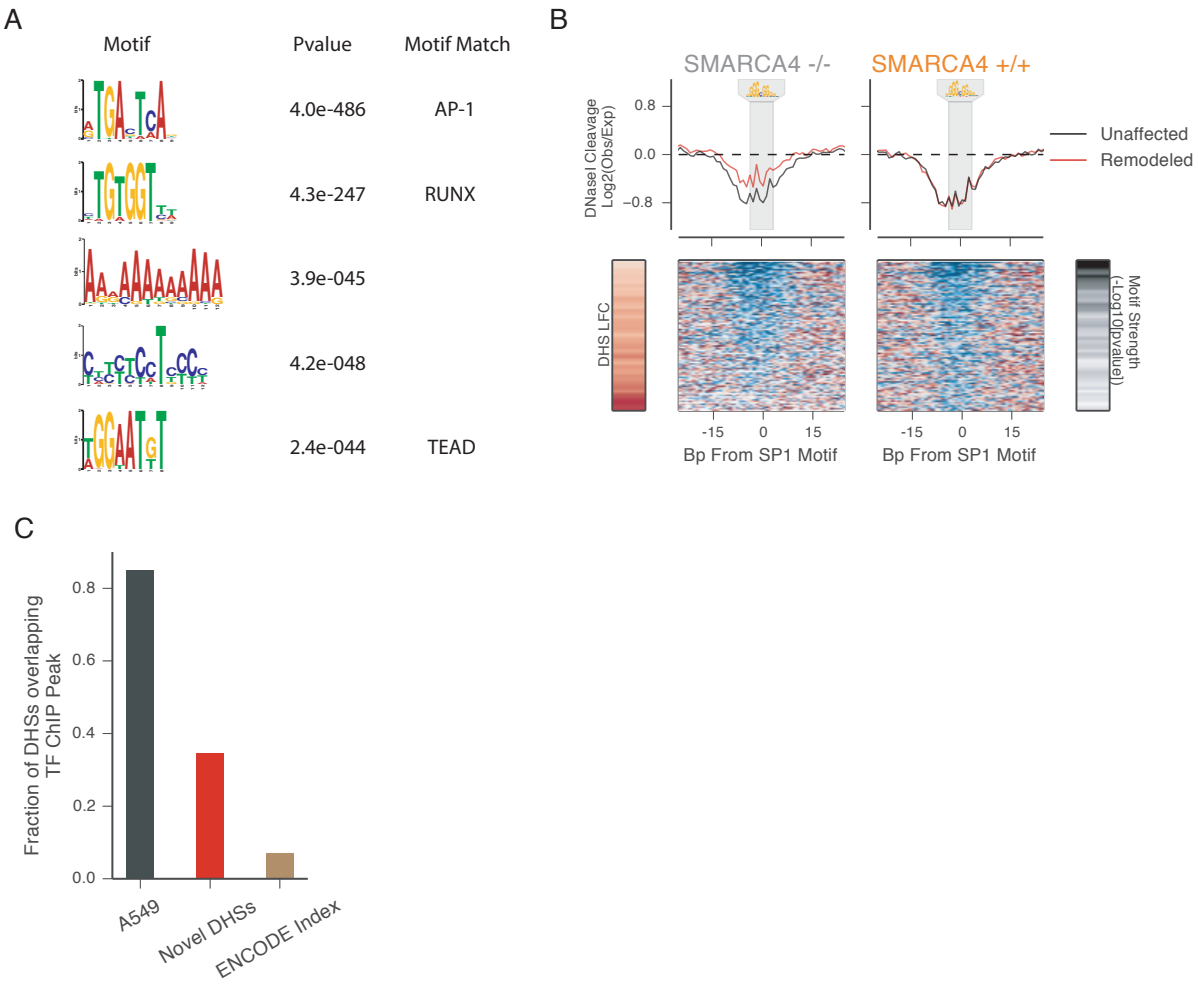


Supplement #6-SMARCA4/SMARCA4 Occupancy at Remodeled DHSs

(a) Quantitative SMARCA4 CUT&RUN signal in SMARCA4$^{+/+}$ samples at activated DHSs, unchanged A549 DHSs, and a background of DHSs active in other cell types. Median signal is plotted with error bars displaying 25th-75th percentile.

(b) Genome wide correlation of SMARCA2/SMARCA4 occupancy at DHSs. Scatterplot of SMARCA2 CUT&RUN signal versus SMARCA4 CUT&RUN signal in SMARCA4$^{+/+}$ samples at A549 DHSs.

(c) Relative SMARCA2 and SMARCA4 binding at SMARCA4 sensitive DHSs. Plotted is the median SMARCA2 CUT&RUN signal in SMARCA4$^{-/-}$ cells (top left), SMARCA2 CUT&RUN signal in SMARCA4$^{+/+}$ cells (top middle), SMARCA4 CUT&RUN signal in SMARCA4$^{-/-}$ cells (bottom left), SMARCA4 CUT&RUN signal in SMARCA4$^{+/+}$ cells (bottom middle) and the ratio of SMARCA4/SMARCA2 CUT&RUN signal in SMARCA4$^{+/+}$ cells (right) for SMARCA4 sensitive and unaffected DHSs that overlapSMARCA4 CUT&RUN peaks. Error bars show 25-75th percentiles.

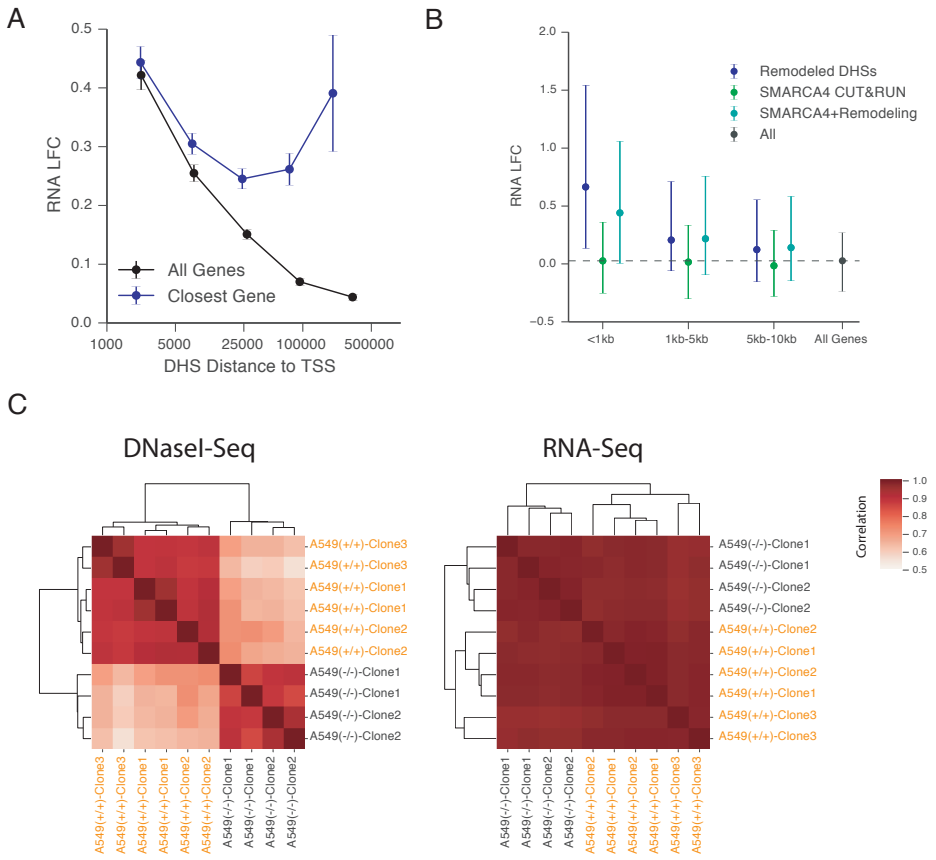


Supplement #7-TF binding in SMARCA4 responsive DHSs

(a) Significant motifs identified by de novo motif search.

(b) Top: Aggregate cleavage profile (log2 observed/expected) around SP1 motifs in SMARCA4 sensitive and unchanging DHSs in SMARCA4$^{-/-}$ and SMARCA4$^{+/+}$ samples. Motifs were selected from DHSs that overlap a SMARCA4 CUT&RUN peak and SP1 ChIP-Seq peak and are footprinted in either SMARCA4$^{-/-}$ and SMARCA4$^{+/+}$ cells. Bottom: DHSs were ordered by ordered by motif footprint occupancy in SMARCA4$^{-/-}$ samples. From left to right heatmaps display the average log fold change in accessibility after SMARCA4 reactivation, DNaseI cleavages around SP1 motif in SMARCA4$^{-/-}$ cells, DNaseI cleavages around SP1 motif in SMARCA4$^{+/+}$ cells, and AP1 motif strength. Ordering of DHSs (rows) is consistent across heatmaps. To highlight the trend with footprint occupancy, for heatmaps displaying DHS log fold change and motif pvalues, DHSs were separated into 25 bins and average values for the bin are shown.

(c) Overlap of A549 DHSs, SMARCA4 dependent DHSs, and DHSs from other cell types with ChIP-Seq peaks in unedited A549.

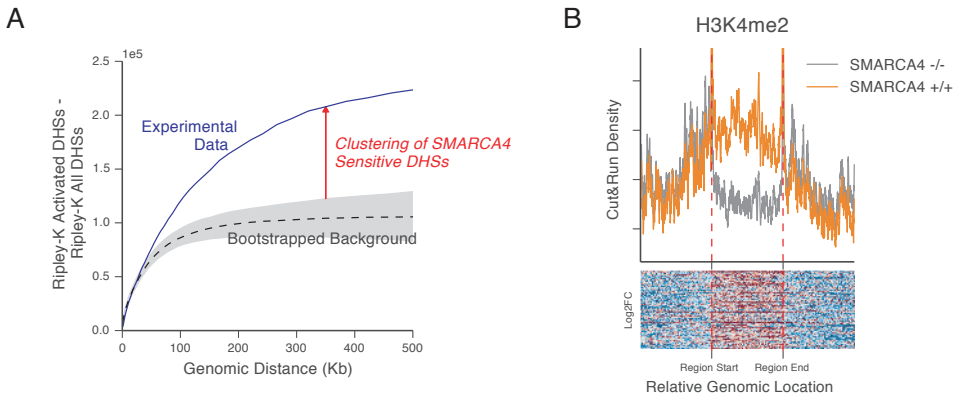


Supplement #8-Comparison of changes in chromatin accessibility and expression after SMARCA4 reactivation

(a) Relationship between gene expression and DHSs as a function of distance between a gene's TSS and the DHS. Plotted is the average change in gene expression for all genes within a given genome distance of a SMARCA4 responsive DHS (black) or genes within a given genomic distance of a SMARCA4 responsive DHS only if it is the closest gene (blue).

(b) Median change in expression for genes neighboring DHSs with increasing accessibility, SMARCA4 CUT&RUN peaks, or both. Error bars show 25th-75th percentile.

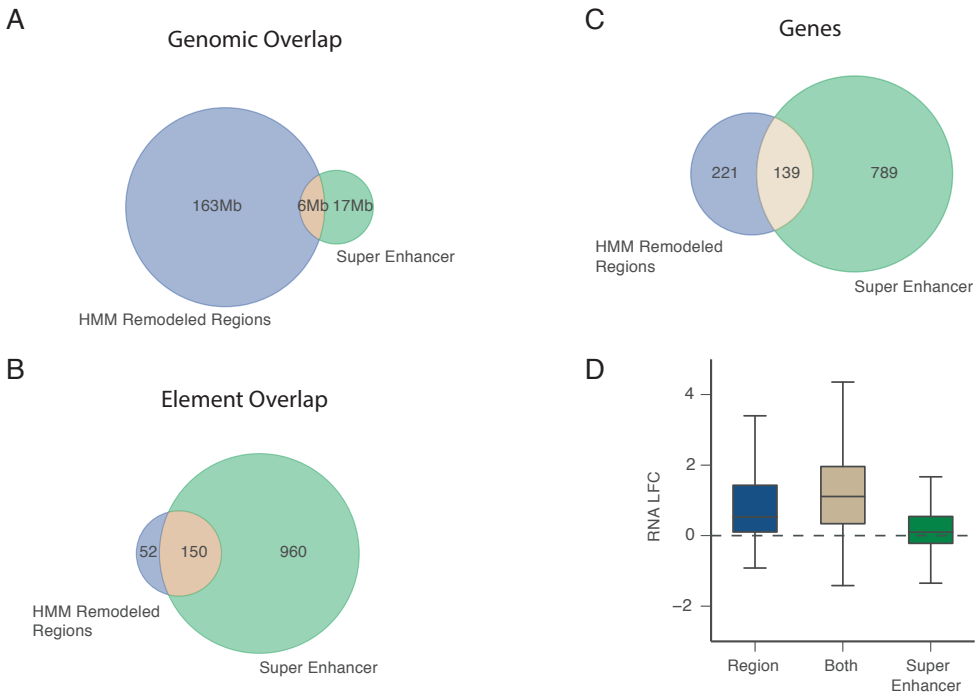
(c) Heatmaps of pairwise correlation between DNaseI and RNA samples. DNaseI and RNA are plotted on the same color scale.



Supplement #9-Genomic clustering of SMARCA4 responsive DHSs

(a) Linear ripley K (a measure of spatial clustering) of increasing DHSs vs all DHSs in the master list. Shaded region represents null distribution +/- 95% confidence interval based on 1000 block permutations of the increasing DHSs.

(b) H3K4me2 signal around identified remodeled regions in SMARCA4$^{-/-}$ and SMARCA4$^{+/+}$ clones. Top: Lineplots of the aggregate (trimmed mean) score over all regions. Bottom: heatmaps of log fold change values for individual regions.



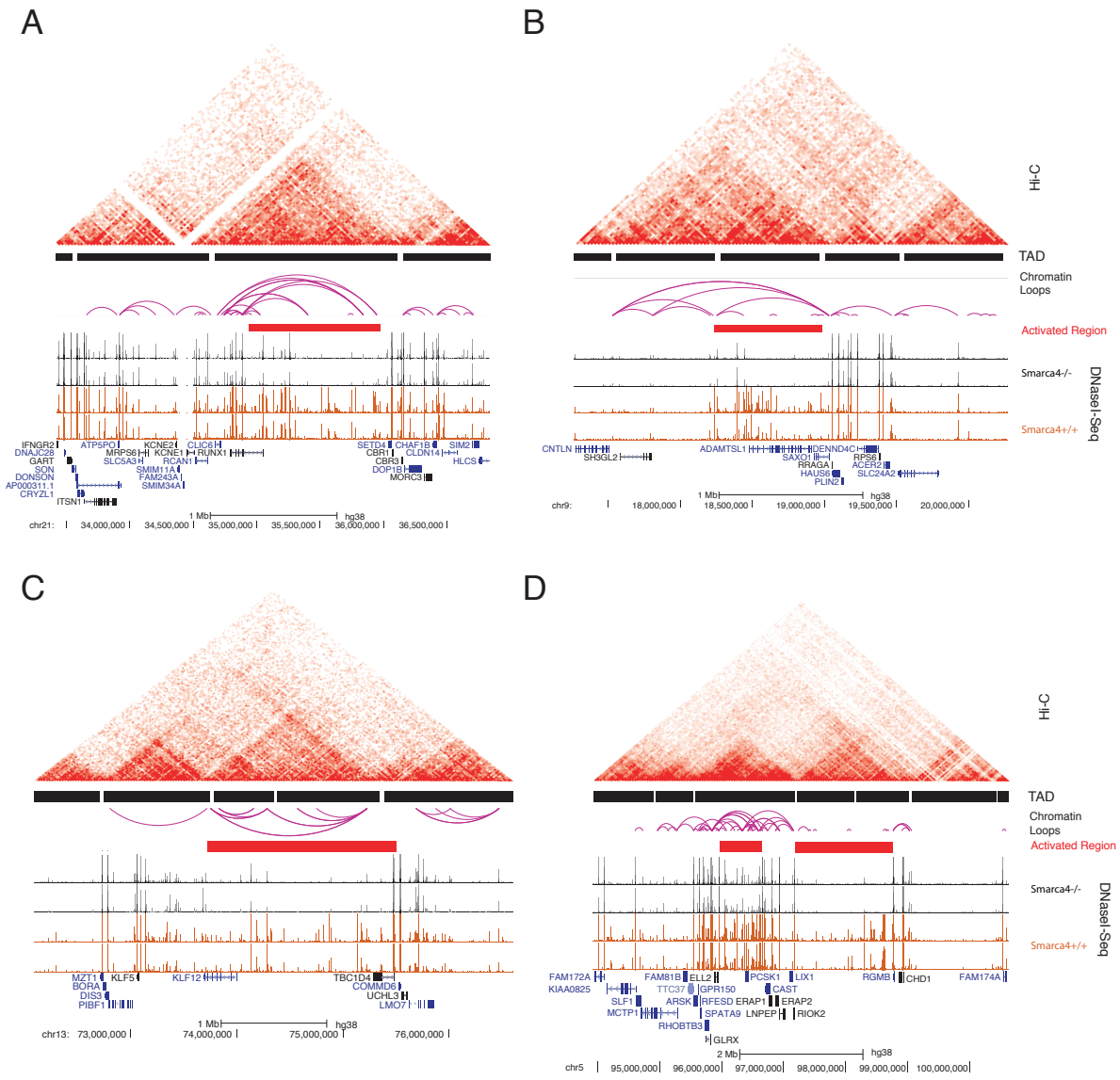
Supplement #10-Comparison of remodeled regions to SMARCA4 super-enhancers

(a) Overlap of SMARCA4 sensitive regions with SMARCA4 super-enhancers identified by the ROSE algorithm by element number.

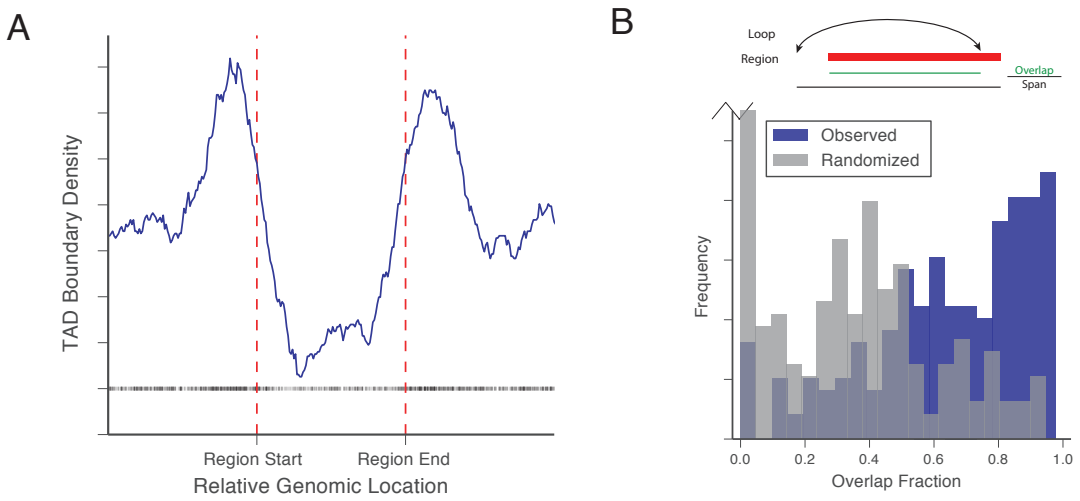
(b) As (a) by genomic region.

(c) Overlap of genes in SMARCA4 remodeled regions with the genes closest to SMARCA4 super-enhancers

(d) Boxplots of change in expression of genes linked to SMARCA4 activated regions, SMARCA4 super enhancers or both.



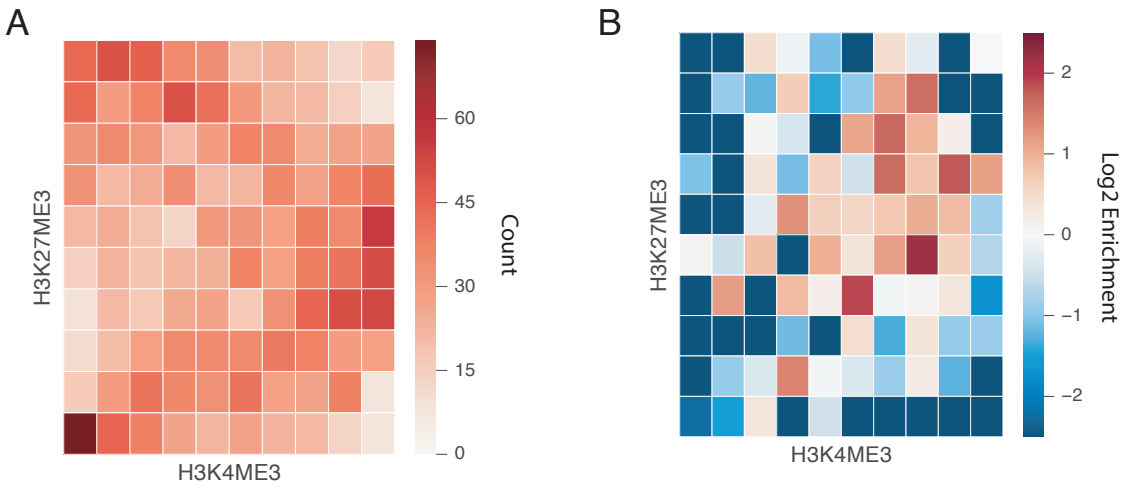
Supplement #11-Regional activation of chromatin accessibility aligns with contact domains
A-D: Hi-C signal, TAD annotations, chromatin loops, and DNaseI cleavage density at example loci identified as regional changes in chromatin accessibility.



Supplement #12-Enrichment of chromatin loops at SMARCA4 activated regions

(a) Density of TAD boundaries relative to the identified regional increases in accessibility.

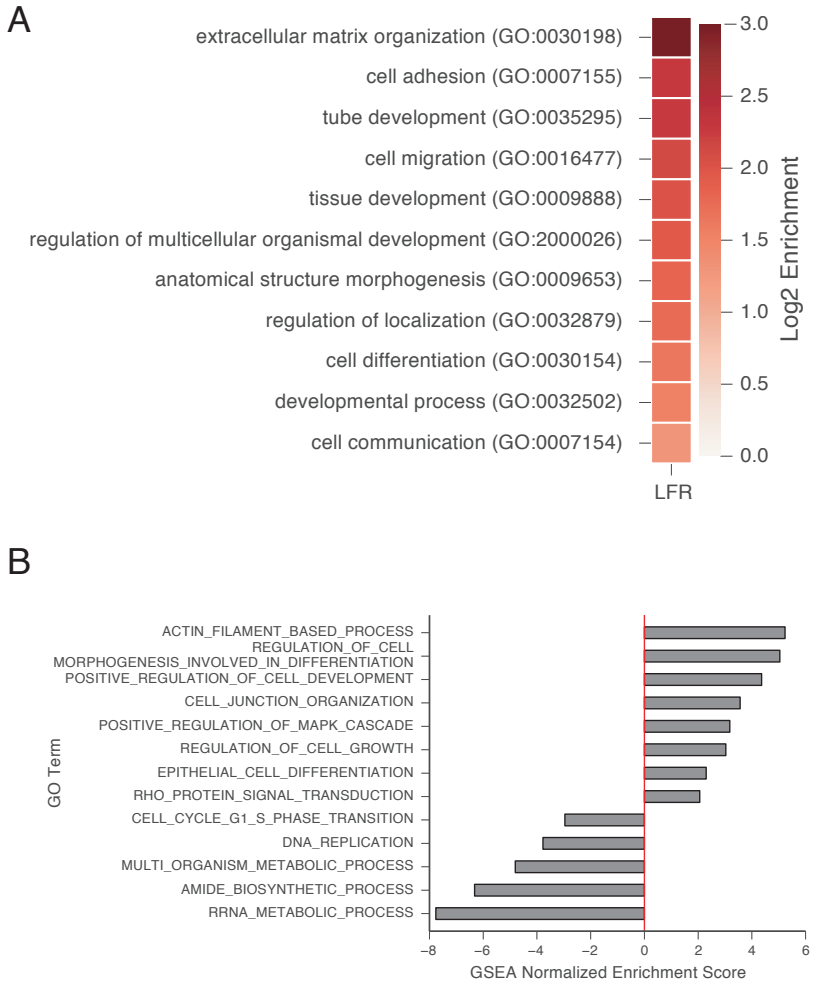
(b) Overlap between Hi-C chromatin loops and regional changes in chromatin accessibility. Histogram of the maximum fraction of genomic overlap between the region and a chromatin loop for each region (blue) with the same quantity for a background of random regions with equal numbers of DHSs (grey) for comparison.



Supplement #13-SMARCA4 remodeled domains are enriched for H3K27me3/H3K4me3 balanced domains

(a) Heatmap of joint distribution for average H3K27me3 and H3K4me3 single across TADs. TADs were binned based on the average signal each histone mark and the number of TADs in each joint bin is plotted.

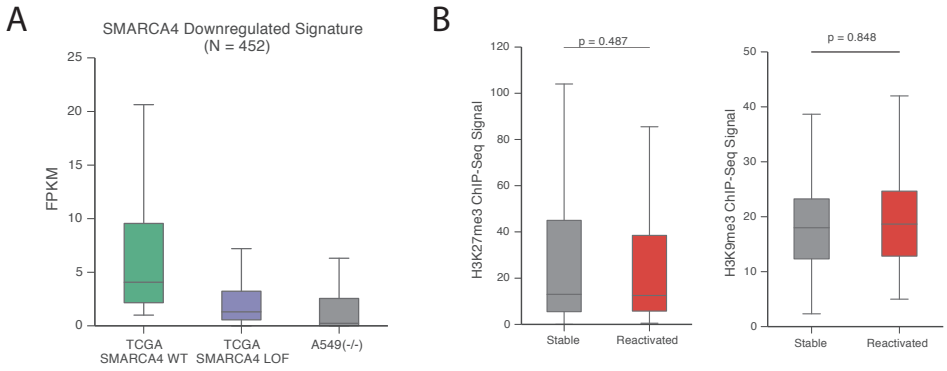
(b) Number of TADs overlapping SMARCA4 remodeled regions in each joint bin of H3K4me3/H3K27me3 signal. Plotted is Log2(observed/expected) for each bin.



Supplement #14-GO enrichment of genes in SMARCA4 remodeled regions

(a) Biological process GO terms associated with protein coding genes found in the SMARCA4 sensitive regions were analyzed to identify enriched terms. Heatmap displays enrichment of significant terms.

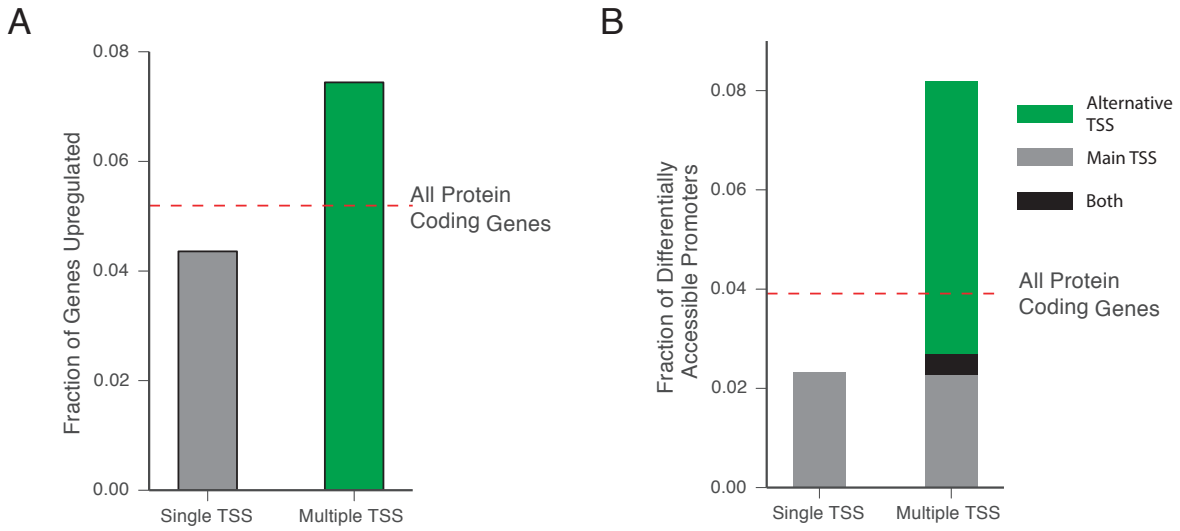
(b) Enrichment of biological process GO terms by GSEA in all (protein-coding) up- and down-regulated genes after SMARCA4 reactivation.



Supplement #15-Chromatin context of TCGA SMARCA4-null expression signature

(a) In A549 expression levels of genes with lower expression in SMARCA4 null TCGA samples genes match the SMARCA4 null tumor samples.

(b) H3K27me3 and H3K9me3 signal around the TSSs are not significantly different between reactivated and stable genes.



Supplement #16-Activation of alternative TSSs

Genes with a single annotated TSS were compared to genes with multiple distinct TSSs. To define distinct TSSs, for each gene the TSS with the highest accessibility was chosen. All TSSs falling within 1kb of the TSS were considered to be associated with that TSS/promoter while annotated TSSs ¿ 1kb were considered alternative TSSs.

(a) Barplot showing the fraction of upregulated genes after SMARCA4 reactivation for genes with a single TSS and multiple distinct TSSs.

(b) Barplot showing the fraction of genes with change in promoter accessibility after SMARCA4 reactivation for genes with a single TSS and multiple distinct TSSs.

Chapter 5: Discussion

5.1 Extensive regulatory plasticity in cancer cells

The chromatin state in cancer cells is altered extensively during oncogenesis. The changes do not follow a stereotyped trajectory; instead we observe diverse patterns of chromatin accessibility in samples of the same cancer. The differences in accessibility between samples of the same cancer are similar magnitude to the changes in accessibility seen during normal differentiation, implying cancer cells do not reflect a deterministic unwinding of the developmental clock to an earlier stage in development. Rather as cancer cells progress, they are reprogrammed to an array of possible epigenetic states. Despite the extensive regulatory diversity, in both SCLC and AML, the number of regulatory elements we discover saturates quickly. And the cancer cells maintain an underlying regulatory structure due to common regulation of many DHSs by a handful lineage specific TFs. There may, therefore, be a finite menu of regulatory programs a single cancer instance can access to drive oncogenic phenotypes or adopt to the tumor environment. However, even a relatively limited number of different regulatory programs could, through combinatorics, lead to an almost unlimited number of distinct regulatory states.

The diversity of possible epigenetic states observed for a single cancer prompts the question of the much of the regulatory plasticity observed between cancer samples might be accessed during the evolution of a single tumor. If a single tumor sample could access even a fraction of the DHSs observed in other sample of the same cancer type though epigenetic drift, the total regulatory repertoire of a single cancer could be large. In AML the repertoire might be similar in magnitude to the entire myeloid lineage or in SCLC similar to all stages of fetal brain development combined. This extensive regulatory potential could allow a cancer to access different regulatory states in response to a specific environment. Our findings suggest it may be possible for a cancer to stochastically access this regulatory variation since the sets of regulatory programs accessed by different patient samples overlap with stochastic expression observed within a single cancer line. When we performed single cell RNA-Seq on a single SCLC line, the TFs that were identified as driving variation between different cancer lines also showed elevated expression stochasticity between cells of the same cell line, and non-TF genes located near elements that were variable between different instances of the same cancer were more likely to have elevated stochasticity in expression.

It would be worthwhile to test this hypothesis, and *in vivo* models that select for metas-

tasis progression from a genetically homogeneous starting population(Morrow et al. 2018) are an interesting avenue to tackle the question. Studying the epigenetic variability that occurs after independent metastases would help establish the amount of regulatory variation possible in the absence of genetic confounders and how this epigenetic variation relates to the alternative regulatory trajectories we observe between genetically heterogeneous cancer samples.

5.2 TF affinity directs the effects TFs and chromatin remodelers to a shared set of labile sites

Regulatory changes in cancer cells occurs primarily at cell type specific regulatory elements from the cancers' lineage of origin. Notably the variability at these elements does not reflect a discrete difference between cell type specific and constitutive elements. Instead we observe a quantitative relationship between the breadth of a DHS's activity across normal development and its variability in cancer samples. The variability of a regulatory element in cancer recapitulates the developmental history of the element, as DHSs most specific to the terminal cell type are the most variable while those that are arise earlier in the cell of origin's lineage are more stable. The quantitative relationship between an element's developmental promiscuity and its plasticity during cancer development points towards a role for the intrinsic stability of each element in response to differences in a cell's trans regulatory environment. The inherent tradeoffs between stability and specificity required to activate an element in a highly specific cellular context may make the element more susceptible to disruption due to the variations in the trans regulatory environment that occur during oncogenesis.

Our other findings also support a role for the intrinsic stability of regulatory elements in determining their response to regulatory perturbations. Considering the binding sites (as defined by ChIP-Seq peaks) of a lineage specific TF, we find only a subset of binding sites appear sensitive to differences in TF expression. The sensitive sites are more cell type restricted, contain low affinity binding motifs, and appear less buffered by the binding of other TFs due to their overall lower motif density. When we apply an unrelated perturbation—LSD1 inhibition—sites that are most sensitive to changing TF dosage preferentially respond. This applies not only to TFs that directly recruits LSD1 (in which case the effect could be explained as a response only at functional binding events), but also sites that are sensitive to the expression of unrelated TFs. We interpret this shared sensitivity of certain regulatory elements to multiple distinct regulatory factors as pointing towards a role for the inherent (in)stability of the element in providing selectivity.

The role for TF affinity and element stability in jointly specifying the regulatory activity of TFs and chromatin remodelers can be explained by a model of nucleosome mediated cooperativity(Mirny 2010). To activate a regulatory element, the combined binding energy of all TFs and chromatin remodelers must be large enough to overcome the binding energy of a nucleosome bound to the DNA. High affinity binding or binding of numerous TFs can evict a nucleosome independent of the activity of a single TF or chromatin remodeler. Only when nucleosome-bound DNA and TF-bound DNA are in close equilibrium will the regulatory site be sensitive either to the dosage of the TF or the activity of a chromatin remodelers.

We observe some additional support for a general link between element stability, TF binding affinity, and chromatin remodeling in our work on SMARCA4. In lung adenocarcinoma, the subset of SMARCA4 binding sites sensitive to SMARCA4 rescue contain motifs with lower affinity for the bound factors and are have lower TF occupancy in the absence of chromatin remodeling by SMARCA4. However, future work to directly test if different chromatin remodelers affect the same set of elements in the same cellular context, for example testing the overlap between the chromatin response to LSD1 inhibition and the chromatin response to bromodomain inhibition or DNA demethylating agents in SCLC, would help confirm the phenomenon.

In our data, the preferential response of low affinity sites explains an additional layer of regulatory specificity to chromatin remodelers beyond that provided by factor recruitment. Both LSD1 and SMARCA4 affect chromatin at only a fraction of sites where they are recruited. For SMARCA4, we show the effect of SMARCA4 binding on gene expression is conditional on this remodeling. Since the low affinity sites the factors are able to remodel are more cell type specific, the relative stability of different classes of elements targets the effects to chromatin remodelers to developmental programs. The preferential response of more labile regulatory elements also helps explain the distinct of the same chromatin remodeler in different cell types. In each cell type the low affinity sites are enriched for binding sites of the lineage specifying TFs active in that cellular context. A preferential response at low affinity sites therefore targets the effects of a chromatin remodeler to the most cell type specific regulatory programs even if the factor were recruited uniformly to all sites.

5.3 Expansion of low affinity sites in cancer cells

Relative to differentiated cell types, cancer cells, like stem cells, have permissive chromatin with greater chromatin accessibility and lower levels of heterochromatin(Carone and Lawrence 2013). Our measurements of chromatin accessibility in cancer cell types suggest

the ectopic activation of a single TF or small number of TFs may maintain this global permissive state. Compared to their normal counterparts, in cancer cells we observe chromatin accessibility dramatically expands around specific lineage specific TFs resulting in activation of elements from other stages of development and cancer specific elements at the highest levels of TF expression. This feature of chromatin in cancer cells contrasts with the chromatin state of stem cells where the permissive environment is actively maintained by chromatin remodelers(Gaspar-Maia et al. 2009) and hypersensitive sites show lower enrichment of the motifs of TFs from multiple lineages. Therefore we propose the permissive chromatin states of cancer cells and normal cell types may be convergent phenotypes maintained by alternative regulatory mechanisms.

Since the expansion of chromatin accessibility in cancer cells appears to be specified by a relatively small number of TFs, it follows that the chromatin state of the cell might be especially sensitive to the disruption of these TFs. Given the effects of many chromatin remodelers converge on lineage specifying TF, elevated dependence on the expression of a small number of lineage specifying TFs could be one source of the preferential sensitivity to epigenetic therapies observed for many cancers. If this hypothesis is correct, we would expect to see cancers where a single lineage specifying TF is epigenetically driven to high expression (in contrast to being overexpression due to amplification of the DNA sequence which would likely be a more stable mechanism of overexpression) to be preferentially affected by epigenetic therapies.

The expansion of low affinity sites during cancer development might also alter the targets of chromatin remodelers in the metastatic context and help explain opposing behaviors of the same chromatin remodeler at different stages of cancer development. In mouse models of both lung adenocarcinoma and pancreatic carcinoma, BAF acts as a tumor suppressor early in cancer development but maintains oncogenic programs during the metastatic phase(Glaros et al. 2008; Roy et al. 2015). Our data, as well as studies of other cancer models(Denny et al. 2016; Morrow et al. 2018; Roe et al. 2017), suggest gain of accessibility at low affinity sites due to upregulation of lineage specific TFs is common at later stages of cancer progression. Under the framework of nucleosome competition, the increased TF expression driving the expansion of accessible chromatin would stabilize "normal" BAF targets while the new, lower affinity sites the TF activated would be relatively lowly occupied and therefore more dependent on BAF activity[1]. In this way the effect of BAF would shift from maintaining enhancers found in the normal cell type to maintaining cancer specific enhancers, and therefore the complex

[1]The idea that changing the TF occupancy can alter a sites dependence on BAF has some direct experimental support as increasing the binding strength of a motif can decrease the dependence of gene expression of SWI/SNF activity(Burns and Peterson 1997).

might shift from tumor suppressive to pro-oncogenic behavior.

If this model of BAF behavior at different cancer stages is correct, loss of function mutations in BAF should not co-occur with amplification or overexpression of lineage specific TFs. In lung cancer, NKX2-1 is recurrently amplified and we observe elevated NKX2-1 expression is associated with expansion of chromatin accessibility at novel cancer specific elements. SMARCA4 is mutated in ~10% of lung adenocarcinomas, but none of the TCGA samples with elevated NKX2-1 expression (fold change ¿2 versus normal lung tissue) have mutations in SMARCA4 or lose SMARCA4 expression. Although this is anecdotal, a more comprehensive study of the relationship between lineage specific TF amplification and loss of BAF activity across cancers may be informative.

Although this model of BAF activity is admittedly speculative, laying out the model in detail highlights how nucleosome mediated cooperatively might lead to predictable, but somewhat counterintuitive, interactions between regulatory factors and chromatin remodelers. These interacting effects can occur without needing to invoke direct interactions between the factors or additional mechanisms of cooperativity.

5.4 Potential for selection on epigenetic variation of regulatory elements?

If trends from normal development hold, mis-expression of a handful of critical genes may underlie may oncogenic phenotypes of mutations in the chromatin machinery. Given the widespread dysregulation of chromatin accessibility we identify in response to perturbing chromatin remodelers, it will likely be difficult to predict key downstream target genes. Integrating data on the chromatin changes with patterns of somatic mutations in regulatory elements may help identify important genes by highlighting elements affected by both genetic and epigenetic mechanisms. In general, however, the search for high impact non-coding variants in cancer has been challenging. The average effect size of regulatory mutations is smaller than for coding variants, so the signatures of selection are smaller. Additionally, interactions between regulatory factors and the DNA repair machinery(Polak et al. 2015) alters the mutation rate at regulatory elements, potentially confounding the search for recurrent mutations in regulatory elements. Caution is therefore warranted, especially when identifying patterns of regulatory mutations by aggregating over many sites.

Additionally, it is not known whether the effect of altering regulatory machinery in cancer deterministically affects the regulation of certain essential genes or if the effects can be

mediated by selection on regulatory states arising after epigenetic drift. For example, the increase in low affinity sites by lineage specific TFs may drive cancer progression through the upregulation of a specific target gene. Alternatively, if these less buffered elements vary stochastically, they could activate different sets of genes in different random states. If a gene that promoted survival was stochastically upregulated and the elevated expression was inherited epigenetically, that epigenetic state could be driven to predominate through selection. In this model, global chromatin dysregulation might act in an analogous manner to mutations in the DNA repair machinery, causing an overall increase in the epigenetic mutation rate rather the dysregulating a specific gene.

At this point there is no reason to prefer the model of epigenetic selection over deterministic regulation of a few key target genes. However, in a few places our results hint at a role for stochasticity in the effects of regulatory dysfunction. As mentioned above, the labile regulatory elements, which are preferentially affected by chromatin remodelers, are associated with increased gene expression stochasticity. Additionally, when we analyze the link between survival and gene expression for all genes upregulated after SMARCA4 rescue, we find equal numbers of genes associated with increased and decreased survival. However, genes that are downregulated in SMARCA4-null patient samples are biased for associations with increased survival. The effect could be explained by different SMARCA4 targets in the oncogenic context compared to the cell of origin. But another possible explanation for the bias is SMARCA4 loss increases the probability that SMARCA4 target genes can be durably repressed. Subpopulations of cancer cells that stochastically silenced the SMARCA4 targets that slow tumor progression(and so would be associated with increased survival) would have a selective advantage and so comprise the bulk of the tumor. It therefore might be interesting to compare single cell expression data in the mutant and wild-type SMARCA4 samples or before and after overexpression of a lineage specific TF oncogene to test if these common alterations can increase the stochasticity of gene expression independently of changes in mean expression.

5.5 Methodological limitations

While the work presented here has advanced our understanding of the connection between chromatin remodelers and cancer development, the models and approaches used have some limitations that might be addressed in future experiments.

First, all experiments were conducted in cancer cell lines, which differ from cancer cells *in vivo*. Cell lines lack the tumor micro-environment, which contributes to certain cancer

phenotypes, and the process of culturing cancer cell lines may select for a subpopulation of cells with the ability to grow in certain media that is not representative of the entire cancer cell population. While these are valid limitations to cancer cell lines as a model system, cell lines do accurately capture many critical cancer phenotypes. Crucially they behave like tumor cells when transplanted into nude mice as they successfully reconstitute tumors of their derived cancers(Daniel et al. 2009). Also, the cancers studied in this work tend to form cancer cell lines at high frequency, which suggests a significant fraction of cancer cells have the potential to be cultured and the selective bottleneck during the act of culturing may not be large(Carney et al. 1985; Drexler, Matsuo, and MacLeod 2000). One possible approach to address the differences between cancer cell lines and cancer cells *in vivo* is to study chromatin changes in tumor tissue samples. However, this approach has its own limitations as cellular heterogeneity can confound the analyses of bulk tumors and single cell approaches are not, at this point, able to map chromatin states with the same resolution as bulk methods(Matthew T. Maurano and John A. Stamatoyannopoulos 2015).

The use of cancer cell lines as a model system also limits the types of phenotypes that can be studied. In this work we focus on cell growth and cell morphology, which are easily assayed in cell line models. Alterations to chromatin remodelers might also affect other hallmarks of cancer such as vascularization, migration, or immune evasion. Therefore, an important next step is to adopt in vivo models to understand the relationship between the observed changes in chromatin state and cancer phenotypes that cannot be studied *in vitro*.

In addition to the choice of model systems, the type of information provided by the regulatory assays we applied affects the questions we can address. One major challenge in applying genome-wide assays is that only a fraction of the observed effects is likely relevant to the phenotypes of interest. We therefore focused primarily on the mechanisms by which chromatin remodelers can regulate specific genes rather than focus on the relationship between the changes in gene expression and cancer phenotypes. Additionally, we primarily studied regulatory activity using chromatin accessibility, which—as a first approximation—can be thought of as a readout of total TF binding(Thurman et al. 2012). For certain regulatory events, however, changes in TF binding may not be the critical step. For example, TFs lacking activating domains complete with activating TFs for the same binding site to repress gene expression(Lemercier et al. 1998). In this situation, the level of polymerase recruitment (and therefore gene expression) changes independently from the amount of TF binding. While recruitment of activating complexes will often feed back to cause changes in chromatin accessibility due to the highly interconnected nature of chromatin regulation, in general, compared to changes in TF binding, changes in the factors that are recruited by

TFs will be relatively opaque to mapping chromatin accessibility. For chromatin modifiers, where changes in subunit composition or binding partners may be a key factor in complex specificity, understanding in more detail the composition of the complexes recruited by TFs is likely to be important. Current methods, however, are not well suited to addressing this question. Proteomics can provide some insight into the composition of chromatin complexes globally, but as of yet we are unable to study the regulatory complexes present at an individual locus in an unbiased manner.

Another limitation of the work is that several conclusions are based on correlations between different regulatory features. Specifically, the overlap between the set of elements affected by chromatin remodelers and the set of elements whose activity is strongly correlated with TF expression led us to propose the inherent plasticity of certain regulatory elements to TF dosage targets the effects of chromatin remodelers to developmental regulatory elements. While additional evidence, such as the motif content of the DHSs, supports this model, experiments to confirm a causal role for an element's sensitivity to TF expression in targeting the effects of chromatin remodelers would be valuable. Our model predicts increasing TF expression will stabilize the activity elements bound by the TF and therefore make those sites less sensitive to the activity of a chromatin remodeler. This could be directly tested by overexpressing a single TF and studying if the effect of inhibiting a chromatin modifier is changed at the higher expression levels. Such an experiment would allow us to make stronger causal statements about the relationship between the stability of a regulatory element to TF dosage the sensitivity to chromatin modifiers.

5.6 Future directions

A consistent finding in our work is a cancer's epigenetic state strongly modifies the effects of chromatin remodelers. Different chromatin environments direct the effects of the same chromatin remodelers towards highly distinct developmental or cancer-specific regulatory pathways. Therefore to fully understand the role of chromatin remodelers in cancer development, it will be important to understand how somatic mutations interact with the cell type of origin to establish oncogenic regulatory states. While in this work we inferred factors that contribute to cancers' chromatin states by studying the variation between distinct epigenetic endpoints, experiments to directly test how key driver mutations contribute to the regulatory state will be a crucial next step.

In many cancers several different chromatin remodeling complexes are mutated. These mutations are often lumped together into a single class of mutations that cause chromatin

dysfunction, but it is unknown whether these mutations might affect the same regulatory pathways or act by distinct mechanisms. Therefore, it would be useful to expand the *SMARCA4* experiments presented in Chapter #4 to other recurrently mutated complexes in the same cancer type. Specifically, the chromatin modifiers *ARID1A*, *ARID2*, and *SETD2* are recurrently mutated in addition to *SMARCA4* in lung adenocarcinoma(Network, Collisson, et al. 2014). Understanding the relationship between chromatin changes induced by different mutated complexes may allow us to understand what (if any) are the key changes in the chromatin state being selected for.

It is also not well understood how the diverse chromatin states we observe in cancer relate to non-chromatin driver mutations. If key driver mutations, such mutations in the *EGFR/KRAS* pathway in lung adenocarcinoma, cause specific regulatory states, there will be interactions between the non-chromatin driver mutations and alterations to the regulatory complexes. In this case, drugs targeting chromatin remodelers may have different effects depending on the mutational status of a cancer. Experiments directly testing the effect of driver mutations on chromatin state and the interaction between driver mutations and chemical or genetic perturbations to chromatin remodelers would help us understand the potential interactions between driver mutations and chromatin remodelers. Understanding these interactions might then allow therapies against chromatin modifiers to be targeted to the specific patient subsets most likely to benefit from the treatment.

A major question for these purposed experiments is selecting an applicable model system. As a first step, the effect of the mutations can be studied in cancer cell lines, but since cell lines represent terminal cancer states, they may not capture the role of the mutations in early cancer development. Recently, there have been attempts to partially reprogrammed cancer cells to pluripotent cells to produce embryoid bodies that model earlier stages of cancer development(Kim et al. 2013). Advances in gene editing have also made these systems tractable to the types of genetic modifications proposed. Although these models are still in their infancy, it will be interesting to see if they prove to be more broadly applicable in the future.

5.7 Concluding remarks

Currently there are limited studies comprehensively mapping the regulatory changes that characterize cancer development[2]. Our results therefore contribute to understanding

[2]DNA methylation data is available for a subset of TCGA samples. However this data is primarily from methylation arrays, which primarily measure methylation state at promoters

how regulatory programs are altered during oncogenesis. But increasingly mapping the differences between the normal and oncogenic states will not be a limiting factor. In most cancers we should expect to see extensive genome-wide changes in chromatin states. This will lead to a similar challenge as the one faced when analyzing somatic mutations, where the large number of passenger mutations obscure the limited number of driver mutations. We should analogously expect a limited set of epigenetic driver events are embedded in the genome-wide regulatory alterations.

When studying genetic mutations, identifying recurrent mutations is the key step to identifying driver mutations since recurrence provides evidence for selection. This approach may prove less effective when studying regulatory changes in cancer since many regulatory events will be incidentally correlated if they are regulated by the same set of factors. Instead it may be valuable to integrate experimental perturbations of regulatory state, such as chemical inhibition of regulatory machinery, with selection on specific phenotypes to break as much of the incidental correlation as possible and more fully understand the specific regulatory states that drive oncogenesis.

In addition to illuminating the role of epigenetic dysfunction in cancer development, studying the response of regulatory states to diverse perturbations will improve our basic understanding of gene regulation. Cells' regulatory states can be highly plastic but genome-wide studies are just beginning to capture the dynamic behavior of the epigenome. Quantifying how chromatin at regulatory elements changes dynamically in response to perturbations will add additional richness to our understanding of gene regulation, pinpoint places where our current models of gene regulation are not sufficient, and allow us to bridge the gap in our understanding between biochemical changes to chromatin state and downstream effects on gene expression.

Appendix A1 : SCLC/AML Tables

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-H3K4me1	Active Motif	AM39297 (Lot 21008001(
Rabbit polyclonal anti-H3K4me2	Active Motif	AM39141 (Lot 01008001(
Chemicals, Peptides, and Recombinant Proteins		
GSK-LSD1	Sigma-Aldrich	SML1072
Critical Commercial Assays		
MagMax mirVana Total RNA isolation kit	Thermo Fisher	Cat# A27828
TruSeq Stranded Total RNA kit	Illumina	Cat# 20020596
TruSeq DNA Library Preparation Kit v2, Set B	Illumina	Cat# FC-121-2002
Deposited Data		
AML and SCLC mutation data	CCLE	https://portals.broadinstitute.org/ccle
Raw and analyzed data	This study	GSEXXX
DHS data from diverse cell lines and tissue	ENCODE Consortium	https://www.encodeproject.org/
SPI1 ChIP-Seq in HL-60	ENCODE Consortium	ENCFF797QGP
ASCL1/NEUROD1 ChIP-Seq in SCLC	[1]	GSE69398
LSD1 ChIP-Seq in NCI-H526	[2]	GSE66298
Experimental Models: Cell Lines		
DMS-114	ATCC	CRL-2066
DMS-153	ATCC	CRL-2064
DMS-53	ATCC	CRL-2062
DMS-79	ATCC	CRL-2049
NCI-H1048	ATCC	CRL-5853
NCI-H1092	ATCC	CRL-5855
NCI-H1105	ATCC	CRL-5856
NCI-H128	ATCC	HTB-120
NCI-H1417	ATCC	CRL-5869
NCI-H1436	ATCC	CRL-5871
NCI-H146	ATCC	HTB-173
NCI-H1694	ATCC	CRL-5888
NCI-H187	ATCC	CRL-5804
NCI-H1930	ATCC	CRL-5906
NCI-H2029	ATCC	CRL-5913
NCI-H2081	ATCC	CRL-5920
NCI-H209	ATCC	HTP-172
NCI-H2107	ATCC	CRL-5983_FL
NCI-H2141	ATCC	CRL-5927
NCI-H2171	ATCC	CRL-5929

NCI-H2196	ATCC	CRL-5932
NCI-H2227	ATCC	CRL-5934
NCI-H345	ATCC	HTB-180
NCI-H446	ATCC	HTB-171
NCI-H510A	ATCC	HTB-184
NCI-H524	ATCC	CRL-5831
NCI-H526	ATCC	CRL-5811
NCI-H69	ATCC	HTB-119
NCI-H748	ATCC	CRL-5841
NCI-H774	ATCC	CRL-5842
NCI-H82	ATCC	HTB-175
NCI-H889	ATCC	CRL-5817
SHP-77	ATCC	CRL-2195
SW1271	ATCC	CRL-2177
COLO-668	Sigma Aldrich	87061209
COR-L47	Sigma Aldrich	92031915
DMS-273	Sigma Aldrich	95062830
AML 193	ATCC	CRL-9589
GF-D8	DSMZ	ACC-615
NOMO-1	DSMZ	ACC-542
SKNO-1	DSMZ	ACC-690
UT-7	DSMZ	ACC-137
KASUMI-1	ATCC	CRL-2724
MOLM-13	DSMZ	ACC-554
OCI-AML2	DSMZ	ACC-99
SIG-M5	DSMZ	ACC-468
SKM-1	DSMZ	ACC-547
GDM-1	ATCC	CRL-2627
HL-60	ATCC	CCL-240
KASUMI-3	ATCC	CRL-2725
KG-1	ATCC	CCL-246
MV-4-11	ATCC	CRL-9591
OCI-M1	DSMZ	ACC-529
OCl-M2	DSMZ	ACC-619
TF-1.CN5a.1	ATCC	CRL-2512
THP-1	ATCC	TIB-202
UCSD-AML1	DSMZ	ACC-691
Software and Algorithms		
bwa	[3]	http://bio-bwa.sourceforge.net/
RNA-STAR	[4]	https://github.com/alexdobin/STAR

hotspot-v2		https://github.com/Altius/hotspot2
DESeq2 package	[5]	
bedops	[6]	https://bedops.readthedocs.io

1. Borromeo, M. D. *et al.* ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs. *Cell Rep.* **16,** 1259–1272 (2016).
2. Mohammad, H. P. *et al.* A DNA Hypomethylation Signature Predicts Antitumor Activity of LSD1 Inhibitors in SCLC. *Cancer Cell* **28,** 57–69 (2015).
3. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
4. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
5. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. (2014).
6. Neph, S. *et al.* BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28,** 1919–1920 (2012).

The Key Resources Table includes the critical reagents and software necessary to reproduce the experiments and analyses described in Chapters #2-3.

	Mean Pearson Correlation	Cell1	Cell2
Fetal Kidney Cortex vs Pelvis	0.9132549000000005	fKidney_renal_cortex	fKidney_renal_pelvis
Islet cell differentiation	0.898898	H9_hESC	ISL1
Endoderm Differentiation	0.891131	H9_hESC	DE
T vs B cell	0.86585875	CD3	CD20
Fetal Atrium vs Ventricle	0.8653463333333334	fLeftAtrium	fLeftVentricle
Muscle Differentiation	0.86484175	HSMM	HSMM_D
Adipocyte Differentiation	0.8580099999999999	AF_MSC	AL
Fetal Stomach vs Small Intestine	0.8303913409090908	fStomach	fIntestine_Sm
Cardiac Differentiation	0.82836375	hESCT0	H7_hESC_T14
Neuronal differentiation	0.799519	H9_hESC	neuronal
NPC Differentiation	0.7828235	H9_hESC	NPC
Fetal Brain vs Fetal Spinal Cord	0.7796718487394957	fBrain	fSpinal_cord
CD34 vs CD4	0.7377563921568626	CD34	CD4
Erythroid Differentiation	0.7125263333333334	CD34_T0	CD34_T8
CD34 vs CD20	0.7096028235294118	CD34	CD20
Inter-AML	0.6311890210526314		
CD34 vs CD14	0.6227618823529414	CD34	CD14
Inter-SCLC	0.5224466141141142		

Supplementary Table #1: Developmental Correlations
Table comparing the average correlation in chromatin accessibility between different SCLC and AML samples to the correlation in chromatin accessibility between of different cell types from normal development. Note the difference in chromatin accessibility between samples of the same cancer type is similar in magnitude to changes in chromatin accessibility that occur during differentiation of normal cell types.

DS-ID	Cell Line	Drug	Time	Batch	Cancer Type	Total Reads	Nuclear Mapping Rate	Nuclear Duplicate Rate	Hotspot Number
DS36058	KG-1	None	None	0	AML	126525362	0.620563867662	4.18011788214	108350
DS35968	MV-4-11	None	None	0	AML	317587060	0.704273398293	12.9871768313	283782
DS35827	OCI-M1	None	None	0	AML	335008258	0.684306367158	11.6907300384	212478
DS35697	OCI-M2	None	None	0	AML	73453148	0.699226668951	49.5600306851	34606
DS42655	KASUMI-3	None	None	0	AML	62152430	0.596870564836	21.1579543076	44789
DS36078	HL-60	None	None	0	AML	481693550	0.457474853047	33.6164431214	199634
DS36370	THP-1	None	None	0	AML	289111328	0.571508425986	23.1846771244	243127
DS36528	TF-1-CN5A-1	None	None	0	AML	352348080	0.555963301971	30.1489404893	156097
DS36532	UCSD_AML1	None	None	0	AML	512549230	0.53089645262	31.3012716213	167480
DS37336	GDM-1	None	None	0	AML	87068296	0.549615556965	6.5324134259	80894
DS39011	AML-193	10nM	Day2	0	AML	60826120	0.625344440842	19.051019847	75510
DS42343	AML-193	10nM	Day2	1	AML	143638790	0.621196363461	25.2820446485	116752
DS42260	AML-193	10nM	Day6	1	AML	73500802	0.612273564035	4.59415254204	102549
DS39134	AML-193	10nM	Day6	0	AML	79175242	0.674782200224	2.72609179149	109683
DS42245	AML-193	None	Day6	1	AML	69055082	0.664153928599	5.52936074854	98051
DS38994	AML-193	None	Day2	0	AML	62023600	0.644799979363	11.2618251238	85716
DS42329	AML-193	None	Day2	1	AML	367798042	0.635389260718	51.5616256897	153738
DS39019	AML-193	500nM	Day2	0	AML	95024410	0.668027909881	20.1713188172	108298
DS42350	AML-193	500nM	Day2	1	AML	87620618	0.618386713502	23.3414494683	96647
DS42266	AML-193	500nM	Day6	1	AML	78079344	0.625248618892	11.8351538608	95107
DS39141	AML-193	500nM	Day6	0	AML	85701882	0.558915147278	10.2047178209	107152
DS39002	AML-193	DMSO	Day2	0	AML	76614484	0.513679658797	9.26133408911	78128
DS42336	AML-193	DMSO	Day2	1	AML	125652214	0.589831214594	27.4745964971	110523
DS42252	AML-193	DMSO	Day6	1	AML	70024132	0.630669666851	6.82805227361	99073
DS43361	GF-D8	None	Day2	1	AML	553116284	0.527929429031	29.0509654783	173111
DS42623	GF-D8	None	Day2	0	AML	71035712	0.603091019908	6.70978735324	81891
DS43388	GF-D8	None	Day6	1	AML	81037636	0.580416412937	9.0007958657	100651
DS42836	GF-D8	None	Day6	0	AML	111326456	0.469394157306	17.6213795824	80279
DS43376	GF-D8	10nM	Day2	1	AML	132572062	0.534978025762	8.56164151785	115863
DS42640	GF-D8	10nM	Day2	0	AML	86728874	0.551576029916	9.0179814325	79349
DS43403	GF-D8	10nM	Day6	1	AML	70311804	0.579903966054	2.94865411357	84903
DS42852	GF-D8	10nM	Day6	0	AML	107732554	0.448860555186	10.8917603351	93852
DS43396	GF-D8	DMSO	Day6	1	AML	83817078	0.632987348951	3.65097073517	112497
DS42843	GF-D8	DMSO	Day6	0	AML	91175844	0.475545606137	9.55126855609	60905
DS43369	GF-D8	DMSO	Day2	1	AML	103490364	0.612030700752	10.1281306008	104762
DS42631	GF-D8	DMSO	Day2	0	AML	60611796	0.57847317377	9.71444883618	72180
DS43382	GF-D8	500nM	Day2	1	AML	88320894	0.559332336468	10.2600787373	93432
DS42647	GF-D8	500nM	Day2	0	AML	85659114	0.456712662239	11.5648024843	77834
DS43411	GF-D8	500nM	Day6	1	AML	65426942	0.701965162914	4.28301742266	76047
DS42860	GF-D8	500nM	Day6	0	AML	104488614	0.505128319532	7.43087203339	96553
DS38280	NOMO-1	None	Day2	0	AML	75327358	0.726426645682	9.85371291562	134579
DS42301	NOMO-1	None	Day2	1	AML	494503702	0.719738615829	39.3518086323	242840
DS38513	NOMO-1	None	Day6	0	AML	75826446	0.730689606631	26.6481060866	118225
DS41906	NOMO-1	None	Day6	1	AML	54345940	0.77036797965	4.8977781022	125130
DS38296	NOMO-1	10nM	Day2	0	AML	87346144	0.568885445017	27.1595765341	86319
DS42315	NOMO-1	10nM	Day2	1	AML	123317682	0.732635843739	10.9009286578	162835
DS38529	NOMO-1	10nM	Day6	0	AML	74373440	0.679263538166	10.760264728	106765
DS41921	NOMO-1	10nM	Day6	1	AML	69145334	0.751201375352	5.62204779286	109180
DS38286	NOMO-1	DMSO	Day2	0	AML	69431888	0.69360499026	15.3881291168	135280
DS42308	NOMO-1	DMSO	Day2	1	AML	96835760	0.725326883374	7.69149933696	148615
DS38521	NOMO-1	DMSO	Day6	0	AML	78154322	0.689350615824	43.0415216648	96720

DS-ID	Cell Line	Drug	Time	Batch	Cancer Type	Total Reads	Nuclear Mapping Rate	Nuclear Duplicate Rate	Hotspot Number
DS41913	NOMO-1	DMSO	Day6	1	AML	53247644	0.782622006713	4.53518361555	118813
DS38538	NOMO-1	500nM	Day6	0	AML	56830742	0.691084131895	9.10254110878	103346
DS41927	NOMO-1	500nM	Day6	1	AML	90661718	0.550515202017	6.89155666043	127032
DS38303	NOMO-1	500nM	Day2	0	AML	184210342	0.401574814947	45.063378419	118632
DS42322	NOMO-1	500nM	Day2	1	AML	118206204	0.746014177056	7.59784392328	174115
DS42113	KASUMI-1	None	Day6	1	AML	118506298	0.341256141509	9.83910586587	88570
DS38640	KASUMI-1	None	Day2	0	AML	87329686	0.536229593222	4.99916055863	113160
DS42272	KASUMI-1	None	Day2	1	AML	496310672	0.4372273623	34.7054620668	188375
DS38648	KASUMI-1	DMSO	Day2	0	AML	80805108	0.524334600233	4.2366249935	106938
DS42279	KASUMI-1	DMSO	Day2	1	AML	105254638	0.399372709828	18.9178612626	105330
DS38746	KASUMI-1	DMSO	Day6	0	AML	127285876	0.574649633554	2.40788331285	131632
DS42121	KASUMI-1	DMSO	Day6	1	AML	78248366	0.514008228619	3.92527117647	91295
DS38751	KASUMI-1	10nM	Day6	0	AML	64045562	0.584776287856	8.84010782134	83550
DS42127	KASUMI-1	10nM	Day6	1	AML	226363244	0.429806448612	9.00439974838	133666
DS38656	KASUMI-1	10nM	Day2	0	AML	91901666	0.624747999672	14.2377229154	102803
DS42286	KASUMI-1	10nM	Day2	1	AML	131329066	0.461512076847	34.3162698534	107057
DS38758	KASUMI-1	500nM	Day6	0	AML	98276184	0.579823429042	13.7632291156	105192
DS42134	KASUMI-1	500nM	Day6	1	AML	95384586	0.523827235566	8.19228739196	100345
DS38664	KASUMI-1	500nM	Day2	0	AML	139592012	0.616598949802	9.3811854984	142281
DS42293	KASUMI-1	500nM	Day2	1	AML	73134348	0.550698421486	13.6010132348	95890
DS38442	MOLM-13	500nM	Day6	0	AML	62078954	0.735549732362	4.99784964171	116222
DS41827	MOLM-13	500nM	Day6	1	AML	63228484	0.609645424995	6.44749224816	109239
DS38230	MOLM-13	500nM	Day2	0	AML	70369386	0.674090377881	12.6239166144	105091
DS41771	MOLM-13	500nM	Day2	1	AML	57788374	0.693130559444	2.70583455383	102540
DS38434	MOLM-13	10nM	Day6	0	AML	49631762	0.729890105453	7.65832695941	95051
DS41820	MOLM-13	10nM	Day6	1	AML	53890898	0.614671219619	2.50657928421	99181
DS38221	MOLM-13	10nM	Day2	0	AML	104823964	0.360756973472	1.98496000738	109632
DS41764	MOLM-13	10nM	Day2	1	AML	69750450	0.669195109135	3.92736969612	111461
DS38416	MOLM-13	None	Day6	0	AML	113682640	0.392636430681	20.1811293526	100972
DS41806	MOLM-13	None	Day6	1	AML	58582082	0.648285972492	1.58053588054	107023
DS38205	MOLM-13	None	Day2	0	AML	332513658	0.569994523353	10.1049778864	258381
DS41750	MOLM-13	None	Day2	1	AML	63157780	0.636025775447	4.45001261639	113490
DS38425	MOLM-13	DMSO	Day6	0	AML	89415620	0.68154792194	12.1602982424	102129
DS41813	MOLM-13	DMSO	Day6	1	AML	78897648	0.585453624676	5.39507834809	130423
DS38213	MOLM-13	DMSO	Day2	0	AML	298914186	0.553729550996	9.79158122007	207199
DS41757	MOLM-13	DMSO	Day2	1	AML	131920090	0.649569568972	10.292935771	151635
DS38496	OCI-AML2	10nM	Day2	0	AML	95553150	0.450763852369	6.79645103237	128129
DS38626	OCI-AML2	10nM	Day6	0	AML	66631622	0.67618831191	3.83131489049	87926
DS41848	OCI-AML2	10nM	Day6	1	AML	49841980	0.761085253836	4.99382664563	87864
DS38480	OCI-AML2	None	Day2	0	AML	105066954	0.393009832568	7.6347345341	123758
DS41721	OCI-AML2	None	Day2	1	AML	542752078	0.552173410564	28.539184127	238933
DS38609	OCI-AML2	None	Day6	0	AML	73273706	0.526035110057	10.6288304061	113303
DS41833	OCI-AML2	None	Day6	1	AML	64612308	0.727391969963	4.05326777206	120543
DS38504	OCI-AML2	500nM	Day2	0	AML	92095762	0.440917292155	5.24106491617	118976
DS38633	OCI-AML2	500nM	Day6	0	AML	64209312	0.566009522108	6.17855414684	84143
DS41855	OCI-AML2	500nM	Day6	1	AML	56801872	0.752396083002	2.53672734202	96677
DS41840	OCI-AML2	DMSO	Day6	1	AML	52335584	0.72556496169	3.09163653857	112909
DS38618	OCI-AML2	DMSO	Day6	0	AML	124076308	0.6478149237	7.18723060266	140206
DS38489	OCI-AML2	DMSO	Day2	0	AML	108182206	0.473996157926	6.44680413316	136298
DS38360	SIG-M5	500nM	Day2	0	AML	117071118	0.590582982218	17.6888296984	107975
DS42109	SIG-M5	500nM	Day2	1	AML	220555670	0.669901698741	8.57807985615	143434

DS-ID	Cell Line	Drug	Time	Batch	Cancer Type	Total Reads	Nuclear Mapping Rate	Nuclear Duplicate Rate	Hotspot Number
DS38563	SIG-M5	DMSO	Day6	0	AML	61884230	0.64305853042	21.6130510472	89602
DS42146	SIG-M5	DMSO	Day6	1	AML	78482732	0.654079218343	4.47201347787	91136
DS38580	SIG-M5	500nM	Day6	0	AML	72776570	0.720298332279	8.80068275134	73056
DS42161	SIG-M5	500nM	Day6	1	AML	77952328	0.719427827736	4.00241621621	76237
DS38344	SIG-M5	DMSO	Day2	0	AML	85721904	0.644533793836	23.7930027411	80839
DS42094	SIG-M5	DMSO	Day2	1	AML	105902970	0.685227033765	9.11456766547	122919
DS38555	SIG-M5	None	Day6	0	AML	80241022	0.656627578846	26.6511336029	96209
DS42139	SIG-M5	None	Day6	1	AML	107308326	0.646320305099	9.18454687476	118364
DS38336	SIG-M5	None	Day2	0	AML	72531230	0.566015190974	13.8226718135	86919
DS42086	SIG-M5	None	Day2	1	AML	326445318	0.702632935296	15.0299070635	182536
DS38571	SIG-M5	10nM	Day6	0	AML	60777706	0.67320122283	10.9151501189	66379
DS42154	SIG-M5	10nM	Day6	1	AML	87856106	0.70981281597	5.98375533323	104927
DS38353	SIG-M5	10nM	Day2	0	AML	61327060	0.647240810174	3.35116871893	101740
DS42102	SIG-M5	10nM	Day2	1	AML	110437416	0.693957109609	3.13509874825	105454
DS38465	SKM-1	10nM	Day6	0	AML	116355724	0.498118270486	7.91981567908	91449
DS42385	SKM-1	10nM	Day6	2	AML	85917162	0.565010585429	16.2465964457	83280
DS41792	SKM-1	10nM	Day2	1	AML	68464916	0.656678378164	5.38005041434	103738
DS38242	SKM-1	None	Day2	0	AML	74362034	0.704029720327	7.40960197912	68604
DS41778	SKM-1	None	Day2	1	AML	611460766	0.590745915495	34.236979087	231968
DS38451	SKM-1	None	Day6	0	AML	51729038	0.682031009353	1.01145926136	62860
DS42371	SKM-1	None	Day6	2	AML	85737214	0.492063667942	17.4464556034	84355
DS38474	SKM-1	500nM	Day6	0	AML	77096198	0.50173851634	10.7462587636	70165
DS42391	SKM-1	500nM	Day6	2	AML	90362412	0.712163106049	11.2600503823	104642
DS38265	SKM-1	500nM	Day2	0	AML	84979078	0.642241870405	13.625152312	87743
DS41799	SKM-1	500nM	Day2	1	AML	71064548	0.671452719294	3.21816879886	112324
DS38458	SKM-1	DMSO	Day6	0	AML	83980338	0.471983334956	7.76055747462	90480
DS42379	SKM-1	DMSO	Day6	2	AML	136177422	0.579627509764	26.1167625169	120181
DS41785	SKM-1	DMSO	Day2	1	AML	69298684	0.655725179428	3.58723812024	113452
DS42237	SKNO-1	500nM	Day2	1	AML	99264724	0.737402302151	8.37142952852	144975
DS39234	SKNO-1	500nM	Day6	0	AML	80094454	0.680689951392	11.7672267294	122737
DS42051	SKNO-1	500nM	Day6	1	AML	128955424	0.578175695813	10.966639633	130269
DS42223	SKNO-1	DMSO	Day2	1	AML	121864722	0.742121579697	9.86032930245	152571
DS39220	SKNO-1	DMSO	Day6	0	AML	68725014	0.663072109378	6.57652331684	109655
DS42037	SKNO-1	DMSO	Day6	1	AML	110148902	0.764224830857	3.74851419945	148540
DS39210	SKNO-1	None	Day6	0	AML	65880326	0.676375159406	5.94368477644	111381
DS42030	SKNO-1	None	Day6	1	AML	113711078	0.758129933479	9.39937990742	144678
DS42217	SKNO-1	None	Day2	1	AML	111616546	0.762916010678	4.69403275593	143463
DS39227	SKNO-1	10nM	Day6	0	AML	113343276	0.643049915021	21.9461312024	121593
DS42045	SKNO-1	10nM	Day6	1	AML	102633744	0.701453140012	8.39930130754	138926
DS42230	SKNO-1	10nM	Day2	1	AML	355947752	0.730339395429	33.9997936473	207875
DS41899	UT-7	500nM	Day6	1	AML	149690978	0.411535810795	4.1038080523	123025
DS38982	UT-7	500nM	Day2	0	AML	115629572	0.550833449422	20.0391737594	100177
DS42080	UT-7	500nM	Day2	1	AML	133983038	0.500143189767	5.18127969526	128977
DS38966	UT-7	DMSO	Day2	0	AML	82558934	0.58312465614	6.22788138876	116885
DS42066	UT-7	DMSO	Day2	1	AML	140287248	0.542419778596	6.17644926029	152500
DS39066	UT-7	DMSO	Day6	0	AML	106830312	0.535480809978	13.1918804707	89946
DS41885	UT-7	DMSO	Day6	1	AML	94680006	0.433104176187	4.79817908982	66726
DS38974	UT-7	10nM	Day2	0	AML	77255146	0.588073317472	8.6331677294	83339
DS42073	UT-7	10nM	Day2	1	AML	119907732	0.572895882978	4.14045368252	136020
DS39382	UT-7	10nM	Day6	0	AML	81306166	0.548653148889	19.1102987468	91545
DS41893	UT-7	10nM	Day6	1	AML	99542316	0.495439346619	4.6889116531	109557

DS-ID	Cell Line	Drug	Time	Batch	Cancer Type	Total Reads	Nuclear Mapping Rate	Nuclear Duplicate Rate	Hotspot Number
DS38958	UT-7	None	Day2	0	AML	79852752	0.641491228756	7.56867371377	104756
DS42059	UT-7	None	Day2	1	AML	347996994	0.589676915428	5.97272706637	230128
DS39059	UT-7	None	Day6	0	AML	143303692	0.565228717206	23.1776368806	137327
DS41879	UT-7	None	Day6	1	AML	86703562	0.497748316269	3.01588041603	83994
DS27661	NCI_H345	None	None	0	SCLC	320305228	0.788650767823	59.9917626043	184996
DS27846	NCI_H128	None	None	0	SCLC	325155052	0.524445635862	72.1437078377	119870
DS25739	DMS_79	None	None	0	SCLC	444262504	0.643347181962	8.27906146153	229470
DS26267	NCI_H69	None	None	0	SCLC	481107442	0.702348703224	17.2514545707	324775
DS25758	NCI_H1048	None	None	0	SCLC	56874214	0.692399968815	1.01220161533	75298
DS25642	NCI_H2081	None	None	0	SCLC	532493480	0.69251344824	12.8681663045	282236
DS25856	NCI_H1436	None	None	0	SCLC	405473226	0.753500079436	19.3244914081	253574
DS25897	NCI_H1105	None	None	0	SCLC	474298380	0.760848363007	20.3747526811	294786
DS25571	NCI_H187	None	None	0	SCLC	703720552	0.503332359689	11.8995223691	212226
DS25463	NCI_H1417	None	None	0	SCLC	465878750	0.797859979662	4.90432317734	223604
DS25810	NCI_H1694	None	None	0	SCLC	509553742	0.708204564613	34.2082085872	226172
DS27500	COR_L47	None	None	0	SCLC	306020742	0.776075936709	42.6473136521	254468
DS26866	NCI_H748	None	None	0	SCLC	395564662	0.73271663989	27.562085713	226654
DS26706	COLO_668	None	None	0	SCLC	367181174	0.672355048356	11.2571059729	138484
DS26323	DMS_53	None	None	0	SCLC	419312516	0.662348671701	12.0229234114	251648
DS26121	NCI_H2029	None	None	0	SCLC	700069188	0.496114801156	14.017884058	269271
DS26256	NCI_H2141	None	None	0	SCLC	718836812	0.572775357531	17.9003569924	346526
DS25841	NCI_H82	None	None	0	SCLC	438998536	0.471943910993	32.8768833512	257997
DS26892	NCI_H2227	None	None	0	SCLC	478976738	0.629305972684	46.8797326568	151047
DS25297	NCI_H526	None	None	0	SCLC	447581034	0.694231900809	9.3450018536	278838
DS26717	NCI_H2107	None	None	0	SCLC	429760254	0.712826254054	10.7128646246	254016
DS26183	NCI_H889	None	None	0	SCLC	541060012	0.766797872322	19.2224675338	300436
DS26741	NCI_H209	None	None	0	SCLC	553777556	0.663616081978	12.9555907828	239244
DS25323	NCI_H446	None	None	0	SCLC	507255120	0.758348643184	5.25986331279	215892
DS25658	NCI_H1930	None	None	0	SCLC	465992452	0.676990939759	9.84540315233	141511
DS27127	SW1271	None	None	0	SCLC	66353934	0.738810241455	1.08484664106	74207
DS26339	NCI_H510A	None	None	0	SCLC	482345782	0.609810814102	13.5729714585	295100
DS27064	DMS_273	None	None	0	SCLC	480947510	0.572481608232	13.2793765341	224214
DS27605	NCI_H774	None	None	0	SCLC	455101980	0.666462971662	75.8609819652	374612
DS26292	NCI_H2171	None	None	0	SCLC	356762500	0.762731873445	19.2082630917	291072
DS25307	NCI_H524	None	None	0	SCLC	431001234	0.78559893868	8.91848585955	299183
DS27544	DMS_153	None	None	0	SCLC	435932818	0.672192897393	20.1005106137	289711
DS25826	NCI_H1092	None	None	0	SCLC	407156666	0.714204025828	13.8469054251	189013
DS25662	NCI_H146	None	None	0	SCLC	372299912	0.668010622576	41.0639873143	213045
DS26482	DMS_114	None	None	0	SCLC	122583024	0.631081233565	22.6288945818	33727
DS27036	NCI_H2196	None	None	0	SCLC	95663116	0.70741987957	6.70089350428	87667
DS25590	SHP_77	None	None	0	SCLC	779033972	0.343729959956	47.2491067914	246402
DS43107	COLO_668	DMSO	Day2	1	SCLC	75519846	0.701183474341	4.1522234284	87293
DS43113	COLO_668	10nM	Day2	1	SCLC	379208434	0.685304235612	11.9559159837	165551
DS40449	COLO_668	10nM	Day2	0	SCLC	69510162	0.801616258641	7.74774608889	74011
DS43589	COR_L47	DMSO	Day2	1	SCLC	71257196	0.7562943678	8.13421009387	115693
DS40477	COR_L47	10nM	Day2	0	SCLC	76738856	0.703117961519	12.5144403448	111461
DS43595	COR_L47	10nM	Day2	1	SCLC	104176904	0.745595952823	7.78923127799	133391
DS44312	DMS_153	DMSO	Day2	1	SCLC	49754014	0.765507884449	1.79058835947	63981
DS40639	DMS_153	10nM	Day2	0	SCLC	79662552	0.707461192054	5.00603781202	114939
DS44317	DMS_153	10nM	Day2	1	SCLC	58542282	0.70111243699	6.14421021051	104161
DS44088	DMS_53	DMSO	Day2	1	SCLC	50853856	0.759760361142	6.49707439115	98701

DS-ID	Cell Line	Drug	Time	Batch	Cancer Type	Total Reads	Nuclear Mapping Rate	Nuclear Duplicate Rate	Hotspot Number
DS40515	DMS_53	10nM	Day2	0	SCLC	71922550	0.604641576251	13.7056985105	107762
DS44093	DMS_53	10nM	Day2	1	SCLC	74352784	0.721232469251	13.3306842275	122278
DS43763	DMS_114	DMSO	Day2	1	SCLC	86473476	0.541428865425	5.78849684775	46552
DS40488	DMS_114	10nM	Day2	0	SCLC	115186578	0.355878512165	5.30503096816	32748
DS43768	DMS_114	10nM	Day2	1	SCLC	66831768	0.71191724271	6.97022162528	37244
DS43581	DMS_273	DMSO	Day2	1	SCLC	82038748	0.580646671985	4.97203914574	82292
DS40453	DMS_273	10nM	Day2	0	SCLC	76864930	0.518616253212	3.12460286255	86988
DS43585	DMS_273	10nM	Day2	1	SCLC	155781666	0.406940467564	16.8618748338	93349
DS43557	DMS_79	DMSO	Day2	1	SCLC	84972412	0.793250472871	5.20196409712	100092
DS40644	DMS_79	10nM	Day2	0	SCLC	94151378	0.778003142981	2.40134384585	106623
DS43563	DMS_79	10nM	Day2	1	SCLC	79887246	0.805066781248	2.60550455691	107023
DS43085	NCI_H1048	DMSO	Day2	1	SCLC	50149032	0.775257516436	2.67227979001	43702
DS40634	NCI_H1048	10nM	Day2	0	SCLC	71409864	0.648369026442	9.63253864843	88535
DS43091	NCI_H1048	10nM	Day2	1	SCLC	87752140	0.775938113874	4.19919127898	68324
DS43306	NCI_H1092	DMSO	Day2	1	SCLC	63271830	0.788528322952	2.81264212053	91977
DS40705	NCI_H1092	10nM	Day2	0	SCLC	107662384	0.757843537999	4.15036211852	123693
DS43312	NCI_H1092	10nM	Day2	1	SCLC	74134068	0.750138708158	4.12289267232	102806
DS43348	NCI_H1105	DMSO	Day2	1	SCLC	73515978	0.670603089848	5.42519821788	116430
DS40711	NCI_H1105	10nM	Day2	0	SCLC	65458182	0.742184437692	0.959033341619	116065
DS43355	NCI_H1105	10nM	Day2	1	SCLC	64119496	0.774013195612	4.6769515514	108251
DS43867	NCI_H128	DMSO	Day2	1	SCLC	80933148	0.769383837633	10.003132234	68229
DS40648	NCI_H128	10nM	Day2	0	SCLC	59355188	0.745475728255	5.69056323909	72124
DS43870	NCI_H128	10nM	Day2	1	SCLC	78234216	0.771623786707	5.8945027631	75049
DS43294	NCI_H1417	DMSO	Day2	1	SCLC	62362770	0.72465562386	23.0818087778	84169
DS40491	NCI_H1417	10nM	Day2	0	SCLC	80041214	0.668161055128	7.63577931378	89074
DS43299	NCI_H1417	10nM	Day2	1	SCLC	64531930	0.707929826367	6.47346324906	104219
DS42809	NCI_H1436	DMSO	Day2	1	SCLC	185894324	0.621237375704	20.2747620629	126655
DS40615	NCI_H1436	10nM	Day2	0	SCLC	71970048	0.688601597153	29.1652886709	78162
DS42815	NCI_H1436	10nM	Day2	1	SCLC	56507204	0.764251758059	2.86335787308	73948
DS43317	NCI_H146	DMSO	Day2	1	SCLC	57071072	0.775506126817	1.63214838774	82322
DS40433	NCI_H146	10nM	Day2	0	SCLC	73403578	0.756494322388	4.99173500735	116862
DS43323	NCI_H146	10nM	Day2	1	SCLC	59590136	0.787337421079	1.73809822611	92578
DS42822	NCI_H1694	DMSO	Day2	1	SCLC	94727722	0.564623606171	14.4091741636	85514
DS40521	NCI_H1694	10nM	Day2	0	SCLC	76425892	0.722561432453	13.6766108073	108151
DS42828	NCI_H1694	10nM	Day2	1	SCLC	121505200	0.668166728667	33.1220145924	104296
DS43235	NCI_H187	DMSO	Day2	1	SCLC	59282106	0.771329513833	29.5044084073	54238
DS40372	NCI_H187	10nM	Day2	0	SCLC	125752746	0.787441269871	2.0297486333	83680
DS43241	NCI_H187	10nM	Day2	1	SCLC	54036712	0.748100513592	1.92190742278	63928
DS43222	NCI_H1930	DMSO	Day2	1	SCLC	75876076	0.745299954626	5.77458147272	60027
DS43229	NCI_H1930	10nM	Day2	1	SCLC	81706566	0.736393474179	4.18437273042	56972
DS43205	NCI_H2029	10nM	Day2	1	SCLC	69500658	0.740955603615	21.6498091477	86552
DS40655	NCI_H2029	10nM	Day2	0	SCLC	80128186	0.715937934749	11.1884733067	116381
DS43208	NCI_H2029	DMSO	Day2	1	SCLC	99807384	0.668076682583	11.3217663508	118718
DS42866	NCI_H2081	DMSO	Day2	1	SCLC	76453210	0.717903669447	6.95554279376	94382
DS40627	NCI_H2081	10nM	Day2	0	SCLC	56881266	0.668125811405	10.6035016144	77702
DS42873	NCI_H2081	10nM	Day2	1	SCLC	86034664	0.74271512236	5.29317982876	94137
DS42608	NCI_H209	DMSO	Day2	1	SCLC	62681692	0.66498163451	2.59985959466	46353
DS40439	NCI_H209	10nM	Day2	0	SCLC	77177206	0.761573410678	3.00709601255	74807
DS42616	NCI_H209	10nM	Day2	1	SCLC	64790068	0.694518456131	4.18514660396	63983
DS43214	NCI_H2107	DMSO	Day2	1	SCLC	52680922	0.738476027432	11.9431652568	76456
DS40762	NCI_H2107	10nM	Day2	0	SCLC	131301066	0.70776109312	34.5465080486	122609

DS-ID	Cell Line	Drug	Time	Batch	Cancer Type	Total Reads	Nuclear Mapping Rate	Nuclear Duplicate Rate	Hotspot Number
DS43217	NCI_H2107	10nM	Day2	1	SCLC	62336748	0.552080002633	23.7672974637	73540
DS43061	NCI_H2141	DMSO	Day2	1	SCLC	84616910	0.746185673762	4.17082845488	108582
DS40527	NCI_H2141	10nM	Day2	0	SCLC	101080196	0.768966039599	3.74052401039	146558
DS43066	NCI_H2141	10nM	Day2	1	SCLC	87809410	0.757795366123	4.18425842831	122542
DS43069	NCI_H2171	DMSO	Day2	1	SCLC	117789434	0.762070662467	9.18587157203	153342
DS40531	NCI_H2171	10nM	Day2	0	SCLC	104209856	0.737965687238	9.57667901317	132522
DS43073	NCI_H2171	10nM	Day2	1	SCLC	82569316	0.731931750531	8.95284965506	131381
DS43755	NCI_H2196	DMSO	Day2	1	SCLC	77883390	0.58086601007	11.7334390455	65161
DS40754	NCI_H2196	10nM	Day2	0	SCLC	90649522	0.687456597951	5.49735923771	81613
DS43758	NCI_H2196	10nM	Day2	1	SCLC	61275606	0.68942789403	5.06118909094	60200
DS43193	NCI_H2227	DMSO	Day2	1	SCLC	80363658	0.608413469681	10.5498322382	79628
DS40698	NCI_H2227	10nM	Day2	0	SCLC	66593836	0.719530017763	6.81743468147	61406
DS43199	NCI_H2227	10nM	Day2	1	SCLC	95945936	0.674836066011	5.3667756753	94248
DS43339	NCI_H345	DMSO	Day2	1	SCLC	110265196	0.446686132948	22.2797086625	82937
DS40759	NCI_H345	10nM	Day2	0	SCLC	89078730	0.539270732755	5.57234889904	79989
DS43343	NCI_H345	10nM	Day2	1	SCLC	99420946	0.490793760904	16.4454194041	87645
DS43429	NCI_H446	DMSO	Day2	1	SCLC	62044044	0.612165448145	8.19942855969	78703
DS43436	NCI_H446	10nM	Day2	1	SCLC	83660266	0.746986532412	6.03093378705	104512
DS43890	NCI_H510A	DMSO	Day2	1	SCLC	71196332	0.759715374101	4.17574584738	93882
DS40621	NCI_H510A	10nM	Day2	0	SCLC	65085814	0.693804520905	3.34609611232	97499
DS43894	NCI_H510A	10nM	Day2	1	SCLC	55148616	0.69440611166	5.83238474584	88658
DS43448	NCI_H524	DMSO	Day2	1	SCLC	68201682	0.805041582406	3.69807480859	137405
DS40443	NCI_H524	10nM	Day2	0	SCLC	77226280	0.742286641283	9.54184733259	149446
DS43454	NCI_H524	10nM	Day2	1	SCLC	84753364	0.746342552256	5.24727341103	157755
DS43420	NCI_H526	DMSO	Day2	1	SCLC	62283294	0.720552480734	2.56671018131	84068
DS40378	NCI_H526	10nM	Day2	0	SCLC	92909850	0.661113778571	11.5410834159	115304
DS43424	NCI_H526	10nM	Day2	1	SCLC	78888814	0.714371038713	6.58168187017	113181
DS43545	NCI_H69	DMSO	Day2	1	SCLC	71162494	0.776978670815	4.17850659068	109956
DS40375	NCI_H69	10nM	Day2	0	SCLC	127675828	0.665571622531	3.8024824198	157056
DS43551	NCI_H69	10nM	Day2	1	SCLC	78343348	0.8048699425	2.59273290245	124761
DS43329	NCI_H748	DMSO	Day2	1	SCLC	60415226	0.774356020782	2.38235154926	83020
DS43335	NCI_H748	10nM	Day2	1	SCLC	83215046	0.624677513247	8.65529767594	70841
DS43442	NCI_H774	DMSO	Day2	1	SCLC	67644926	0.797284987791	2.84400156315	57825
DS40746	NCI_H774	10nM	Day2	0	SCLC	90322744	0.51438621041	45.3699609051	27066
DS43444	NCI_H774	10nM	Day2	1	SCLC	85517440	0.775531517314	3.7457689041	72903
DS43096	NCI_H82	DMSO	Day2	1	SCLC	66697016	0.631404349484	3.19478981989	93716
DS40482	NCI_H82	10nM	Day2	0	SCLC	80677424	0.634065832345	11.3897879139	104953
DS43101	NCI_H82	10nM	Day2	1	SCLC	93565978	0.465958919384	7.71735071105	99866
DS45304	NCI_H889	DMSO	Day2	1	SCLC	55002856	0.713399246032	8.58128225299	96201
DS40430	NCI_H889	10nM	Day2	0	SCLC	83905200	0.632324337467	8.08935959273	120066
DS45309	NCI_H889	10nM	Day2	1	SCLC	88292012	0.737035984637	8.47408481278	130384
DS43570	SHP_77	DMSO	Day2	1	SCLC	72952282	0.790953516711	3.38751968221	103990
DS40659	SHP_77	10nM	Day2	0	SCLC	88135996	0.587561272922	4.0378912177	99311
DS43575	SHP_77	10nM	Day2	1	SCLC	50995758	0.725468773305	4.0172013981	84183
DS43698	SW1271	DMSO	Day2	1	SCLC	115513084	0.716530622626	10.5977610222	85386
DS40611	SW1271	10nM	Day2	0	SCLC	65600266	0.664522671295	8.48407665989	69307
DS43704	SW1271	10nM	Day2	1	SCLC	55380212	0.716076529285	4.4271767003	66527
DS27171	NCI_H526	500nM	Day2	0	SCLC	108187656	0.567790728362	13.70442327	160382
DS33118	NCI_H526	500nM	Day2	1	SCLC	100070012	0.59746911992	10.5373683895	142157
DS33617	NCI_H526	500nM	Day2	2	SCLC	379451634	0.812092225699	27.3523654193	237792
DS33111	NCI_H526	10nM	Day2	1	SCLC	107523828	0.559118058929	32.6131580697	137323

DS-ID	Cell Line	Drug	Time	Batch	Cancer Type	Total Reads	Nuclear Mapping Rate	Nuclear Duplicate Rate	Hotspot Number
DS33608	NCI_H526	10nM	Day2	2	SCLC	315180762	0.774131176192	17.3337122977	207111
DS27207	NCI_H526	500nM	Day7	0	SCLC	76658128	0.747721650599	3.33448583193	113861
DS27574	NCI_H526	500nM	Day7	0	SCLC	456184552	0.717673916323	16.8075852026	246731
DS33327	NCI_H526	500nM	Day7	1	SCLC	90097660	0.673281747828	5.37938392489	135792
DS33643	NCI_H526	500nM	Day7	2	SCLC	123058918	0.783895775843	9.72363283212	135794
DS27204	NCI_H526	DMSO	Day7	0	SCLC	100643040	0.670935774595	4.3571674237	157282
DS33313	NCI_H526	DMSO	Day7	1	SCLC	90828968	0.518739594179	6.28240393159	144673
DS33627	NCI_H526	DMSO	Day7	2	SCLC	100600388	0.78402202584	12.7351947918	132339
DS33101	NCI_H526	DMSO	Day2	1	SCLC	122698162	0.500906639498	28.6971445188	123966
DS33600	NCI_H526	DMSO	Day2	2	SCLC	118300722	0.68713878179	5.99965943737	154589
DS27582	NCI_H526	10nM	Day7	0	SCLC	478133318	0.696733411914	14.8807363608	260379
DS33319	NCI_H526	10nM	Day7	1	SCLC	80312522	0.70098294261	3.60567787236	126033
DS33636	NCI_H526	10nM	Day7	2	SCLC	44665204	0.813644912492	2.64559506655	79869
DS30387	NCI_H510A	10nM	Day2	0	SCLC	382242104	0.701598351395	17.310694893	261930
DS33741	NCI_H510A	10nM	Day2	1	SCLC	84812760	0.72806639001	5.27702977134	130188
DS30521	NCI_H510A	10nM	Day7	0	SCLC	379670878	0.674508198651	23.7913786312	221191
DS34004	NCI_H510A	10nM	Day7	1	SCLC	113974140	0.555894100188	15.6608134102	162869
DS30379	NCI_H510A	DMSO	Day2	0	SCLC	83004832	0.720173784582	3.05662105516	119259
DS33732	NCI_H510A	DMSO	Day2	1	SCLC	130557818	0.751385535564	5.0315835219	172114
DS30395	NCI_H510A	500nM	Day2	0	SCLC	123459168	0.668398283714	31.167351447	154673
DS33748	NCI_H510A	500nM	Day2	1	SCLC	96397188	0.725772809887	6.27645472376	127672
DS30528	NCI_H510A	500nM	Day7	0	SCLC	532390662	0.646201484278	26.2911065539	266214
DS34011	NCI_H510A	500nM	Day7	1	SCLC	84042892	0.566723239367	22.7257518834	109977
DS30512	NCI_H510A	DMSO	Day7	0	SCLC	102382216	0.720117661841	16.3385821485	143543
DS33995	NCI_H510A	DMSO	Day7	1	SCLC	78591280	0.71834895678	5.2199480643	128242
DS30493	NCI_H2029	500nM	Day7	0	SCLC	610412684	0.736148071261	14.2805736376	338898
DS30372	NCI_H2029	500nM	Day2	0	SCLC	608919268	0.740182956405	39.3317471367	311257
DS35108	NCI_H2029	500nM	Day2	1	SCLC	93749116	0.638858525343	24.7947327961	114333
DS30355	NCI_H2029	DMSO	Day2	0	SCLC	87920364	0.642269975133	7.04983883967	143476
DS35093	NCI_H2029	DMSO	Day2	1	SCLC	94482718	0.727710542789	6.50132271958	119067
DS30484	NCI_H2029	10nM	Day7	0	SCLC	702635188	0.720927973223	20.0755439901	334138
DS35144	NCI_H2029	10nM	Day7	1	SCLC	123217022	0.590353644483	37.2608868832	135312
DS30364	NCI_H2029	10nM	Day2	0	SCLC	455047226	0.684175000333	34.3066642543	267904
DS35100	NCI_H2029	10nM	Day2	1	SCLC	78736706	0.629488030652	14.2277307392	103769
DS30477	NCI_H2029	DMSO	Day7	0	SCLC	91893480	0.618641888413	5.11882005777	117250
DS35136	NCI_H2029	DMSO	Day7	1	SCLC	79518620	0.578039810047	35.0849847954	99989
DS34519	NCI_H1417	DMSO	Day2	1	SCLC	78221988	0.620034484421	43.1424693399	88298
DS34135	NCI_H1417	DMSO	Day2	0	SCLC	62097728	0.719619532618	9.33020441098	81943
DS34805	NCI_H1417	DMSO	Day2	2	SCLC	105989488	0.701751026479	21.6462015392	133765
DS34399	NCI_H1417	DMSO	Day7	1	SCLC	68590812	0.789405904686	10.074806645	91958
DS34374	NCI_H1417	DMSO	Day7	1	SCLC	72505524	0.783802596889	11.1406792094	94011
DS34390	NCI_H1417	500nM	Day7	1	SCLC	88320454	0.805445225633	6.90896734776	88898
DS34412	NCI_H1417	500nM	Day7	1	SCLC	410617458	0.815665265747	10.1627874671	183885
DS34383	NCI_H1417	10nM	Day7	1	SCLC	351903016	0.832448801746	13.8768646726	170589
DS34405	NCI_H1417	10nM	Day7	1	SCLC	54049036	0.742554853337	17.4186546425	59327
DS34143	NCI_H1417	10nM	Day2	0	SCLC	48552826	0.83447711159	33.1880104715	65718
DS34810	NCI_H1417	10nM	Day2	2	SCLC	383934470	0.594273699884	61.6803002012	172804
DS34817	NCI_H1417	500nM	Day2	2	SCLC	354164746	0.650358881287	40.0098199925	169164
DS34151	NCI_H1417	500nM	Day2	0	SCLC	58757290	0.845054937013	30.9023852631	76550
DS30893	NCI_H187	DMSO	Day2	0	SCLC	104177950	0.761094415853	10.4224360486	105060
DS30662	NCI_H187	DMSO	Day2	1	SCLC	140199264	0.704056135416	3.91484991027	118705

DS-ID	Cell Line	Drug	Time	Batch	Cancer Type	Total Reads	Nuclear Mapping Rate	Nuclear Duplicate Rate	Hotspot Number
DS30925	NCI_H187	DMSO	Day7	0	SCLC	75944432	0.702491447958	4.87018314456	105073
DS30813	NCI_H187	DMSO	Day7	1	SCLC	131246006	0.750082528226	13.1510506501	113376
DS34068	NCI_H187	DMSO	Day7	2	SCLC	69708440	0.697312176259	9.86852846282	80044
DS30942	NCI_H187	500nM	Day7	0	SCLC	595641528	0.713926951044	31.4493020886	219398
DS30829	NCI_H187	500nM	Day7	1	SCLC	156734736	0.729311325091	6.0907657737	129218
DS34084	NCI_H187	500nM	Day7	2	SCLC	99168644	0.522306829163	24.9459905175	83836
DS30908	NCI_H187	500nM	Day2	0	SCLC	100960454	0.736847023291	4.17570824766	107911
DS30676	NCI_H187	500nM	Day2	1	SCLC	491854728	0.60390933967	9.36738761444	166360
DS33845	NCI_H187	500nM	Day2	2	SCLC	60933828	0.677802976698	45.7727621496	38906
DS30900	NCI_H187	10nM	Day2	0	SCLC	445346066	0.773336527014	25.6529435133	194571
DS30670	NCI_H187	10nM	Day2	1	SCLC	118851858	0.763530949596	4.00922467228	107627
DS30933	NCI_H187	10nM	Day7	0	SCLC	93584838	0.714519311344	5.66901079129	101044
DS30821	NCI_H187	10nM	Day7	1	SCLC	550721356	0.704954303606	20.8671428287	201111
DS34075	NCI_H187	10nM	Day7	2	SCLC	76277360	0.780588683195	2.46316318147	82996
DS30702	NCI_H2081	500nM	Day7	0	SCLC	486659580	0.773283879463	11.5753790793	236965
DS34059	NCI_H2081	500nM	Day7	1	SCLC	91762988	0.594128103152	24.5455334171	79041
DS30632	NCI_H2081	500nM	Day2	0	SCLC	394342666	0.759179718078	15.2524345611	236515
DS30693	NCI_H2081	10nM	Day7	0	SCLC	421215448	0.711281538753	30.3724793307	194335
DS34051	NCI_H2081	10nM	Day7	1	SCLC	113998470	0.552287166661	16.2507934416	94250
DS30625	NCI_H2081	10nM	Day2	0	SCLC	419799920	0.703001563221	9.89212388181	226282
DS33812	NCI_H2081	10nM	Day2	1	SCLC	128571508	0.714017338896	3.19277016697	111639
DS30616	NCI_H2081	DMSO	Day2	0	SCLC	114809494	0.678481955508	13.7409862385	130364
DS33803	NCI_H2081	DMSO	Day2	1	SCLC	97029438	0.733201793872	5.92028990743	73087
DS30686	NCI_H2081	DMSO	Day7	0	SCLC	78132626	0.7073058059	8.2284959049	115052
DS34043	NCI_H2081	DMSO	Day7	1	SCLC	79001894	0.685663890539	5.38267583303	87268
DS30222	NCI_H446	DMSO	Day2	0	SCLC	104337828	0.658828224793	13.1764489006	70877
DS33707	NCI_H446	DMSO	Day2	1	SCLC	97257776	0.787433428459	6.84109782479	84314
DS30246	NCI_H446	DMSO	Day7	0	SCLC	94314634	0.740514902491	5.51692325949	120466
DS33970	NCI_H446	DMSO	Day7	1	SCLC	72858844	0.685322072911	6.65133588084	100650
DS30261	NCI_H446	500nM	Day2	0	SCLC	579911472	0.475341784582	65.3730952076	186813
DS33987	NCI_H446	500nM	Day2	1	SCLC	67854860	0.679323485451	20.145984198	73794
DS30237	NCI_H446	500nM	Day7	0	SCLC	501672538	0.676432804859	31.4325055133	270212
DS33723	NCI_H446	500nM	Day7	1	SCLC	107485686	0.749507427436	5.38929848716	112218
DS30229	NCI_H446	10nM	Day2	0	SCLC	589119990	0.633413665016	48.7723086335	236331
DS33716	NCI_H446	10nM	Day2	1	SCLC	129196100	0.811612393873	5.93046029209	96719
DS30254	NCI_H446	10nM	Day7	0	SCLC	587596920	0.505623664603	53.6523816186	176950
DS33979	NCI_H446	10nM	Day7	1	SCLC	58122546	0.75321886278	14.5636087879	100096
DS30208	NCI_H524	10nM	Day2	0	SCLC	144247142	0.567746915915	7.94370705141	201368
DS34306	NCI_H524	10nM	Day2	1	SCLC	388039106	0.838481165865	24.217767008	252911
DS30216	NCI_H524	500nM	Day2	0	SCLC	155292146	0.747536285576	10.6091858353	256747
DS34311	NCI_H524	500nM	Day2	1	SCLC	403107324	0.853079642879	12.8071008362	266202
DS30277	NCI_H524	10nM	Day7	0	SCLC	419939408	0.77366407108	15.5406162746	304757
DS34027	NCI_H524	10nM	Day7	1	SCLC	85468424	0.642456329837	19.6875781396	154584
DS30286	NCI_H524	500nM	Day7	0	SCLC	539900196	0.787852886795	23.2897030691	349163
DS34035	NCI_H524	500nM	Day7	1	SCLC	80323370	0.677317995995	13.4305633449	140874
DS30269	NCI_H524	DMSO	Day7	0	SCLC	113318054	0.821641077599	5.2186109607	197830
DS34019	NCI_H524	DMSO	Day7	1	SCLC	62678614	0.708882075791	4.29924585903	121324
DS30200	NCI_H524	DMSO	Day2	0	SCLC	89815574	0.732258438832	15.0627509837	180255
DS34299	NCI_H524	DMSO	Day2	1	SCLC	77551572	0.720507483717	7.52547654749	131846
DS33515	SHP_77	10nM	Day7	1	SCLC	482592088	0.805831408492	10.4883186094	235700
DS34428	SHP_77	10nM	Day7	2	SCLC	216001678	0.584855549131	7.61630076455	145664

DS-ID	Cell Line	Drug	Time	Batch	Cancer Type	Total Reads	Nuclear Mapping Rate	Nuclear Duplicate Rate	Hotspot Number
DS33357	SHP_77	10nM	Day2	1	SCLC	54257302	0.68003097537	6.00362970661	109562
DS34121	SHP_77	10nM	Day2	2	SCLC	72973526	0.736683437772	15.9987237713	92329
DS27175	SHP_77	500nM	Day2	0	SCLC	86758412	0.632447583296	20.3310258977	112837
DS33365	SHP_77	500nM	Day2	1	SCLC	350107672	0.657521312472	14.3644232382	262533
DS34128	SHP_77	500nM	Day2	2	SCLC	74851296	0.59931905521	21.6565349544	57363
DS27199	SHP_77	500nM	Day7	0	SCLC	119371090	0.609353588042	2.00492438727	138014
DS33522	SHP_77	500nM	Day7	1	SCLC	516870678	0.633862461047	17.176938189	246720
DS34437	SHP_77	500nM	Day7	2	SCLC	120566894	0.539259890032	22.6955118893	108157
DS27194	SHP_77	DMSO	Day7	0	SCLC	93401190	0.601441287847	2.10132803487	116052
DS33506	SHP_77	DMSO	Day7	1	SCLC	197959430	0.665579144171	12.2829010605	175237
DS34421	SHP_77	DMSO	Day7	2	SCLC	80481782	0.681817656572	5.16599694937	106798
DS33348	SHP_77	DMSO	Day2	1	SCLC	134924956	0.524171688446	4.96662780283	166073
DS34113	SHP_77	DMSO	Day2	2	SCLC	103527378	0.799595156365	9.42649787708	115876
DS30552	NCI_H1930	500nM	Day7	0	SCLC	504867062	0.716916454336	24.8122843803	137417
DS34460	NCI_H1930	500nM	Day7	1	SCLC	100664282	0.555260285868	12.9838336887	50034
DS30435	NCI_H1930	DMSO	Day2	0	SCLC	75301356	0.704158873314	3.65297165339	70111
DS34156	NCI_H1930	DMSO	Day2	1	SCLC	54842478	0.83094370754	11.5800281109	33996
DS30448	NCI_H1930	500nM	Day2	0	SCLC	433644368	0.659553586085	16.4472412593	128990
DS34170	NCI_H1930	500nM	Day2	1	SCLC	49828278	0.815570467838	29.0850453236	26690
DS30537	NCI_H1930	DMSO	Day7	0	SCLC	77433056	0.716841060748	5.89828770663	66916
DS34442	NCI_H1930	DMSO	Day7	1	SCLC	84923242	0.589838951273	6.83278341458	50350
DS30544	NCI_H1930	10nM	Day7	0	SCLC	506621868	0.757303405624	20.2057633048	137688
DS34451	NCI_H1930	10nM	Day7	1	SCLC	74873584	0.708074265551	6.15778713687	39727
DS30442	NCI_H1930	10nM	Day2	0	SCLC	527028704	0.739043192608	41.8340906452	148927
DS34163	NCI_H1930	10nM	Day2	1	SCLC	60137500	0.833477547287	19.6387042374	34712

Supplementary Table #2: DNaseI-Seq Libraries
Table containing sequencing statistics and metadata for DNaseI-Seq experiments performed for Chapters #2-3.

DS-ID	Cell Line	Drug	Time	Total Reads	Ribosomal Fraction
DS26864	NCI_H748	None	None	59169472.0	0.0120768358386
DS27050	DMS_273	None	None	142989094.0	0.018245363524
DS27497	COR_L47	None	None	135507524.0	0.00425802186453
DS27533	DMS_153	None	None	97265884.0	0.00404415180147
DS27836	NCI_H128	None	None	83400948.0	0.0174166365591
DS28590	NCI_H526	None	None	162958432.0	0.0208947518592
DS28591	NCI_H524	None	None	116148978.0	0.00745248055476
DS28592	NCI_H446	None	None	213476086.0	0.00491484559071
DS28593	NCI_H1417	None	None	158881198.0	0.00252647893554
DS28594	NCI_H187	None	None	138941758.0	0.127219651273
DS28595	SHP_77	None	None	199124878.0	0.00585960936534
DS28669	NCI_H2081	None	None	327444380.0	0.0127710666465
DS28670	NCI_H1930	None	None	88903294.0	0.0200931362566
DS28671	NCI_H146	None	None	502619074.0	0.0927288764214
DS28672	DMS_79	None	None	330700930.0	0.0218642687216
DS28673	NCI_H1048	None	None	174123252.0	0.0352139069858
DS28674	NCI_H1694	None	None	281090312.0	0.0230200463117
DS28702	NCI_H1092	None	None	137702052.0	0.0199845533166
DS28703	NCI_H82	None	None	192772006.0	0.0481854818692
DS28704	NCI_H1436	None	None	197304946.0	0.105573369661
DS28705	NCI_H1105	None	None	189421938.0	0.0265174037022
DS28706	NCI_H2029	None	None	175578188.0	0.0139885485092
DS28707	NCI_H889	None	None	275858648.0	0.0226110076491
DS28759	NCI_H2141	None	None	179441872.0	0.0108797906433
DS28760	NCI_H69	None	None	218273900.0	0.0204979523434
DS28762	DMS_53	None	None	215277674.0	0.0231559358078
DS28763	NCI_H510A	None	None	164529538.0	0.144783898925
DS28764	COLO_668	None	None	173797960.0	0.0443052956433
DS28765	NCI_H2107	None	None	178366822.0	0.0450925228684
DS28766	NCI_H209	None	None	235589656.0	0.050795201297
DS28767	NCI_H2227	None	None	221214328.0	0.142752869064
DS28768	SW1271	None	None	199742436.0	0.0807290144394
DS30196	NCI_H524	DMSO	Day2	96642310.0	0.0126245533659
DS30204	NCI_H524	10nM	Day2	104293538.0	0.0219325764939
DS30212	NCI_H524	500nM	Day2	93715022.0	0.00941057240535
DS30220	NCI_H446	DMSO	Day2	97218746.0	0.0172963555815
DS30227	NCI_H446	10nM	Day2	92594802.0	0.0143337635735
DS30234	NCI_H446	500nM	Day2	125499456.0	0.0164973145382
DS30242	NCI_H446	DMSO	Day7	75889452.0	0.0342435994926
DS30250	NCI_H446	10nM	Day7	92733402.0	0.0140048781991
DS30258	NCI_H446	500nM	Day7	94380148.0	0.0101850444227
DS30266	NCI_H524	DMSO	Day7	110024628.0	0.0122360059241
DS30274	NCI_H524	10nM	Day7	103541624.0	0.0330790832487
DS30282	NCI_H524	500nM	Day7	109725804.0	0.0182171551917
DS30350	NCI_H2029	DMSO	Day2	111217542.0	0.0118987524468
DS30358	NCI_H2029	10nM	Day2	94096416.0	0.0173712461057
DS30366	NCI_H2029	500nM	Day2	82196220.0	0.0213812508653
DS30374	NCI_H510A	DMSO	Day2	100929520.0	0.0134442529797
DS30382	NCI_H510A	10nM	Day2	118617464.0	0.00451365239102
DS30390	NCI_H510A	500nM	Day2	85389246.0	0.00422074227005
DS30431	NCI_H1930	DMSO	Day2	102018830.0	0.0172297604276
DS30438	NCI_H1930	10nM	Day2	90712842.0	0.00719291762461
DS30445	NCI_H1930	500nM	Day2	103695040.0	0.00526125454024
DS30472	NCI_H2029	DMSO	Day7	108428030.0	0.0577769973318
DS30480	NCI_H2029	10nM	Day7	97877064.0	0.0147383047779
DS30488	NCI_H2029	500nM	Day7	76124492.0	0.016207740342
DS30508	NCI_H510A	DMSO	Day7	90745388.0	0.00743537511791
DS30516	NCI_H510A	10nM	Day7	132791356.0	0.00588845557086
DS30524	NCI_H510A	500nM	Day7	154286658.0	0.0149220809488
DS30532	NCI_H1930	DMSO	Day7	104554110.0	0.00826833110626
DS30540	NCI_H1930	10nM	Day7	71290544.0	0.00668453308478
DS30548	NCI_H1930	500nM	Day7	84151236.0	0.00762058919729
DS30613	NCI_H2081	DMSO	Day2	74116878.0	0.0144926233941
DS30621	NCI_H2081	10nM	Day2	86218502.0	0.0111120696576
DS30629	NCI_H2081	500nM	Day2	96156780.0	0.0159683383741
DS30659	NCI_H187	DMSO	Day2	142433190.0	0.00521982271127
DS30666	NCI_H187	10nM	Day2	100115800.0	0.0040855689112
DS30673	NCI_H187	500nM	Day2	99448164.0	0.00354910524039
DS30680	NCI_H2081	DMSO	Day7	81459828.0	0.0144985083936
DS30690	NCI_H2081	10nM	Day7	101556142.0	0.0158684838579
DS30698	NCI_H2081	500nM	Day7	96893632.0	0.0050167383549
DS30808	NCI_H187	DMSO	Day7	111507672.0	0.00829098109052

DS-ID	Cell Line	Drug	Time	Total Reads	Ribosomal Fraction
DS30816	NCI_H187	10nM	Day7	104178276.0	0.0245296821767
DS30824	NCI_H187	500nM	Day7	109684660.0	0.00668122598
DS30889	NCI_H187	DMSO	Day2	99625842.0	0.0150012483709
DS30896	NCI_H187	10nM	Day2	93996230.0	0.0082558204728
DS30903	NCI_H187	500nM	Day2	96962432.0	0.00907642250557
DS30920	NCI_H187	DMSO	Day7	99546104.0	0.00557799831121
DS30928	NCI_H187	10nM	Day7	105486542.0	0.00449731303165
DS30936	NCI_H187	500nM	Day7	70664696.0	0.00663876060544
DS31212	NCI_H774	None	None	76299864.0	0.0762298606456
DS31213	NCI_H345	None	None	103127952.0	0.00769760268293
DS31214	NCI_H2196	None	None	84537348.0	0.0103744442042
DS32528	DMS_114	None	None	72139342.0	0.0304706965583
DS33098	NCI_H526	DMSO	Day2	65247080.0	0.0307507401097
DS33106	NCI_H526	10nM	Day2	65764538.0	0.0199493532517
DS33114	NCI_H526	500nM	Day2	73018130.0	0.0264283678588
DS33307	NCI_H526	DMSO	Day7	109534874.0	0.0530430518412
DS33315	NCI_H526	10nM	Day7	83964436.0	0.00389614955551
DS33323	NCI_H526	500nM	Day7	89141122.0	0.0077355319804
DS33343	SHP_77	DMSO	Day2	74281592.0	0.00539245308582
DS33351	SHP_77	10nM	Day2	79170568.0	0.0379404881875
DS33359	SHP_77	500nM	Day2	66692402.0	0.012429121986
DS33501	SHP_77	DMSO	Day7	75629474.0	0.0457806172234
DS33509	SHP_77	10nM	Day7	77570132.0	0.0283156924369
DS33517	SHP_77	500nM	Day7	79059554.0	0.0419464040994
DS34132	NCI_H1417	DMSO	Day2	56606872.0	0.00508482433016
DS34139	NCI_H1417	10nM	Day2	68590038.0	0.0033342159688
DS34146	NCI_H1417	500nM	Day2	68991010.0	0.00494377455845
DS34378	NCI_H1417	10nM	Day7	84345894.0	0.00784400957325
DS34394	NCI_H1417	DMSO	Day7	90511914.0	0.00504539104101
DS34408	NCI_H1417	500nM	Day7	98243986.0	0.00570589633853
DS38200	MOLM-13	None	Day2	120089416.0	0.00686655017125
DS38208	MOLM-13	DMSO	Day2	107739700.0	0.0220323056404
DS38216	MOLM-13	10nM	Day2	81926342.0	0.00551004706154
DS38224	MOLM-13	500nM	Day2	82048666.0	0.00698049130988
DS38235	SKM-1	None	Day2	85386834.0	0.00582108478223
DS38243	SKM-1	DMSO	Day2	163169018.0	0.00134790294564
DS38251	SKM-1	10nM	Day2	76370020.0	0.00122459048721
DS38259	SKM-1	500nM	Day2	72178308.0	0.00859590668155
DS38273	NOMO-1	None	Day2	53794348.0	0.038951192419
DS38281	NOMO-1	DMSO	Day2	69869216.0	0.0317440230044
DS38289	NOMO-1	10nM	Day2	68342220.0	0.00900939419293
DS38297	NOMO-1	500nM	Day2	67649546.0	0.0060772913391
DS38331	SIG-M5	None	Day2	15976242.0	0.263830004578
DS38339	SIG-M5	DMSO	Day2	13707386.0	0.180044831305
DS38347	SIG-M5	10nM	Day2	148227614.0	0.0976362609466
DS38355	SIG-M5	500nM	Day2	103821198.0	0.172153359278
DS38412	MOLM-13	None	Day6	96243068.0	0.00989756477838
DS38420	MOLM-13	DMSO	Day6	82501540.0	0.0130867375324
DS38428	MOLM-13	10nM	Day6	89954314.0	0.00988713003803
DS38436	MOLM-13	500nM	Day6	76986304.0	0.00734709903725
DS38444	SKM-1	None	Day6	90851542.0	0.00624106082866
DS38452	SKM-1	DMSO	Day6	76216636.0	0.00596546927104
DS38460	SKM-1	10nM	Day6	57675878.0	0.00755716280557
DS38468	SKM-1	500nM	Day6	63297954.0	0.0136924804868
DS38476	OCI-AML2	None	Day2	69552642.0	0.00766363411472
DS38484	OCI-AML2	DMSO	Day2	91136006.0	0.0177370511497
DS38492	OCI-AML2	10nM	Day2	82924042.0	0.00970560504033
DS38500	OCI-AML2	500nM	Day2	81916998.0	0.00654821359542
DS38508	NOMO-1	None	Day6	115522316.0	0.0152005782156
DS38516	NOMO-1	DMSO	Day6	72375270.0	0.0153858355209
DS38524	NOMO-1	10nM	Day6	80341994.0	0.0357783253425
DS38532	NOMO-1	500nM	Day6	83770410.0	0.0213901304769
DS38551	SIG-M5	None	Day6	127145828.0	0.133013723423
DS38559	SIG-M5	DMSO	Day6	80680280.0	0.142370998217
DS38567	SIG-M5	10nM	Day6	64758996.0	0.0365496401457
DS38575	SIG-M5	500nM	Day6	75319078.0	0.0992395844251
DS38604	OCI-AML2	None	Day6	74670348.0	0.0191620909548
DS38620	OCI-AML2	10nM	Day6	64722260.0	0.00958106221878
DS38628	OCI-AML2	500nM	Day6	89325236.0	0.0157280077043
DS38636	KASUMI-1	None	Day2	66037386.0	0.0154764454184
DS38644	KASUMI-1	DMSO	Day2	96800464.0	0.0185683200857
DS38652	KASUMI-1	10nM	Day2	155162268.0	0.0119181552567

DS-ID	Cell Line	Drug	Time	Total Reads	Ribosomal Fraction
DS38660	KASUMI-1	500nM	Day2	119366330.0	0.0121701153081
DS38732	KASUMI-1	None	Day6	71320924.0	0.0106095092094
DS38740	KASUMI-1	DMSO	Day6	135236830.0	0.00874221911294
DS38748	KASUMI-1	10nM	Day6	70851662.0	0.00985244354607
DS38755	KASUMI-1	500nM	Day6	71271332.0	0.00888466627788
DS38952	UT-7	None	Day2	76547936.0	0.00610744096353
DS38960	UT-7	DMSO	Day2	70960314.0	0.00447692494709
DS38968	UT-7	10nM	Day2	66865366.0	0.00612005922468
DS38976	UT-7	500nM	Day2	63966738.0	0.00891353878323
DS38990	AML-193	None	Day2	219443010.0	0.00780496949983
DS38998	AML-193	DMSO	Day2	63357878.0	0.0103458326051
DS39006	AML-193	10nM	Day2	62326244.0	0.0100645564331
DS39014	AML-193	500nM	Day2	89785842.0	0.0367880049507
DS39053	UT-7	None	Day6	144791400.0	0.0089037194198
DS39061	UT-7	DMSO	Day6	145068772.0	0.00724897567893
DS39077	UT-7	500nM	Day6	71658060.0	0.00676144456046
DS39085	SKNO-1	None	Day2	82263282.0	0.0144556109492
DS39093	SKNO-1	DMSO	Day2	67084142.0	0.0497475543475
DS39101	SKNO-1	10nM	Day2	85495820.0	0.00211039557256
DS39109	SKNO-1	500nM	Day2	95104936.0	0.0212967705483
DS39117	AML-193	None	Day6	86199148.0	0.0153834003092
DS39124	AML-193	DMSO	Day6	82631100.0	0.00762151296546
DS39131	AML-193	10nM	Day6	91568458.0	0.0103655125436
DS39138	AML-193	500nM	Day6	98012584.0	0.0149601810314
DS39207	SKNO-1	None	Day6	120338926.0	0.0300598328425
DS39215	SKNO-1	DMSO	Day6	51132296.0	0.028925319528
DS39223	SKNO-1	10nM	Day6	97556232.0	0.0101048798195
DS39230	SKNO-1	500nM	Day6	119186066.0	0.0366011912836
DS39375	UT-7	10nM	Day6	61141332.0	0.00669004070765
DS42082	SIG-M5	None	Day2	77475796.0	0.0188353276164
DS42089	SIG-M5	DMSO	Day2	84069620.0	0.00930483568262
DS42096	SIG-M5	10nM	Day2	116348990.0	0.00573545159266
DS42103	SIG-M5	500nM	Day2	57483972.0	0.0519393127531
DS42137	SIG-M5	None	Day6	118832636.0	0.102413061005
DS42144	SIG-M5	DMSO	Day6	74342354.0	0.0330148544933
DS42151	SIG-M5	10nM	Day6	84186866.0	0.00759557910138
DS42158	SIG-M5	500nM	Day6	69107330.0	0.00280560687267
DS42503	GDM-1	None	None	95494544.0	0.00398902370799
DS42504	THP-1	None	None	76488502.0	0.00179987836603
DS42505	MV-4-11	None	None	75328340.0	0.00161331047518
DS42506	KG-1	None	None	98885838.0	0.0124950551564
DS42507	OCI-M1	None	None	80068398.0	0.00493805308806
DS42508	TF-1-CN5A-1	None	None	118433690.0	0.00269246022817
DS42618	UCSD_AML1	None	None	86304598.0	0.00585783390127
DS42619	GF-D8	None	Day2	69941148.0	0.00112588944065
DS42627	GF-D8	DMSO	Day2	81221238.0	0.00122564002287
DS42635	GF-D8	10nM	Day2	100116930.0	0.000956181936462
DS42643	GF-D8	500nM	Day2	68123052.0	0.000920305214746
DS42651	HL-60	None	None	73886514.0	0.0160509128905
DS42652	KASUMI-3	None	None	65490874.0	0.00207873237422
DS42658	OCI-M2	None	None	72320602.0	0.0024267773656
DS42831	GF-D8	None	Day6	133965264.0	0.000810598186109
DS42839	GF-D8	DMSO	Day6	70583142.0	0.00071521327288
DS42847	GF-D8	10nM	Day6	74563824.0	0.00123274256964
DS42855	GF-D8	500nM	Day6	137386458.0	0.000897715843289
DS54655	NCI_H2171	None	None	72970628.0	0.00858052091864

Supplementary Table #3: RNA-Seq Libraries
Table containing sequencing statistics and metadata for RNA-Seq experiments performed for Chapters #2-3.

DS_ID	Antibody	Cell Line	Drug	Time	Total Reads	Aligned Nuclear Reads	Nuclear Mapping Rate	Nuclear Duplicate Rate	Median Insert Size
DS32937	H3K4me1	NCI_H524	DMSO	Day7	123698668	105485632	0.852762876962	4.83947235582	201
DS32938	H3K4me2	NCI_H524	DMSO	Day2	38304428	33051342	0.862859562868	1.4467612238	168
DS32939	H3K4me1	NCI_H524	10nM	Day2	52315404	41393266	0.791225200134	3.30282708303	224
DS32940	H3K4me2	NCI_H524	10nM	Day2	45728870	38110568	0.833402793465	1.68955760512	222
DS32941	H3K4me1	NCI_H524	500nM	Day2	92484708	64040702	0.692446388002	4.35713212513	238
DS32942	H3K4me2	NCI_H524	500nM	Day2	36295872	30486618	0.839947253506	0.946566129441	216
DS32944	H3K4me1	NCI_H2029	DMSO	Day2	42924094	35509906	0.827272114351	2.5934171721	182
DS32945	H3K4me2	NCI_H2029	DMSO	Day2	45498158	38335574	0.842574198279	0.866531958019	190
DS32946	H3K4me1	NCI_H2029	500nM	Day2	37195864	30759248	0.826953448373	1.62292654229	191
DS32947	H3K4me2	NCI_H2029	10nM	Day2	35170102	28516650	0.810820793184	1.39614576046	205
DS32949	H3K4me2	NCI_H2029	500nM	Day2	42587362	30041038	0.705397953506	3.60078769582	213
DS33188	H3K4me1	NCI_H2029	DMSO	Day7	33314052	26949904	0.808965057748	1.2837225691	191
DS33189	H3K4me2	NCI_H2029	DMSO	Day7	37319340	32676210	0.875583812575	0.880071464836	201
DS33190	H3K4me1	NCI_H2029	10nM	Day7	38833008	31941230	0.822527835083	2.56854228845	210
DS33191	H3K4me2	NCI_H2029	10nM	Day7	38740824	32360374	0.835304225847	0.875515221178	176
DS33192	H3K4me1	NCI_H2029	500nM	Day7	48035160	36131186	0.752182068302	2.07617319841	202
DS33193	H3K4me2	NCI_H2029	500nM	Day7	45965588	40715074	0.88577293953	0.797272282988	193
DS33195	H3K4me1	NCI_H2081	DMSO	Day2	55718608	44912320	0.806056030689	3.68542083776	193
DS33196	H3K4me2	NCI_H2081	DMSO	Day2	41944994	35738684	0.852036931988	0.505267625411	185
DS33197	H3K4me1	NCI_H2081	10nM	Day2	35984142	29548324	0.821148493689	1.18745144395	201
DS33198	H3K4me2	NCI_H2081	10nM	Day2	31983548	28225886	0.882512659321	0.746449553435	198
DS33199	H3K4me1	NCI_H2081	500nM	Day2	66953674	53052296	0.792373186272	2.08692193077	202
DS33200	H3K4me2	NCI_H2081	500nM	Day2	34466328	28086460	0.814895627988	1.12068235014	192
DS33202	H3K4me1	NCI_H2081	DMSO	Day7	66971886	44993754	0.671830475253	5.17975450548	198
DS33203	H3K4me2	NCI_H2081	DMSO	Day7	54002748	43396722	0.803602105582	2.38848454959	206
DS33204	H3K4me1	NCI_H2081	10nM	Day7	73615550	52765076	0.716765357319	4.23488824313	192
DS33205	H3K4me2	NCI_H2081	10nM	Day7	40085668	32621792	0.813801880513	2.68023289462	188
DS33206	H3K4me1	NCI_H2081	500nM	Day7	72795970	45065454	0.619065231221	5.21789040448	213
DS33207	H3K4me2	NCI_H2081	500nM	Day7	40110522	30915686	0.770762494689	1.45112743091	193
DS34087	H3K4me1	NCI_H524	DMSO	Day7	49568460	33244004	0.670668485565	4.06248898298	227
DS34088	H3K4me2	NCI_H524	DMSO	Day7	41506710	29660274	0.714589857881	1.4541875102	235
DS34089	H3K4me1	NCI_H524	10nM	Day7	52291658	30164098	0.576843403971	5.03788974562	247
DS34090	H3K4me2	NCI_H524	10nM	Day7	42700850	32324568	0.75700057493	1.40013626787	232
DS34091	H3K4me1	NCI_H524	500nM	Day7	45469862	34448518	0.757612107994	1.22891788843	199
DS34092	H3K4me2	NCI_H524	500nM	Day7	42369592	38375270	0.905726682475	0.482513868958	219
DS34227	H3K4me2	NCI_H510A	DMSO	Day7	58987506	43608100	0.739276890262	3.63825527826	194
DS34228	H3K4me2	NCI_H510A	10nM	Day7	52683882	35932058	0.682031327912	2.1772312624	189
DS34229	H3K4me2	NCI_H510A	500nM	Day7	112373390	64960352	0.57807593061	2.90194856087	167
DS34230	H3K4me2	NCI_H1930	DMSO	Day7	93346218	59951192	0.642245538003	19.3005336741	169
DS34231	H3K4me2	NCI_H1930	10nM	Day7	110549952	65503466	0.592523694628	2.52483128145	174
DS34232	H3K4me2	NCI_H1930	500nM	Day7	69281228	47103388	0.679886736419	1.8476462882	174
DS34233	H3K4me2	NCI_H187	DMSO	Day7	69190360	45743972	0.661132157717	6.95096175732	170
DS34234	H3K4me2	NCI_H187	10nM	Day7	75553884	44730836	0.592038868578	0.765078479642	196
DS34235	H3K4me2	NCI_H187	500nM	Day7	72033494	45744844	0.635049633994	4.9656918712	177
DS34236	H3K4me2	NCI_H187	DMSO	Day7	53382886	34843564	0.652710383624	1.35260560602	209
DS34237	H3K4me2	NCI_H187	10nM	Day7	82274238	55490964	0.674463420737	6.29324803224	174
DS34238	H3K4me2	NCI_H187	500nM	Day7	73096852	46782952	0.640013225193	9.86879579553	167
DS34242	H3K4me1	NCI_H1930	DMSO	Day7	201290648	55814194	0.277281605254	11.4468230071	155
DS34243	H3K4me1	NCI_H1930	10nM	Day7	99516684	51333534	0.515828421293	25.0857655738	167
DS34244	H3K4me1	NCI_H1930	500nM	Day7	104425692	55621482	0.532641737246	26.1489400804	165
DS34246	H3K4me1	NCI_H187	10nM	Day7	141445932	64886216	0.458735115832	35.429416935	171

DS_ID	Antibody	Cell Line	Drug	Time	Total Reads	Aligned Nuclear Reads	Nuclear Mapping Rate	Nuclear Duplicate Rate	Median Insert Size
DS34247	H3K4me1	NCI_H187	500nM	Day7	140856166	41957046	0.297871560695	8.39252601339	192
DS34248	H3K4me1	NCI_H187	DMSO	Day7	67016058	39237640	0.585496091101	3.15749876904	174
DS34249	H3K4me1	NCI_H187	10nM	Day2	65448666	44130992	0.674284056454	4.65540860718	189
DS34250	H3K4me1	NCI_H187	500nM	Day2	125243768	48246990	0.385224676409	4.42831770438	206
DS34251	H3K4me2	NCI_H446	DMSO	Day7	63182516	47809888	0.756694905914	4.28415979556	185
DS34252	H3K4me2	NCI_H446	10nM	Day7	68862096	46394016	0.67372355323	3.24346139813	188
DS34253	H3K4me2	NCI_H446	500nM	Day7	54200608	38707634	0.714154977745	6.08267092739	196
DS34254	H3K4me2	NCI_H510A	DMSO	Day2	50666384	33012148	0.651559187646	13.2779908778	195
DS34255	H3K4me2	NCI_H510A	10nM	Day2	58766342	40410398	0.687645285119	3.15950859974	185
DS34256	H3K4me2	NCI_H510A	500nM	Day2	61347634	45381974	0.739751006534	1.7723468794	184
DS34257	H3K4me1	NCI_H510A	500nM	Day2	53555214	37198028	0.694573417259	5.56896188153	201
DS34258	H3K4me1	NCI_H510A	DMSO	Day7	60844558	40194170	0.660604190764	12.712216722	195
DS34259	H3K4me1	NCI_H510A	10nM	Day7	76991966	37974816	0.49323089113	12.5860780998	197
DS34261	H3K4me1	NCI_H446	DMSO	Day7	77933454	39099862	0.501708316431	2.40200336257	184
DS34262	H3K4me1	NCI_H446	10nM	Day7	68694632	38083374	0.554386462104	8.26357979732	188
DS34263	H3K4me1	NCI_H446	500nM	Day7	117492040	60523852	0.515131510186	10.8359725683	169
DS34264	H3K4me1	NCI_H510A	DMSO	Day2	125068332	64117854	0.512662581924	32.8415295995	165
DS34265	H3K4me1	NCI_H510A	10nM	Day2	86745576	48078216	0.554244011245	7.83047357664	211
DS34977	H3K4me1	SHP_77	DMSO	Day2	49524904	37707404	0.761382677289	4.61080269541	185
DS34978	H3K4me2	SHP_77	DMSO	Day2	34966138	29178300	0.834473055045	16.355990582	195
DS34979	H3K4me1	SHP_77	10nM	Day2	63076816	37986124	0.60222006133	8.51096047599	169
DS34980	H3K4me2	SHP_77	10nM	Day2	50939638	38173212	0.749381297134	3.78876160591	299
DS34981	H3K4me1	SHP_77	500nM	Day2	40539964	32837998	0.810015470167	29.7021152142	242
DS34982	H3K4me2	SHP_77	500nM	Day2	58369966	49577564	0.849367703932	4.77901253882	220
DS34984	H3K4me1	SHP_77	DMSO	Day7	37278830	31317646	0.840091977136	1.1043231027	167
DS34985	H3K4me2	SHP_77	DMSO	Day7	75303954	47253566	0.627504446845	18.7080526367	155
DS34986	H3K4me1	SHP_77	10nM	Day7	69959644	49782420	0.711587669028	21.0584459333	158
DS34987	H3K4me2	SHP_77	10nM	Day7	37843044	31826468	0.841012366764	21.3928608101	190
DS34988	H3K4me1	SHP_77	500nM	Day7	52714864	37288256	0.707357530127	11.9500895939	171
DS34989	H3K4me2	SHP_77	500nM	Day7	35441136	27443290	0.774334377995	8.39531266113	184
DS34991	H3K4me1	NCI_H187	DMSO	Day2	54813108	39462264	0.719942098521	18.5589047805	173
DS34992	H3K4me2	NCI_H187	DMSO	Day2	48841116	41536432	0.850439863004	3.9244584128	160
DS34993	H3K4me1	NCI_H187	10nM	Day2	39980012	33804974	0.845546869771	12.8014356704	173
DS34994	H3K4me2	NCI_H187	10nM	Day2	40064906	29399902	0.733806838334	7.73752919312	166
DS34995	H3K4me1	NCI_H187	500nM	Day2	52043600	34810234	0.668866757872	14.9700286416	194
DS34996	H3K4me2	NCI_H187	500nM	Day2	64541180	50098358	0.776223149313	10.6576946095	172
DS35062	H3K4me2	NCI_H1417	DMSO	Day7	98543562	82445734	0.836642519579	10.9231437008	148
DS35064	H3K4me2	NCI_H1417	10nM	Day7	70484300	55598646	0.788808940431	18.9934733303	158
DS35066	H3K4me2	NCI_H1417	500nM	Day7	78097958	64762922	0.829252437048	6.3506028959	176
DS35069	H3K4me2	NCI_H446	DMSO	Day2	75979334	58354060	0.768025421228	18.0861074619	197
DS35071	H3K4me2	NCI_H446	10nM	Day2	73559998	47672482	0.64807617314	39.7162035742	191
DS35073	H3K4me2	NCI_H446	500nM	Day2	115491650	88188542	0.763592363604	6.06411204757	158
DS35075	H3K4me1	NCI_H1930	DMSO	Day2	48632468	36538574	0.751320578672	13.4290571931	309
DS35076	H3K4me2	NCI_H1930	DMSO	Day2	42666696	32685302	0.766061238958	23.8126054335	302
DS35077	H3K4me1	NCI_H1930	10nM	Day2	46396044	37447884	0.807135280758	11.0990730478	287
DS35078	H3K4me2	NCI_H1930	10nM	Day2	110174388	89404756	0.811484026578	4.43001264944	157
DS35079	H3K4me1	NCI_H1930	500nM	Day2	69556914	55624176	0.799692982354	9.67257834076	253
DS35080	H3K4me2	NCI_H1930	500nM	Day2	45057934	34053874	0.755779747913	16.9818916931	286
DS35157	H3K4me2	NCI_H526	DMSO	Day2	60614746	48987392	0.808176149084	4.97057283637	147
DS35159	H3K4me2	NCI_H526	10nM	Day2	119419088	84686734	0.709155759086	16.2562698427	196
DS35161	H3K4me2	NCI_H526	500nM	Day2	54571512	43355630	0.7944736807	7.41572893763	209

DS_ID	Antibody	Cell Line	Drug	Time	Total Reads	Aligned Nuclear Reads	Nuclear Mapping Rate	Nuclear Duplicate Rate	Median Insert Size
DS35163	H3K4me1	NCI_H526	DMSO	Day7	45198000	40214016	0.889729988053	20.5848876173	234
DS35164	H3K4me2	NCI_H526	DMSO	Day7	50976866	45752776	0.897520377184	25.1107604924	227
DS35165	H3K4me1	NCI_H526	10nM	Day7	42636550	30574868	0.717104643786	5.67452327186	166
DS35166	H3K4me2	NCI_H526	10nM	Day7	124314530	101476528	0.816288554524	6.3855712525	177
DS35167	H3K4me1	NCI_H526	500nM	Day7	51650316	45177558	0.874681153935	27.9679968537	211
DS35168	H3K4me2	NCI_H526	500nM	Day7	55326768	45719086	0.826346588689	3.6714119788	150
DS35270	H3K4me1	NCI_H1417	DMSO	Day2	56798584	51257508	0.902443413026	62.2597766555	261
DS35271	H3K4me2	NCI_H1417	DMSO	Day2	34134028	31399106	0.91987696266	32.7557096689	267
DS35273	H3K4me2	NCI_H1417	10nM	Day2	63204342	58738000	0.929334886518	50.8505771392	228
DS35274	H3K4me1	NCI_H1417	500nM	Day2	59805906	48443250	0.810007794214	14.0421297085	153
DS35275	H3K4me2	NCI_H1417	500nM	Day2	74826758	66399780	0.887380153501	8.66882089067	172
DS35277	H3K4me1	NCI_H1417	10nM	Day7	39103130	32459380	0.830096721158	16.7528338496	159
DS36195	H3K4me1	NCI_H2029	500nM	Day2	54644810	39421550	0.721414348407	14.9837233696	154
DS36198	H3K4me1	NCI_H1417	DMSO	Day7	88896528	71901252	0.808819575046	15.6895459901	152
DS36199	H3K4me1	NCI_H1417	500nM	Day7	53916110	39863622	0.739363837636	51.9972871507	172
DS36249	H3K4me1	NCI_H446	DMSO	Day2	37255394	26550380	0.712658682391	17.1004030827	174

Supplementary Table #4: Chip-Seq Libraries
Table containing sequencing statistics and metadata for histone ChIP-Seq experiments performed for Chapters #2-3.

Appendix A2 : SMARCA4 Tables

DS-ID	clone_ID	Clone	Total Reads	Aligned Nuclear Reads	Nuclear Mapping Rate	Nuclear Duplicate Rate	SPOT	Hotspot Number	Hotspot Coverage (Bp)
DS62068	A3	WT	82100174	35757728	0.435537786	13.62931112	0.6954	73242	30957873
DS62113	A3	WT	81576474	42730132	0.523804596	8.379403087	0.5805	72367	31206386
DS62073	B8	WT	260458804	127451240	0.489333584	14.63560496	0.5351	89616	43074752
DS62118	B8	WT	133150566	66290020	0.497857591	9.507675514	0.6293	84140	37828699
DS62078	E9	Rescue	113297102	54913020	0.484681594	6.284986694	0.5814	96058	41489731
DS62123	E9	Rescue	93097004	44993816	0.483300365	5.740877813	0.5462	83838	32038179
DS62128	C12	Rescue	147458648	69925906	0.474206884	6.555075597	0.5629	112363	46108620
DS62149	C12	Rescue	352400134	192455072	0.546126557	12.87358226	0.5018	133135	58010819
DS62133	F3	Rescue	138212502	81486932	0.589577143	7.610697627	0.5042	115540	45743470
DS62154	F3	Rescue	127891098	84769052	0.662822146	4.467566772	0.3775	113470	43047811

Supplementary Table #1: DNaseI-Seq Library Statistics
Table containing sequencing statistics and metadata for DNaseI-Seq experiments performed for Chapter #4

DS-ID	clone_ID	Clone	Total Reads	Ribosomal Fraction	Nuclear Duplicate Rate
DS62081	A3	WT	37741048	0.002260986	19.13560566
DS62136	A3	WT	21387064	0.003123009	19.07689757
DS62082	B8	WT	19496110	0.002126783	13.00062813
DS62137	B8	WT	23606732	0.001931907	15.14737593
DS62083	E9	Rescue	19386874	0.001954209	14.73615514
DS62138	E9	Rescue	42126714	0.00217838	17.044498
DS62157	C12	Rescue	30192364	0.002025082	14.87589216
DS62139	C12	Rescue	43113384	0.0023672	18.32239776
DS62158	F3	Rescue	9700602	0.00273715	11.39960602
DS62140	F3	Rescue	26039012	0.003401819	16.72635612

Supplementary Table #2: RNA-Seq Library Statistics
Table containing sequencing statistics and metadata for RNA-Seq experiments performed for Chapter #4

	Antibody	clone_ID	Clone	Total Reads	Aligned Nuclear Reads	Nuclear Mapping Rate	Nuclear Duplicate Rate	Median Insert Size
DS65989	H3K27me3	A3	WT	14356010	12571940	0.875726612	1.805671997	175
DS65990	H3K27me3	B8	WT	16059896	13983776	0.870726436	1.705791054	168
DS65991	H3K27me3	C12	Rescue	17835566	15638850	0.876835083	1.873795068	166
DS65992	H3K27me3	E9	Rescue	15286732	13534414	0.885370006	1.938539785	165
DS65993	H3K27me3	F3	Rescue	15535854	13539026	0.871469698	1.924820884	171
DS65994	H3K4me1	A3	WT	13275202	11917610	0.897734739	1.869082811	172
DS65995	H3K4me1	B8	WT	13594488	12273226	0.902808991	1.909310559	168
DS65996	H3K4me1	C12	Rescue	12992438	11750824	0.904435642	1.87384306	167
DS65997	H3K4me1	E9	Rescue	10380306	9257664	0.891848853	1.749404601	163
DS65998	H3K4me1	F3	Rescue	14371786	13020186	0.905954625	1.938927754	164
DS65999	H3K4me2	A3	WT	17061756	15585638	0.913483817	2.581966808	164
DS66000	H3K4me2	B8	WT	12826120	11620962	0.906038771	2.14302396	158
DS66001	H3K4me2	C12	Rescue	20311292	18620400	0.916751135	2.760026637	157
DS66002	H3K4me2	E9	Rescue	18610562	17018368	0.914446753	3.227359991	152
DS66003	H3K4me2	F3	Rescue	20426588	18684266	0.914703229	2.832019197	160
DS66514	SMARCA4	A3	WT	21210270	15387936	0.725494584	43.00862702	111
DS66515	SMARCA4	B8	WT	19863482	13661520	0.687770654	54.78930602	101
DS66516	SMARCA4	C12	Rescue	24714828	20029624	0.810429431	22.73058146	88
DS66517	SMARCA4	E9	Rescue	24659906	20466306	0.82994258	15.15720521	89
DS66518	SMARCA4	F3	Rescue	25808664	21141410	0.819159411	20.08004197	87
DS66519	SMARCA2	A3	WT	22266478	18083030	0.812119007	20.18917184	92
DS66520	SMARCA2	B8	WT	24096138	20317310	0.843177027	23.36750288	101
DS66521	SMARCA2	C12	Rescue	25568084	18805740	0.735516201	33.93028937	71
DS66522	SMARCA2	E9	Rescue	24379774	19673216	0.806948251	23.72461117	90
DS66523	SMARCA2	F3	Rescue	31411738	24103992	0.767356203	8.826977706	112

Supplementary Table #3: CUT&RUN Library Statistics
Table containing sequencing statistics and metadata for CUT&RUN experiments performed for Chapter #4

TF	Peak-Number	File-Name
SREBF1-human	3429	ENCFF624DDK
YY1-human	17078	ENCFF613DTQ
FOXA1-human	33874	ENCFF297HAX
POLR2AphosphoS2	6272	ENCFF156MIR
PHF8-human	17048	ENCFF907WHF
TCF12-human	31794	ENCFF228CDD
HDAC2-human	4167	ENCFF814DAF
EHMT2-human	2352	ENCFF199OOU
ETS1-human	9988	ENCFF896WFR
HES2-human	3242	ENCFF558XCJ
SREBF2-human	838	ENCFF483YCC
ELF1-human	11737	ENCFF935ZUW
CHD2-human	3440	ENCFF310IDS
MAZ-human	4323	ENCFF661NNJ
CBX8-human	2819	ENCFF330OCU
CEBPB-human	47003	ENCFF047UIF
CHD4-human	3542	ENCFF766YPH
RAD21-human	26062	ENCFF897QCA
CBX2-human	141	ENCFF208AXT
NR3C1-human	665	ENCFF963CGV
CTCF-human	43844	ENCFF535MZG
SIN3A-human	38272	ENCFF567BJI
FOSL2-human	33138	ENCFF808RWZ
TAF1-human	17246	ENCFF886KDK
JUND-human	21802	ENCFF587VEY
RFX5-human	6479	ENCFF179WDI
ZFP36-human	11426	ENCFF137JHO
ELK1-human	470	ENCFF605JXG
SP1-human	43742	ENCFF404OSB
RNF2-human	4801	ENCFF110EOX
KDM5A-human	5209	ENCFF149INM
MYC-human	9437	ENCFF542GMN
JUNB-human	11264	ENCFF565QYS
ZC3H11A-human	7858	ENCFF415SIS
SIX5-human	8553	ENCFF189NMX
ATF3-human	11014	ENCFF851UTY

REST-human	9886	ENCFF706DRE
USF2-human	11956	ENCFF593EOW
SMC3-human	23810	ENCFF256LDD
BCL3-human	12352	ENCFF093ZAB
GABPA-human	17425	ENCFF520GJC
POLR2A-human	31834	ENCFF664KTN
RCOR1-human	862	ENCFF993WZP
MAFK-human	70080	ENCFF813WJW
ESRRA-human	2631	ENCFF558UWY
KDM1A-human	7245	ENCFF316CBQ
EP300-human	4797	ENCFF727TYG
CREB1-human	3289	ENCFF576PUH
JUN-human	1767	ENCFF127HJG
ZBTB33-human	11349	ENCFF593ZJA
NFE2L2-human	7256	ENCFF418TUX

Supplementary Table #4: ENCODE TF ChIP-Seq Data Used
Table of ENCODE transcription factor ChIP-Seq experiments analyzed for overlap with SMARCA4 remodeled DHSs in Chapter #4.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                RNA_LFR   R-squared:                       0.246
Model:                            OLS   Adj. R-squared:                  0.245
Method:                 Least Squares   F-statistic:                     414.9
Date:                Tue, 23 Apr 2019   Prob (F-statistic):               0.00
Time:                        20:57:35   Log-Likelihood:                -13494.
No. Observations:               16544   AIC:                         2.702e+04
Df Residuals:                   16530   BIC:                         2.712e+04
Df Model:                          13
Covariance Type:            nonrobust
================================================================================================================
                                                   coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------------------------
Intercept                                       -0.0039      0.010     -0.383      0.702      -0.024       0.016
H3K4me3_Peak[T.True]                             0.0530      0.018      3.019      0.003       0.019       0.087
H3K27me3_Peak[T.True]                           -0.0047      0.024     -0.196      0.845      -0.051       0.042
Bivalent[T.True]                                -0.0814      0.028     -2.925      0.003      -0.136      -0.027
Methylation_Fraction                            -0.0012      0.010     -0.119      0.905      -0.021       0.019
CpG_Island                                       0.0395      0.019      2.125      0.034       0.003       0.076
DHS_Score                                        0.5440      0.018     30.996      0.000       0.510       0.578
H3K4me3_Peak[T.True]:DHS_Score                  -0.0729      0.035     -2.096      0.036      -0.141      -0.005
H3K27me3_Peak[T.True]:DHS_Score                  0.0538      0.040      1.360      0.174      -0.024       0.131
Bivalent[T.True]:DHS_Score                       0.4889      0.047     10.387      0.000       0.397       0.581
Methylation_Fraction:DHS_Score                  -0.1010      0.020     -4.994      0.000      -0.141      -0.061
CpG_Island:DHS_Score                            -0.0478      0.057     -0.845      0.398      -0.159       0.063
Methylation_Fraction:CpG_Island                  0.0463      0.019      2.416      0.016       0.009       0.084
Methylation_Fraction:CpG_Island:DHS_Score        0.0338      0.057      0.596      0.551      -0.077       0.145
==============================================================================
Omnibus:                     2444.944   Durbin-Watson:                   2.034
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            39894.936
Skew:                          -0.017   Prob(JB):                         0.00
Kurtosis:                      10.607   Cond. No.                         32.9
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Supplementary Table #5: Promoter State Regression
Results of multiple regression to study the relationship between genes' promoter states and changes in gene expression after SMARCA4 reactivation. Linear regression was performed on genes' change in expression with the genes' weighted change in chromatin accessibility, promoter state, and interaction terms between the two as potential explanatory variables. The promoter state for each gene was characterized by CpG island status, CpG methylation, overlap with H3K27me3 peaks, overlap with H3K4me3 peaks, and bivalency(both H3K4me3 and H3K27me3 peaks) in SMARCA4$^{-/-}$ A549s.

References

Abate-Shen, Cory (2002). "Deregulated homeobox gene expression in cancer: cause or consequence?" In: *Nature Reviews Cancer* 2.10, pp. 777–785. ISSN: 1474-1768. DOI: 10.1038/nrc907.

Abranches, Elsa et al. (July 2014). "Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency". In: *Development* 141.14, 2770 LP –2779. DOI: 10.1242/dev.108910.

Adamo, Antonio et al. (May 2011). "LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells". In: *Nature Cell Biology* 13, p. 652.

Alon, Uri (June 2007). "Network motifs: theory and experimental approaches". In: *Nat Rev Genet* 8.6, pp. 450–461. ISSN: 1471-0056.

Alver, Burak H et al. (Mar. 2017). "The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers". In: *Nature Communications* 8, p. 14648.

Anders, Simon, Paul Theodor Pyl, and Wolfgang Huber (Jan. 2015). "HTSeq—a Python framework to work with high-throughput sequencing data". In: *Bioinformatics* 31.2, pp. 166–169. ISSN: 1367-4803.

Attanasio, Catia et al. (Apr. 2014). "Tissue-specific SMARCA4 binding at active and repressed regulatory elements during embryogenesis". In: *Genome Research*. DOI: 10.1101/gr.168930.113.

Audergon, Pauline N C B et al. (Apr. 2015). "Epigenetics. Restricted epigenetic inheritance of H3K9 methylation". eng. In: *Science (New York, N.Y.)* 348.6230, pp. 132–135. ISSN: 1095-9203. DOI: 10.1126/science.1260638.

Augustyn, Alexander et al. (Oct. 2014). "ASCL1 is a lineage oncogene providing therapeutic targets for high-grade neuroendocrine lung cancers". In: *Proceedings of the National Academy of Sciences of the United States of America* 111.41, pp. 14788–14793. ISSN: 0027-8424. DOI: 10.1073/pnas.1410419111.

Bailey, Timothy L et al. (July 2009). "MEME Suite: tools for motif discovery and searching". In: *Nucleic Acids Research* 37.Web Server issue, W202–W208. ISSN: 0305-1048. DOI: 10.1093/nar/gkp335.

Bao, Xiaomin et al. (2015). "A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63". In: *Genome Biology* 16.1, p. 284. ISSN: 1474-760X. DOI: 10.1186/s13059-015-0840-9.

Barretina, Jordi et al. (Mar. 2012). "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity". In: *Nature* 483, p. 603.

Basta, Jeannine and Michael Rauchman (Jan. 2015). "The nucleosome remodeling and deacetylase complex in development and disease". eng. In: *Translational research : the journal of laboratory and clinical medicine* 165.1, pp. 36–47. ISSN: 1878-1810. DOI: 10.1016/j.trsl.2014.05.003.

Baylin, Stephen B and Peter A Jones (2011). "A decade of exploring the cancer epigenome - biological and translational implications." In: *Nature reviews. Cancer* 11.10, pp. 726–34. ISSN: 1474-1768. DOI: 10.1038/nrc3130.A.

Beard, J (1902). "EMBRYOLOGICAL ASPECTS AND ETIOLOGY OF CARCINOMA." In: *The Lancet* 159.4112, pp. 1758–1761. ISSN: 0140-6736. DOI: https://doi.org/10.1016/S0140-6736(01)89936-0.

Behjati, Sam et al. (Dec. 2013). "Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone". eng. In: *Nature genetics* 45.12, pp. 1479–1482. ISSN: 1546-1718. DOI: 10.1038/ng.2814.

Bhagwat, Anand S. et al. (Apr. 2016). "BET Bromodomain Inhibition Releases the Mediator Complex from Select cis-Regulatory Elements". In: *Cell Reports* 15.3, pp. 519–530. ISSN: 2211-1247. DOI: 10.1016/J.CELREP.2016.03.054.

Blattler, Adam et al. (2014). "Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes". In: *Genome Biology* 15.9, p. 469. ISSN: 1474-760X. DOI: 10.1186/s13059-014-0469-0.

Borczuk, Alain C. et al. (2003). "Non-Small-Cell Lung Cancer Molecular Signatures Recapitulate Lung Developmental Pathways". In: *The American Journal of Pathology* 163.5, pp. 1949–1960. ISSN: 00029440. DOI: 10.1016/S0002-9440(10)63553-5.

Borges, Michael et al. (Apr. 1997). "An achaete-scute homologue essential for neuroendocrine differentiation in the lung". In: *Nature* 386.6627, pp. 852–855.

Borromeo, Mark D. et al. (2016). "ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs". In: *Cell Reports* 16.5, pp. 1259–1272. ISSN: 22111247. DOI: 10.1016/j.celrep.2016.06.081.

Bossen, Claudia et al. (May 2015). "The chromatin remodeler Brg1 activates enhancer repertoires to establish B cell identity and modulate cell growth". In: *Nature Immunology* 16, p. 775.

Boulay, Gaylor et al. (2017). "Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain". In: *Cell* 171.1, 163–178.e19. ISSN: 0092-8674. DOI: https://doi.org/10.1016/j.cell.2017.07.036.

Brandeis, Michael et al. (1994). "Spl elements protect a CpG island from de novo methylation". In: *Nature* 371.6496, pp. 435–438. ISSN: 1476-4687. DOI: 10.1038/371435a0.

Brennecke, Philip et al. (Nov. 2013). "Accounting for technical noise in single-cell RNA-seq experiments". In: *Nat Meth* 10.11, pp. 1093–1095. ISSN: 1548-7091.

Bultman, Scott et al. (Dec. 2000). "A Brg1 Null Mutation in the Mouse Reveals Functional Differences among Mammalian SWI/SNF Complexes". In: *Molecular Cell* 6.6, pp. 1287–1295. ISSN: 1097-2765. DOI: 10.1016/S1097-2765(00)00127-1.

Burns, L G and C L Peterson (Aug. 1997). "The yeast SWI-SNF complex facilitates binding of a transcriptional activator to nucleosomal sites in vivo". eng. In: *Molecular and cellular biology* 17.8, pp. 4811–4819. ISSN: 0270-7306.

Butler, Jennifer E F and James T Kadonaga (Oct. 2001). "Enhancer–promoter specificity mediated by DPE or TATA core promoter motifs". In: *Genes & Development* 15.19, pp. 2515–2519. DOI: 10.1101/gad.924301.

Carney, Desmond N et al. (June 1985). "Establishment and Identification of Small Cell Lung Cancer Cell Lines Having Classic and Variant Features". In: *Cancer Research* 45.6, 2913 LP –2923.

Carone, Dawn M and Jeanne B Lawrence (2013). "Heterochromatin instability in cancer: From the Barr body to satellites and the nuclear periphery". In: *Seminars in Cancer Biology* 23.2, pp. 99–108. ISSN: 1044-579X. DOI: https://doi.org/10.1016/j.semcancer.2012.06.008.

Castro, Diogo S et al. (May 2011). "A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets". In: *Genes & Development* 25.9, pp. 930–945. DOI: 10.1101/gad.627811.

Cavalli, Giacomo and Renato Paro (Oct. 1999). "Epigenetic Inheritance of Active Chromatin After Removal of the Main Transactivator". In: *Science* 286.5441, 955 LP –958. DOI: 10.1126/science.286.5441.955.

Cermak, Tomas et al. (July 2011). "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting". eng. In: *Nucleic acids research* 39.12, e82–e82. ISSN: 1362-4962. DOI: 10.1093/nar/gkr218.

Chanda, Soham et al. (2014). "Generation of Induced Neuronal Cells by the Single Reprogramming Factor ASCL1". In: *Stem Cell Reports* 3.2, pp. 282–296. ISSN: 2213-6711. DOI: https://doi.org/10.1016/j.stemcr.2014.05.020.

Chang, Hannah H et al. (May 2008). "Transcriptome-wide noise controls lineage choice in mammalian progenitor cells". eng. In: *Nature* 453.7194, pp. 544–547. ISSN: 1476-4687. DOI: 10.1038/nature06965.

Christensen, Camilla L. et al. (2014). "Targeting Transcriptional Addictions in Small Cell Lung Cancer with a Covalent CDK7 Inhibitor". In: *Cancer Cell* 26.6, pp. 909–922. ISSN: 1535-6108. DOI: https://doi.org/10.1016/j.ccell.2014.10.019.

Coleman, Rory T and Gary Struhl (Apr. 2017). "Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene". In: *Science* 356.6333, eaai8236. DOI: 10.1126/science.aai8236.

Conery, Andrew R et al. (2016). "Bromodomain inhibition of the transcriptional coactivators CBP/EP300 as a therapeutic strategy to target the IRF4 network in multiple myeloma". In: *eLife* 5. Ed. by Scott A Armstrong, e10483. ISSN: 2050-084X. DOI: 10.7554/eLife.10483.

Consortium, The ENCODE Project et al. (Sept. 2012). "An integrated encyclopedia of DNA elements in the human genome". In: *Nature* 489, p. 57.

Costa, Yael et al. (Mar. 2013). "NANOG-dependent function of TET1 and TET2 in establishment of pluripotency". eng. In: *Nature* 495.7441, pp. 370–374. ISSN: 1476-4687. DOI: 10.1038/nature11925.

Courey, Albert J et al. (1989). "Synergistic activation by the glutamine-rich domains of human transcription factor Sp1". In: *Cell* 59.5, pp. 827–836. ISSN: 0092-8674. DOI: https://doi.org/10.1016/0092-8674(89)90606-5.

Crocker, Justin et al. (2015). "Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness". In: *Cell* 160.1, pp. 191–203. ISSN: 0092-8674. DOI: https://doi.org/10.1016/j.cell.2014.11.041.

Cusan, Monica et al. (Jan. 2018). "LSD1 inhibition exerts its anti-leukemic effect by recommissioning PU.1- and C/EBPα-dependent enhancers in AML". In: *Blood*, blood–2017–09–807024. DOI: 10.1182/blood-2017-09-807024.

Daniel, Vincent C et al. (Apr. 2009). "A Primary Xenograft Model of Small-Cell Lung Cancer Reveals Irreversible Changes in Gene Expression Imposed by Culture ¡em¿In vitro¡/em¿". In: *Cancer Research* 69.8, 3364 LP –3373. DOI: 10.1158/0008-5472.CAN-08-4210.

Davis, Robert L, Harold Weintraub, and Andrew B Lassar (Dec. 1987). "Expression of a single transfected cDNA converts fibroblasts to myoblasts". In: *Cell* 51.6, pp. 987–1000. ISSN: 0092-8674. DOI: 10.1016/0092-8674(87)90585-X.

Dawson, Mark A. and Tony Kouzarides (Sept. 2016). "Cancer Epigenetics: From Mechanism to Therapy". In: *Cell* 150.1, pp. 12–27. ISSN: 0092-8674. DOI: 10.1016/j.cell.2012.06.013.

Delmore, Jake E et al. (Sept. 2011). "BET bromodomain inhibition as a therapeutic strategy to target c-Myc". eng. In: *Cell* 146.6, pp. 904–917. ISSN: 1097-4172. DOI: 10.1016/j.cell.2011.08.017.

Denny, Sarah K. et al. (2016). "Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility". In: *Cell* 166.2, pp. 328–342. ISSN: 10974172. DOI: 10.1016/j.cell.2016.05.052.

Di Croce, Luciano et al. (Feb. 2002). "Methyltransferase Recruitment and DNA Hypermethylation of Target Promoters by an Oncogenic Transcription Factor". In: *Science* 295.5557, 1079 LP –1082.

Dobin, Alexander et al. (Jan. 2013). "STAR: ultrafast universal RNA-seq aligner". In: *Bioinformatics* 29.1, pp. 15–21. ISSN: 1367-4803.

Dooley, Alison L et al. (July 2011). "Nuclear factor I/B is an oncogene in small cell lung cancer". In: *Genes & Development* 25.14, pp. 1470–1475. DOI: 10.1101/gad.2046711.

Dorighi, Kristel M et al. (May 2017). "Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation". eng. In: *Molecular cell* 66.4, 568–576.e4. ISSN: 1097-4164. DOI: 10.1016/j.molcel.2017.04.018.

Dowen, Jill M. et al. (Oct. 2014). "Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes". In: *Cell* 159.2, pp. 374–387. ISSN: 0092-8674. DOI: 10.1016/j.cell.2014.09.030.

Drexler, Hans G, Yoshinobu Matsuo, and Roderick A F MacLeod (2000). "Continuous hematopoietic cell lines as model systems for leukemia–lymphoma research". In: *Leukemia Research* 24.11, pp. 881–911. ISSN: 0145-2126. DOI: https://doi.org/10.1016/S0145-2126(00)00070-9.

Du, Jiamu et al. (Aug. 2015). "DNA methylation pathways and their crosstalk with histone methylation". In: *Nature Reviews Molecular Cell Biology* 16, p. 519.

Dyson, H Jane and Peter E Wright (2005). "Intrinsically unstructured proteins and their functions". In: *Nature Reviews Molecular Cell Biology* 6.3, pp. 197–208. ISSN: 1471-0080. DOI: 10.1038/nrm1589.

Elgin, Sarah C R and Gunter Reuter (2013). "Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila". eng. In: *Cold Spring Harbor perspectives in biology* 5.8, a017780–a017780. ISSN: 1943-0264. DOI: 10.1101/cshperspect.a017780.

Fan, Yuhong et al. (2005). "Histone H1 Depletion in Mammals Alters Global Chromatin Structure but Causes Specific Changes in Gene Regulation". In: *Cell* 123.7, pp. 1199–1212. ISSN: 0092-8674. DOI: https://doi.org/10.1016/j.cell.2005.10.028.

Fandy, Tamer E et al. (Sept. 2009). "Early epigenetic changes and DNA damage do not predict clinical response in an overlapping schedule of 5-azacytidine and entinostat in

patients with myeloid malignancies". In: *Blood* 114.13, pp. 2764–2773. ISSN: 0006-4971. DOI: 10.1182/blood-2009-02-203547.

Farley, Emma K et al. (Oct. 2015). "Suboptimization of developmental enhancers". In: *Science* 350.6258, 325 LP –328.

Felsenfeld, G et al. (Sept. 1996). "Chromatin structure and gene expression". In: *Proceedings of the National Academy of Sciences* 93.18, 9384 LP –9388. DOI: 10.1073/pnas.93.18.9384.

Fisher, William W et al. (Dec. 2012). "DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.52, pp. 21330–5. ISSN: 1091-6490. DOI: 10.1073/pnas.1209589110.

Funato, Kosuke et al. (Dec. 2014). "Use of human embryonic stem cells to model pediatric gliomas with H3.3K27M histone mutation". In: *Science* 346.6216, 1529 LP –1533. DOI: 10.1126/science.1253799.

Galas, D J and A Schmitz (Sept. 1978). "DNAse footprinting: a simple method for the detection of protein-DNA binding specificity". eng. In: *Nucleic acids research* 5.9, pp. 3157–3170. ISSN: 0305-1048.

Galloway, Jenna L et al. (Jan. 2005). "Loss of Gata1 but Not Gata2 Converts Erythropoiesis to Myelopoiesis in Zebrafish Embryos". In: *Developmental Cell* 8.1, pp. 109–116. ISSN: 1534-5807. DOI: 10.1016/j.devcel.2004.12.001.

Garraway, Levi A and William R Sellers (Aug. 2006). "Lineage dependency and lineage-survival oncogenes in human cancer". In: *Nat Rev Cancer* 6.8, pp. 593–602. ISSN: 1474-175X.

Garraway, Levi A, Hans R Widlund, et al. (July 2005). "Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma". In: *Nature* 436.7047, pp. 117–122. ISSN: 0028-0836.

Gaspar-Maia, Alexandre et al. (July 2009). "Chd1 regulates open chromatin and pluripotency of embryonic stem cells". In: *Nature* 460, p. 863.

George, Julie et al. (July 2015). "Comprehensive genomic profiles of small cell lung cancer". In: *Nature* 524, p. 47.

Gerstein, Mark B et al. (Sept. 2012). "Architecture of the human regulatory network derived from ENCODE data". In: *Nature* 489, p. 91.

Glaros, Selina et al. (May 2008). "Targeted Knockout of BRG1 Potentiates Lung Cancer Development". In: *Cancer Research* 68.10, 3689 LP –3696. DOI: 10.1158/0008-5472.CAN-07-6652.

Goodison, Steve et al. (July 2005). "The RhoGAP protein DLC-1 functions as a metastasis suppressor in breast cancer cells". eng. In: *Cancer research* 65.14, pp. 6042–6053. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-04-3043.

Graf, Thomas and Tariq Enver (Dec. 2009). "Forcing cells to change lineages ". In: *Nature* 462.7273, pp. 587–594. ISSN: 0028-0836.

Gross, David S and William T Garrard (June 1988). "NUCLEASE HYPERSENSITIVE SITES IN CHROMATIN". In: *Annual Review of Biochemistry* 57.1, pp. 159–197. ISSN: 0066-4154. DOI: 10.1146/annurev.bi.57.070188.001111.

Groudine, Mark and Harold Weintraub (Aug. 1982). "Propagation of globin DNAase i-hypersensitive sites in absence of factors required for induction: A possible mechanism for determination". In: *Cell* 30.1, pp. 131–139. ISSN: 0092-8674. DOI: 10.1016/0092-8674(82)90019-8.

Hakim, Ofir et al. (May 2011). "Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements". In: *Genome Research* 21.5, pp. 697–706. DOI: 10.1101/gr.111153.110.

Hanahan, Douglas and Robert A Weinberg (Jan. 2000). "The Hallmarks of Cancer". In: *Cell* 100.1, pp. 57–70. ISSN: 0092-8674. DOI: 10.1016/S0092-8674(00)81683-9.

Hanahan, Douglas and Robert A. Weinberg (2011). "Hallmarks of Cancer: The Next Generation". In: *Cell* 144.5, pp. 646–674. ISSN: 00928674. DOI: 10.1016/j.cell.2011.02.013.

Harris, William J. et al. (2012). "The Histone Demethylase KDM1A Sustains the Oncogenic Potential of MLL-AF9 Leukemia Stem Cells". In: *Cancer Cell* 21.4, pp. 473–487. ISSN: 15356108. DOI: 10.1016/j.ccr.2012.03.014.

Hartman, Mariusz L and Malgorzata Czyz (2015). "MITF in melanoma: mechanisms behind its expression and activity". eng. In: *Cellular and molecular life sciences : CMLS* 72.7, pp. 1249–1260. ISSN: 1420-9071. DOI: 10.1007/s00018-014-1791-0.

Hawkins, R D, G C Hon, and B Ren (2010). "Next-generation genomics: an integrative approach". In: *Nat Rev Genet* 11.

Hecht, Andreas et al. (Feb. 1995). "Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: A molecular model for the formation of heterochromatin in yeast". In: *Cell* 80.4, pp. 583–592. ISSN: 0092-8674. DOI: 10.1016/0092-8674(95)90512-X.

Helin, Kristian and Dashyant Dhanak (Oct. 2013). "Chromatin proteins and modifications as drug targets". In: *Nature* 502.7472, pp. 480–488. ISSN: 0028-0836.

Hesselberth, Jay R et al. (Apr. 2009). "Global mapping of protein-DNA interactions in vivo by digital genomic footprinting". In: *Nat Meth* 6.4, pp. 283–289. ISSN: 1548-7091.

Ho, Lena and Gerald R Crabtree (Jan. 2010). "Chromatin remodelling during development". In: *Nature* 463, p. 474.

Hochedlinger, Konrad et al. (Aug. 2004). "Reprogramming of a melanoma genome by nuclear transplantation". In: *Genes & Development* 18.15, pp. 1875–1885. DOI: 10.1101/gad.1213504.

Hodges, Courtney, Jacob G Kirkland, and Gerald R Crabtree (Aug. 2016). "The Many Roles of BAF (mSWI/SNF) and PBAF Complexes in Cancer". In: *Cold Spring Harbor perspectives in medicine* 6.8, a026930. ISSN: 2157-1422. DOI: 10.1101/cshperspect.a026930.

Hodges, H Courtney et al. (2018). "Dominant-negative SMARCA4 mutants alter the accessibility landscape of tissue-unrestricted enhancers". In: *Nature Structural & Molecular Biology* 25.1, pp. 61–72. ISSN: 1545-9985. DOI: 10.1038/s41594-017-0007-3.

Hota, Swetansu K et al. (Jan. 2019). "Dynamic BAF chromatin remodeling complex subunit inclusion promotes temporally distinct gene expression programs in cardiogenesis". In: *Development*, dev.174086. DOI: 10.1242/dev.174086.

Hu, Gangqing et al. (Oct. 2011). "Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1". In: *Genome Research* 21.10, pp. 1650–1658. DOI: 10.1101/gr.121145.111.

Irizarry, Rafael A et al. (Feb. 2009). "The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores". In: *Nat Genet* 41.2, pp. 178–186. ISSN: 1061-4036.

Ishiguro, Kazuya et al. (Jan. 2019). "DOT1L inhibition blocks multiple myeloma cell proliferation by suppressing IRF4-MYC signaling". eng. In: *Haematologica* 104.1, pp. 155–165. ISSN: 1592-8721. DOI: 10.3324/haematol.2018.191262.

Ishikawa, Yoshinori et al. (Feb. 2017). "A Novel LSD1 Inhibitor T-3775440 Disrupts GFI1B-Containing Complex Leading to Transdifferentiation and Impaired Growth of AML Cells". In: *Molecular Cancer Therapeutics* 16.2, 273 LP –284.

Ito, Taiji et al. (Jan. 2001). "Identification of SWI·SNF Complex Subunit BAF60a as a Determinant of the Transactivation Potential of Fos/Jun Dimers". In: *Journal of Biological Chemistry* 276.4, pp. 2852–2857. DOI: 10.1074/jbc.M009633200.

Iwasaki, Hiromi et al. (Nov. 2006). "The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages". eng. In: *Genes & development* 20.21, pp. 3010–3021. ISSN: 0890-9369. DOI: 10.1101/gad.1493506.

Jabbour, Elias et al. (June 2008). "Evolution of decitabine development: accomplishments, ongoing investigations, and future strategies". eng. In: *Cancer* 112.11, pp. 2341–2351. ISSN: 1097-0142. DOI: 10.1002/cncr.23463.

Jancalek, Radim (Oct. 2014). "The role of the TP73 gene and its transcripts in neuro-oncology". In: *British Journal of Neurosurgery* 28.5, pp. 598–605. ISSN: 0268-8697. DOI: 10.3109/02688697.2014.908162.

Jiang, Jin and Michael Levine (Mar. 1993). "Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen". In: *Cell* 72.5, pp. 741–752. ISSN: 0092-8674. DOI: 10.1016/0092-8674(93)90402-C.

Jiang, Tianyun et al. (Jan. 2009). "Achaete-Scute Complex Homologue 1 Regulates Tumor-Initiating Capacity in Human Small Cell Lung Cancer". In: *Cancer Research* 69.3, 845 LP –854.

Johann, Pascal D et al. (2016). "Atypical Teratoid/Rhabdoid Tumors Are Comprised of Three Epigenetic Subgroups with Distinct Enhancer Landscapes". In: *Cancer Cell* 29.3, pp. 379–393. ISSN: 1535-6108. DOI: https://doi.org/10.1016/j.ccell.2016.02.001.

John, Sam, Peter J Sabo, Thomas A Johnson, et al. (Mar. 2008). "Interaction of the Glucocorticoid Receptor with the Chromatin Landscape". In: *Molecular Cell* 29.5, pp. 611–624. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2008.02.010.

John, Sam, Peter J Sabo, Robert E Thurman, et al. (Mar. 2011). "Chromatin accessibility pre-determines glucocorticoid receptor binding patterns". In: *Nat Genet* 43.3, pp. 264–268. ISSN: 1061-4036.

Johnston, R N, S B Pai, and R B Pai (1992). "The origin of the cancer cell: oncogeny reverses phylogeny." eng. In: *Biochemistry and cell biology = Biochimie et biologie cellulaire* 70.10-11, pp. 831–834. ISSN: 0829-8211 (Print).

Kadam, Shilpa et al. (Oct. 2000). "Functional selectivity of recombinant mammalian SWI/SNF subunits". In: *Genes & Development* 14.19, pp. 2441–2451. DOI: 10.1101/gad.828000.

Kadoch, Cigall and Gerald R. Crabtree (Mar. 2013). "Reversible Disruption of mSWI/SNF (BAF) Complexes by the SS18-SSX Oncogenic Fusion in Synovial Sarcoma". In: *Cell* 153.1, pp. 71–85. ISSN: 0092-8674. DOI: 10.1016/j.cell.2013.02.036.

Kadonaga, James T (1998). "Eukaryotic Transcription: An Interlaced Network of Transcription Factors and Chromatin-Modifying Machines". In: *Cell* 92.3, pp. 307–313. ISSN: 00928674. DOI: 10.1016/S0092-8674(00)80924-1.

Kandoth, Cyriac et al. (Oct. 2013). "Mutational landscape and significance across 12 major cancer types". In: *Nature* 502, p. 333.

Karsunky, Holger et al. (Jan. 2002). "Inflammatory reactions and severe neutropenia in mice lacking the transcriptional repressor Gfi1". In: *Nature Genetics* 30, p. 295.

Kazanjian, Avedis et al. (Oct. 2004). "Growth Factor Independence-1 Is Expressed in Primary Human Neuroendocrine Lung Carcinomas and Mediates the Differentiation of Murine Pulmonary Neuroendocrine Cells". In: *Cancer Research* 64.19, 6874 LP –6882.

Kelso, Timothy W R et al. (Oct. 2017). "Chromatin accessibility underlies synthetic lethality of SWI/SNF subunits in ARID1A-mutant cancers". In: *eLife* 6. ISSN: 2050-084X. DOI: 10.7554/eLife.30506.

Kent, W James et al. (June 2002). "The human genome browser at UCSC". In: *Genome research* 12.6, pp. 996–1006. ISSN: 1088-9051. DOI: 10.1101/gr.229102.

Kerenyi, Marc A et al. (June 2013). "Histone demethylase Lsd1 represses hematopoietic stem and progenitor cell signatures during blood cell maturation". In: *eLife* 2. Ed. by Sean J Morrison, e00633. ISSN: 2050-084X. DOI: 10.7554/eLife.00633.

Kho, Alvin T et al. (Mar. 2004). "Conserved mechanisms across development and tumorigenesis revealed by a mouse development perspective of human cancers". In: *Genes & Development* 18.6, pp. 629–640. DOI: 10.1101/gad.1182504.

Kim, Jungsun et al. (2013). "An iPSC Line from Human Pancreatic Ductal Adenocarcinoma Undergoes Early to Invasive Stages of Pancreatic Cancer Progression". In: *Cell Reports* 3.6, pp. 2088–2099. ISSN: 2211-1247. DOI: https://doi.org/10.1016/j.celrep.2013.05.036.

Krom, Mariken de et al. (2002). "Stochastic Patterns in Globin Gene Expression Are Established prior to Transcriptional Activation and Are Clonally Inherited". In: *Molecular Cell* 9.6, pp. 1319–1326. ISSN: 1097-2765. DOI: https://doi.org/10.1016/S1097-2765(02)00558-0.

Kulesa, Paul M et al. (Mar. 2006). "Reprogramming metastatic melanoma cells to assume a neural crest cell-like phenotype in an embryonic microenvironment". In: *Proceedings of the National Academy of Sciences of the United States of America* 103.10, pp. 3752–3757. DOI: 10.1073/pnas.0506977103.

Kwei, K A et al. (June 2008). "Genomic profiling identifies TITF1 as a lineage-specific oncogene amplified in lung cancer". eng. In: *Oncogene* 27.25, pp. 3635–3640. ISSN: 1476-5594. DOI: 10.1038/sj.onc.1211012.

Ladewig, Julia, Philipp Koch, and Oliver Brüstle (Mar. 2013). "Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies". In: *Nature Reviews Molecular Cell Biology* 14, p. 225.

Laurent, Benoit et al. (Mar. 2015). "A Specific LSD1/KDM1A Isoform Regulates Neuronal Differentiation through H3K9 Demethylation". In: *Molecular Cell* 57.6, pp. 957–970. ISSN: 10972765. DOI: 10.1016/j.molcel.2015.01.010.

Le Dily, François, Davide Baù, et al. (Oct. 2014). "Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation". In: *Genes & Development* 28.19, pp. 2151–2162. DOI: 10.1101/gad.241422.114.

Le Dily, François, Enrique Vidal, et al. (Jan. 2019). "Hormone-control regions mediate steroid receptor–dependent genome organization". In: *Genome Research* 29.1, pp. 29–39. DOI: 10.1101/gr.243824.118.

Lee, Ryan S et al. (Aug. 2012). "A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers". In: *The Journal of Clinical Investigation* 122.8, pp. 2983–2988. ISSN: 0021-9738. DOI: 10.1172/JCI64400.

Lemercier, Claudie et al. (Mar. 1998). "The basic helix–loop–helix transcription factor Mist1 functions as a transcriptional repressor of MyoD". In: *The EMBO Journal* 17.5, 1412 LP –1422. DOI: 10.1093/emboj/17.5.1412.

Lenhart, Ryan et al. (Oct. 2015). "Sensitivity of Small Cell Lung Cancer to BET Inhibition Is Mediated by Regulation of ASCL1 Gene Expression". In: *Molecular Cancer Therapeutics* 14.10, 2167 LP –2174.

Ley, Timothy J et al. (Nov. 2010). "DNMT3A Mutations in Acute Myeloid Leukemia". In: *New England Journal of Medicine* 363.25, pp. 2424–2433. ISSN: 0028-4793. DOI: 10.1056/NEJMoa1005143.

Li, H et al. (2009). "The Ink4/Arf locus is a barrier for iPS cell reprogramming". In: *Nature* 460. DOI: 10.1038/nature08290.

Li, Qiliang et al. (Nov. 2002). "Locus control regions". In: *Blood* 100.9, 3077 LP –3086. DOI: 10.1182/blood-2002-04-1104.

Li, X and M Noll (Jan. 1994). "Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the Drosophila embryo". eng. In: *The EMBO journal* 13.2, pp. 400–406. ISSN: 0261-4189.

Lienert, Florian et al. (Oct. 2011). "Identification of genetic elements that autonomously determine DNA methylation states". In: *Nature Genetics* 43, p. 1091.

Lim, Soyoung et al. (Mar. 2010). "Lysine-specific demethylase 1 (LSD1) is highly expressed in ER-negative breast cancers and a biomarker predicting aggressive biology." In: *Carcinogenesis* 31.3, pp. 512–20. ISSN: 1460-2180. DOI: 10.1093/carcin/bgp324.

Lin, Biaoyang et al. (Aug. 2009). "Integrated Expression Profiling and ChIP-seq Analyses of the Growth Inhibition Response Program of the Androgen Receptor". In: *PLOS ONE* 4.8, e6589.

Linnoila, R Ilona et al. (Aug. 2000). "Constitutive achaete-scute homologue-1 promotes airway dysplasia and lung neuroendocrine tumors in transgenic mice". In: *Cancer Research* 60.15, pp. 4005–4009. ISSN: 00085472.

Local, Andrea et al. (2018). "Identification of H3K4me1-associated proteins at mammalian enhancers". In: *Nature Genetics* 50.1, pp. 73–82. ISSN: 1546-1718. DOI: 10.1038/s41588-017-0015-6.

Lohuizen, M van et al. (1991). "Sequence similarity between the mammalian bmi-1 proto-oncogene and the Drosophila regulatory genes Psc and Su(z)2". eng. In: *Nature* 353.6342, pp. 353–355. ISSN: 0028-0836. DOI: 10.1038/353353a0.

Love, Michael I, Wolfgang Huber, and Simon Anders (2014). "Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2". In:

Lovén, Jakob et al. (Apr. 2013). "Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers". In: *Cell* 153.2, pp. 320–334. ISSN: 0092-8674. DOI: 10.1016/j.cell.2013.03.036.

Lu, Chao et al. (May 2016). "Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape". In: *Science* 352.6287, 844 LP –849.

Lv, Tangfeng et al. (Apr. 2012). "Over-Expression of LSD1 Promotes Proliferation, Migration and Invasion in Non-Small Cell Lung Cancer". In: *PLoS ONE* 7.4. Ed. by Pan-Chyr Yang, e35065. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0035065.

Macleod, D et al. (Oct. 1994). "Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island." In: *Genes & Development* 8.19, pp. 2282–2292. DOI: 10.1101/gad.8.19.2282.

Maiques-Diaz, Alba et al. (Mar. 2018). "Enhancer Activation by Pharmacologic Displacement of LSD1 from GFI1 Induces Differentiation in Acute Myeloid Leukemia". In: *Cell Reports* 22.13, pp. 3641–3659. ISSN: 2211-1247. DOI: 10.1016/J.CELREP.2018.03.012.

Mallo, Moisés and Claudio R Alonso (Oct. 2013). "The regulation of Hox gene expression during animal development". In: *Development* 140.19, 3951 LP –3963. DOI: 10.1242/dev.068346.

Mansour, Marc R et al. (Dec. 2014). "An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element." eng. In: *Science (New York, N.Y.)* 346.6215, pp. 1373–1377. ISSN: 1095-9203. DOI: 10.1126/science.1259037.

Matsumoto, Steven et al. (2006). "Brg1 is required for murine neural stem cell maintenance and gliogenesis". In: *Developmental Biology* 289.2, pp. 372–383. ISSN: 0012-1606. DOI: https://doi.org/10.1016/j.ydbio.2005.10.044.

Matys, V et al. (Jan. 2006). "TRANSFAC(®) and its module TRANSCompel(®): transcriptional gene regulation in eukaryotes". In: *Nucleic Acids Research* 34.Database issue, pp. D108–D110. ISSN: 0305-1048. DOI: 10.1093/nar/gkj143.

Maurano, Matthew T. and John A. Stamatoyannopoulos (2015). "Taking Stock of Regulatory Variation". In: *Cell Systems* 1.1, pp. 18–21. ISSN: 2405-4712. DOI: https://doi.org/10.1016/j.cels.2015.07.011.

Maurano, Matthew T et al. (2015). "Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo." In: *Nature genetics* 47.12, pp. 1393–401. ISSN: 1546-1718. DOI: 10.1038/ng.3432.

McArthur, M, S Gerum, and G Stamatoyannopoulos (Oct. 2001). "Quantification of DNaseI-sensitivity by real-time PCR: quantitative analysis of DNaseI-hypersensitivity of the mouse beta-globin LCR". eng. In: *Journal of molecular biology* 313.1, pp. 27–34. ISSN: 0022-2836. DOI: 10.1006/jmbi.2001.4969.

McBride, Matthew J et al. (June 2018). "The SS18-SSX Fusion Oncoprotein Hijacks BAF Complex Targeting and Function to Drive Synovial Sarcoma". In: *Cancer Cell* 33.6, 1128–1141.e7. ISSN: 1535-6108. DOI: 10.1016/j.ccell.2018.05.002.

McBryant, Steven J, Valerie H Adams, and Jeffrey C Hansen (2006). "Chromatin architectural proteins". In: *Chromosome Research* 14.1, pp. 39–51. ISSN: 1573-6849. DOI: 10.1007/s10577-006-1025-x.

Medina, Pedro P, Julian Carretero, et al. (Apr. 2005). "Transcriptional targets of the chromatin-remodelling factor SMARCA4/BRG1 in lung cancer cells". In: *Human Molecular Genetics* 14.7, pp. 973–982. ISSN: 0964-6906.

Medina, Pedro P, Octavio A Romero, et al. (May 2008). "Frequent BRG1/SMARCA4–inactivating mutations in human lung cancer cell lines". In: *Human Mutation* 29.5, pp. 617–622. ISSN: 1059-7794. DOI: 10.1002/humu.20730.

Metzger, Eric et al. (Sept. 2005). "LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription". In: *Nature* 437.7057, pp. 436–439. ISSN: 0028-0836.

Mirny, Leonid A (Dec. 2010). "Nucleosome-mediated cooperativity between transcription factors". In: *Proceedings of the National Academy of Sciences* 107.52, 22534 LP –22539. DOI: 10.1073/pnas.0913805107.

Mohammad, Helai P. et al. (2015). "A DNA Hypomethylation Signature Predicts Antitumor Activity of LSD1 Inhibitors in SCLC". In: *Cancer Cell* 28.1, pp. 57–69. ISSN: 18783686. DOI: 10.1016/j.ccell.2015.06.002.

Mootha, Vamsi K et al. (June 2003). "PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes". In: *Nature Genetics* 34, p. 267.

Morrow, James J et al. (Feb. 2018). "Positively selected enhancer elements endow osteosarcoma cells with metastatic competence". eng. In: *Nature medicine* 24.2, pp. 176–185. ISSN: 1546-170X. DOI: 10.1038/nm.4475.

Murry, Charles E and Gordon Keller (Feb. 2008). "Differentiation of Embryonic Stem Cells to Clinically Relevant Populations: Lessons from Embryonic Development". In: *Cell* 132.4, pp. 661–680. ISSN: 0092-8674. DOI: 10.1016/j.cell.2008.02.008.

Muruganujan, Anushya et al. (Nov. 2016). "PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements". In: *Nucleic Acids Research* 45.D1, pp. D183–D189. ISSN: 0305-1048. DOI: 10.1093/nar/gkw1138.

Nagl, Norman G et al. (Oct. 2005). "The p270 (ARID1A/SMARCF1) Subunit of Mammalian SWI/SNF-Related Complexes Is Essential for Normal Cell Cycle Arrest". In: *Cancer Research* 65.20, 9236 LP –9244. DOI: 10.1158/0008-5472.CAN-05-1225.

Nakayama, Robert T et al. (Sept. 2017). "SMARCB1 is required for widespread BAF complex–mediated activation of enhancers and bivalent promoters". In: *Nature Genetics* 49, p. 1613.

Narlikar, Geeta J., Ramasubramanian Sundaramoorthy, and Tom Owen-Hughes (Aug. 2013). "Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes". In: *Cell* 154.3, pp. 490–503. ISSN: 0092-8674. DOI: 10.1016/j.cell.2013.07.011.

Neph, Shane et al. (July 2012). "BEDOPS: high-performance genomic feature operations". In: *Bioinformatics* 28.14, pp. 1919–1920. ISSN: 1367-4803.

Nerlov, C and T Graf (Aug. 1998). "PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors". eng. In: *Genes & development* 12.15, pp. 2403–2412. ISSN: 0890-9369.

Network, The Cancer Genome Atlas Research, Eric A Collisson, et al. (July 2014). "Comprehensive molecular profiling of lung adenocarcinoma". In: *Nature* 511, p. 543.

Network, The Cancer Genome Atlas Research, John N Weinstein, et al. (Oct. 2013). "The Cancer Genome Atlas Pan-Cancer analysis project". In: *Nat Genet* 45.10, pp. 1113–1120. ISSN: 1061-4036.

Newburger, Daniel E and Martha L Bulyk (Jan. 2009). "UniPROBE: an online database of protein binding microarray data on protein–DNA interactions". In: *Nucleic Acids Research* 37.Database issue, pp. D77–D82. ISSN: 0305-1048. DOI: 10.1093/nar/gkn660.

Onder, Tamer T et al. (Mar. 2012). "Chromatin-modifying enzymes as modulators of reprogramming". In: *Nature* 483, p. 598.

Ong, Chin-Tong and Victor G Corces (Mar. 2014). "CTCF: an architectural protein bridging genome topology and function". In: *Nature Reviews Genetics* 15, p. 234.

Orkin, Stuart H (Mar. 1995). "Transcription Factors and Hematopoietic Development". In: *Journal of Biological Chemistry* 270.10, pp. 4955–4958. DOI: 10.1074/jbc.270.10.4955.

Orvis, Tess et al. (Aug. 2014). "BRG1/SMARCA4 Inactivation Promotes Non–Small Cell Lung Cancer Aggressiveness by Altering Chromatin Organization". In: *Cancer Research*. DOI: 10.1158/0008-5472.CAN-14-0061.

Osada, Hirotaka et al. (Dec. 2005). "ASH1 Gene Is a Specific Therapeutic Target for Lung Cancers with Neuroendocrine Features". In: *Cancer Research* 65.23, 10680 LP –10685.

Pabst, Thomas et al. (Mar. 2001). "Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-α (C/EBPα), in acute myeloid leukemia". In: *Nature Genetics* 27, p. 263.

Park, Kwon-Sik et al. (Aug. 2011). "Characterization of the cell of origin for small cell lung cancer". In: *Cell Cycle* 10.16, pp. 2806–2815. ISSN: 1538-4101. DOI: 10.4161/cc.10.16.17012.

Parker, Stephen C J et al. (Oct. 2013). "Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants". In: *Proceedings of the National Academy of Sciences* 110.44, 17921 LP –17926. DOI: 10.1073/pnas.1317023110.

Polak, Paz et al. (Feb. 2015). "Cell-of-origin chromatin organization shapes the mutational landscape of cancer". In: *Nature* 518, p. 360.

Pomerantz, Mark M et al. (Nov. 2015). "The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis". In: *Nat Genet* 47.11, pp. 1346–1351. ISSN: 1061-4036.

Portales-Casamar, Elodie et al. (Jan. 2010). "JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles". In: *Nucleic Acids Research* 38.Database issue, pp. D105–D110. ISSN: 0305-1048. DOI: 10.1093/nar/gkp950.

Qian, Xiaolan et al. (May 2007). "Oncogenic inhibition by a deleted in liver cancer gene requires cooperation between tensin binding and Rho-specific GTPase-activating protein activities". In: *Proceedings of the National Academy of Sciences* 104.21, 9012 LP –9017. DOI: 10.1073/pnas.0703033104.

Raab, Jesse R, Samuel Resnick, and Terry Magnuson (Dec. 2016). "Genome-Wide Transcriptional Regulation Mediated by Biochemically Distinct SWI/SNF Complexes". In: *PLOS Genetics* 11.12, e1005748.

Ragunathan, Kaushik, Gloria Jih, and Danesh Moazed (Apr. 2015). "Epigenetics. Epigenetic inheritance uncoupled from sequence-specific recruitment". eng. In: *Science (New York, N.Y.)* 348.6230, p. 1258699. ISSN: 1095-9203. DOI: 10.1126/science.1258699.

Raj, Arjun et al. (Oct. 2006). "Stochastic mRNA synthesis in mammalian cells". eng. In: *PLoS biology* 4.10, e309–e309. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.0040309.

Ramagopalan, Sreeram V et al. (Oct. 2010). "A ChIP-seq defined genome-wide map of vitamin D receptor binding: Associations with disease and evolution". In: *Genome Research* 20.10, pp. 1352–1360. DOI: 10.1101/gr.107920.110.

Rao, Suhas S P et al. (2017). "Cohesin Loss Eliminates All Loop Domains". In: *Cell* 171.2, 305–320.e24. ISSN: 0092-8674. DOI: https://doi.org/10.1016/j.cell.2017.09.026.

Rao, Suhas S.P. et al. (Dec. 2014). "A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping". In: *Cell* 159.7, pp. 1665–1680. ISSN: 0092-8674. DOI: 10.1016/j.cell.2014.11.021.

Raser, Jonathan M and Erin K O'Shea (June 2004). "Control of Stochasticity in Eukaryotic Gene Expression". In: *Science* 304.5678, 1811 LP –1814. DOI: 10.1126/science.1098641.

— (Sept. 2005). "Noise in Gene Expression: Origins, Consequences, and Control". In: *Science* 309.5743, 2010 LP –2013. DOI: 10.1126/science.1105891.

Reddy, Timothy E et al. (Dec. 2009). "Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation". In: *Genome Research* 19.12, pp. 2163–2171. DOI: 10.1101/gr.097022.109.

Ren, Yixuan et al. (Dec. 2015). "Targeting Transcriptional Dependency in Acute Myeloid Leukemia (AML) with a Covalent Inhibitor of Transcriptional Kinase CDK7". In: *Blood* 126.23, 1354 LP –1354.

Rhodes, Jennifer et al. (Jan. 2005). "Interplay of Pu.1 and Gata1 Determines Myelo-Erythroid Progenitor Cell Fate in Zebrafish". In: *Developmental Cell* 8.1, pp. 97–108. ISSN: 1534-5807. DOI: 10.1016/j.devcel.2004.11.014.

Rickels, Ryan et al. (Nov. 2017). "Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability". eng. In: *Nature genetics* 49.11, pp. 1647–1653. ISSN: 1546-1718. DOI: 10.1038/ng.3965.

Ringrose, Leonie and Renato Paro (Nov. 2004). "Epigenetic Regulation of Cellular Memory by the Polycomb and Trithorax Group Proteins". In: *Annual Review of Genetics* 38.1, pp. 413–443. ISSN: 0066-4197. DOI: 10.1146/annurev.genet.38.072902.091907.

Ripley, B D (1976). "The Second-Order Analysis of Stationary Point Processes". In: *Journal of Applied Probability* 13.2, pp. 255–266. ISSN: 00219002. DOI: 10.2307/3212829.

Roadmap Epigenomics Consortium et al. (Feb. 2015). "Integrative analysis of 111 reference human epigenomes". In: *Nature* 518.7539, pp. 317–330. ISSN: 0028-0836.

Rodriguez-Nieto, Salvador et al. (Feb. 2011). "Massive parallel DNA pyrosequencing analysis of the tumor suppressor BRG1/SMARCA4 in lung primary tumors". In: *Human Mutation* 32.2, E1999–E2017. ISSN: 1059-7794. DOI: 10.1002/humu.21415.

Roe, Jae-Seok et al. (Aug. 2017). "Enhancer Reprogramming Promotes Pancreatic Cancer Metastasis". eng. In: *Cell* 170.5, 875–888.e20. ISSN: 1097-4172. DOI: 10.1016/j.cell.2017.07.007.

Rosenbauer, Frank et al. (May 2004). "Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1". In: *Nature Genetics* 36, p. 624.

Roy, Nilotpal et al. (Mar. 2015). "Brg1 promotes both tumor-suppressive and oncogenic activities at distinct stages of pancreatic cancer formation". In: *Genes & Development* 29.6, pp. 658–671. DOI: 10.1101/gad.256628.114.

Rudin, Charles M et al. (Oct. 2012). "Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer". In: *Nat Genet* 44.10, pp. 1111–1116. ISSN: 1061-4036.

Sabo, Peter J et al. (2006). "Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays". In: *Nature Methods* 3.7, pp. 511–518. ISSN: 1548-7105. DOI: 10.1038/nmeth890.

Saintigny, Pierre et al. (Sept. 2012). "Global evaluation of Eph receptors and ephrins in lung adenocarcinomas identifies EphA4 as an inhibitor of cell migration and invasion". eng. In: *Molecular cancer therapeutics* 11.9, pp. 2021–2032. ISSN: 1538-8514. DOI: 10.1158/1535-7163.MCT-12-0030.

Sakuma, Tetsushi et al. (Apr. 2013). "Efficient TALEN construction and evaluation methods for human cell and animal applications". In: *Genes to cells : devoted to molecular & cellular mechanisms* 18.4, pp. 315–326. ISSN: 1356-9597. DOI: 10.1111/gtc.12037.

Saleque, Shireen et al. (Aug. 2007). "Epigenetic Regulation of Hematopoietic Differentiation by Gfi-1 and Gfi-1b Is Mediated by the Cofactors CoREST and LSD1". In: *Molecular Cell* 27.4, pp. 562–572. ISSN: 1097-2765. DOI: 10.1016/J.MOLCEL.2007.06.039.

Sankar, Savita et al. (Sept. 2014). "Reversible LSD1 Inhibition Interferes with Global EWS/ETS Transcriptional Activity and Impedes Ewing Sarcoma Tumor Growth". In: *Clinical Cancer Research* 20.17, 4584 LP –4597.

Sashida, Goro and Atsushi Iwama (Jan. 2017). "Multifaceted role of the polycomb-group gene EZH2 in hematological malignancies". In: *International Journal of Hematology* 105.1, pp. 23–30. ISSN: 1865-3774. DOI: 10.1007/s12185-016-2124-x.

Schenk, Tino et al. (Apr. 2012). "Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia". In: *Nat Med* 18.4, pp. 605–611. ISSN: 1078-8956.

Schildhaus, Hans-Ulrich et al. (2011). "Lysine-specific demethylase 1 is highly expressed in solitary fibrous tumors, synovial sarcomas, rhabdomyosarcomas, desmoplastic small

round cell tumors, and malignant peripheral nerve sheath tumors". In: *Human Pathology* 42.11, pp. 1667–1675. ISSN: 00468177. DOI: 10.1016/j.humpath.2010.12.025.

Schmidt, Dominic et al. (July 2009). "ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions". In: *Methods (San Diego, Calif.)* 48.3, pp. 240–248. ISSN: 1046-2023. DOI: 10.1016/j.ymeth.2009.03.001.

Schübeler, Dirk (Jan. 2015). "Function and information content of DNA methylation". In: *Nature* 517, p. 321.

Schwartzentruber, Jeremy et al. (Feb. 2012). "Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma". In: *Nature* 482.7384, pp. 226–231. ISSN: 0028-0836.

Scully, Steve et al. (Sept. 2012). "Transdifferentiation of Glioblastoma Stem-Like Cells into Mural Cells Drives Vasculogenic Mimicry in Glioblastomas". In: *The Journal of Neuroscience* 32.37, 12950 LP –12960.

Sekimata, Masayuki et al. (June 1999). "Morphological Changes and Detachment of Adherent Cells Induced by p122, a GTPase-activating Protein for Rho". In: *Journal of Biological Chemistry* 274.25, pp. 17757–17762. DOI: 10.1074/jbc.274.25.17757.

Serna, Ivana L de la et al. (July 2006). "The Microphthalmia-associated Transcription Factor Requires SWI/SNF Enzymes to Activate Melanocyte-specific Genes". In: *Journal of Biological Chemistry* 281.29, pp. 20233–20241. DOI: 10.1074/jbc.M512052200.

Shaffer, Arthur L et al. (July 2008). "IRF4 addiction in multiple myeloma". In: *Nature* 454.7201, pp. 226–231. ISSN: 0028-0836.

Shaffer, Sydney M et al. (June 2017). "Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance". In: *Nature* 546.7658, pp. 431–435. ISSN: 1476-4687. DOI: 10.1038/nature22794.

Sharma, Sreenath V et al. (Apr. 2010). "A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations". In: *Cell* 141.1, pp. 69–80. ISSN: 1097-4172. DOI: 10.1016/j.cell.2010.02.027.

Shen, Lanlan et al. (Feb. 2010). "DNA Methylation Predicts Survival and Response to Therapy in Patients With Myelodysplastic Syndromes". In: *Journal of Clinical Oncology* 28.4, pp. 605–613. ISSN: 0732-183X. DOI: 10.1200/JCO.2009.23.4781.

Shi, Yu-Jiang et al. (2005). "Regulation of LSD1 Histone Demethylase Activity by Its Associated Factors".

Shi, Junwei and Christopher R. Vakoc (2014). "The Mechanisms behind the Therapeutic Activity of BET Bromodomain Inhibition". In: *Molecular Cell* 54.5, pp. 728–736. ISSN: 1097-2765. DOI: https://doi.org/10.1016/j.molcel.2014.05.016.

Shogren-Knaak, Michael et al. (Feb. 2006). "Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions". In: *Science* 311.5762, 844 LP –847. DOI: 10.1126/science.1124000.

Simcox, Amanda A and James H Sang (1983). "When does determination occur in Drosophila embryos?" In: *Developmental Biology* 97.1, pp. 212–221. ISSN: 0012-1606. DOI: https://doi.org/10.1016/0012-1606(83)90078-7.

Skene, Peter J, Jorja G Henikoff, and Steven Henikoff (Apr. 2018). "Targeted in situ genome-wide profiling with high efficiency for low cell numbers". In: *Nature Protocols* 13, p. 1006.

Skene, Peter J and Steven Henikoff (2017). "An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites". In: *eLife* 6. Ed. by Danny Reinberg, e21856. ISSN: 2050-084X. DOI: 10.7554/eLife.21856.

Skibinski, Adam et al. (2014). "The Hippo Transducer TAZ Interacts with the SWI/SNF Complex to Regulate Breast Epithelial Lineage Commitment". In: *Cell Reports* 6.6, pp. 1059–1072. ISSN: 2211-1247. DOI: https://doi.org/10.1016/j.celrep.2014.02.038.

Smith, Zachary D and Alexander Meissner (Feb. 2013). "DNA methylation: roles in mammalian development". In: *Nature Reviews Genetics* 14, p. 204.

Smyth, Gordon K, Wei Shi, and Yang Liao (Nov. 2013). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features". In: *Bioinformatics* 30.7, pp. 923–930. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btt656.

Soda, Yasushi et al. (Mar. 2011). "Transdifferentiation of glioblastoma cells into vascular endothelial cells". In: *Proceedings of the National Academy of Sciences* 108.11, pp. 4274–4280. DOI: 10.1073/pnas.1016030108.

Song, Jikui et al. (Feb. 2011). "Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation". eng. In: *Science (New York, N.Y.)* 331.6020, pp. 1036–1040. ISSN: 1095-9203. DOI: 10.1126/science.1195380.

Sprussel, A et al. (Sept. 2012). "Lysine-specific demethylase 1 restricts hematopoietic progenitor proliferation and is essential for terminal differentiation". In: *Leukemia* 26.9, pp. 2039–2051. ISSN: 0887-6924.

Stankunas, Kryn et al. (Feb. 2008). "Endocardial Brg1 Represses ¡em¿ADAMTS1¡/em¿ to Maintain the Microenvironment for Myocardial Morphogenesis". In: *Developmental Cell* 14.2, pp. 298–311. ISSN: 1534-5807. DOI: 10.1016/j.devcel.2007.11.018.

Stergachis, Andrew B et al. (Aug. 2013). "Developmental fate and cellular maturity encoded in human regulatory DNA landscapes." In: *Cell* 154.4, pp. 888–903. ISSN: 1097-4172. DOI: 10.1016/j.cell.2013.07.020.

Stewart, David J (Dec. 2013). "Wnt Signaling Pathway in Non–Small Cell Lung Cancer". In: *JNCI: Journal of the National Cancer Institute* 106.1. ISSN: 0027-8874. DOI: 10.1093/jnci/djt356.

Subramanian, Aravind et al. (Oct. 2005). "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles". In: *Proceedings of the National Academy of Sciences* 102.43, 15545 LP –15550. DOI: 10.1073/pnas.0506580102.

Sun, Xuxu et al. (2017). "Arid1a Has Context-Dependent Oncogenic and Tumor Suppressor Functions in Liver Cancer". In: *Cancer Cell* 32.5, 574–589.e6. ISSN: 1535-6108. DOI: https://doi.org/10.1016/j.ccell.2017.10.007.

Sutherland, Kate D. et al. (June 2011). "Cell of Origin of Small Cell Lung Cancer: Inactivation of ¡em¿Trp53¡/em¿ and ¡em¿Rb1¡/em¿ in Distinct Cell Types of Adult Mouse Lung". In: *Cancer Cell* 19.6, pp. 754–764. ISSN: 1535-6108. DOI: 10.1016/j.ccr.2011.04.019.

Takagi, Shinji et al. (June 2017). "LSD1 Inhibitor T-3775440 Inhibits SCLC Cell Proliferation by Disrupting LSD1 Interactions with SNAG Domain Proteins INSM1 and GFI1B". In: *Cancer Research*.

Takahashi, K and S Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors". In: *Cell* 126. DOI: 10.1016/j.cell.2006.07.024.

Tanay, Amos (Aug. 2006). "Extensive low-affinity transcriptional interactions in the yeast genome". In: *Genome Research* 16.8, pp. 962–972. DOI: 10.1101/gr.5113606.

Thambyrajah, Roshana et al. (Jan. 2016). "GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1". In: *Nat Cell Biol* 18.1, pp. 21–32. ISSN: 1465-7392.

Thiery, Jean Paul (2002). "Epithelial–mesenchymal transitions in tumour progression". In: *Nature Reviews Cancer* 2.6, pp. 442–454. ISSN: 1474-1768. DOI: 10.1038/nrc822.

Thiery, Jean Paul and Jonathan P Sleeman (Feb. 2006). "Complex networks orchestrate epithelial-mesenchymal transitions". In: *Nat Rev Mol Cell Biol* 7.2, pp. 131–142. ISSN: 1471-0072.

Thomas, Jean O (1999). "Histone H1: location and role". In: *Current Opinion in Cell Biology* 11.3, pp. 312–317. ISSN: 0955-0674. DOI: https://doi.org/10.1016/S0955-0674(99)80042-8.

Thurman, Robert E et al. (Sept. 2012). "The accessible chromatin landscape of the human genome". In: *Nature* 489.7414, pp. 75–82. ISSN: 0028-0836.

Ting, Chou-Chik et al. (Nov. 1972). "Expression of fetal antigens in tumor cells". In: *Proceedings of the National Academy of Sciences* 69.11, 3493 LP –3493.

Torrino, Stéphanie et al. (Feb. 2019). "UBTD1 is a mechano-regulator controlling cancer aggressiveness". In: *EMBO reports*, e46570. DOI: 10.15252/embr.201846570.

Vachtenheim, Jiri, Lubica Ondrušová, and Jan Borovanský (2010). "SWI/SNF chromatin remodeling complex is critical for the expression of microphthalmia-associated transcription factor in melanoma cells". In: *Biochemical and Biophysical Research Communications* 392.3, pp. 454–459. ISSN: 0006-291X. DOI: https://doi.org/10.1016/j.bbrc.2010.01.048.

Versteege, Isabella et al. (July 1998). "Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer". In: *Nature* 394.6689, pp. 203–206. ISSN: 0028-0836.

Vierbuchen, Thomas et al. (2017). "AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection". In: *Molecular Cell* 68.6, 1067–1082.e12. ISSN: 1097-2765. DOI: https://doi.org/10.1016/j.molcel.2017.11.026.

Vierstra, Jeff and John A Stamatoyannopoulos (Feb. 2016). "Genomic footprinting". In: *Nature Methods* 13, p. 213.

Visvader, J E et al. (Dec. 1992). "GATA-1 but not SCL induces megakaryocytic differentiation in an early myeloid line". eng. In: *The EMBO journal* 11.12, pp. 4557–4564. ISSN: 0261-4189.

Vockley, Christopher M et al. (Aug. 2016). "Direct GR Binding Sites Potentiate Clusters of TF Binding across the Human Genome". eng. In: *Cell* 166.5, 1269–1281.e19. ISSN: 1097-4172. DOI: 10.1016/j.cell.2016.07.049.

Voncken, Jan Willem et al. (Mar. 2003). "Rnf2(Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition". In: *Proceedings of the National Academy of Sciences* 100.5, 2468 LP –2473. DOI: 10.1073/pnas.0434312100.

Waddington, Conrad H (1935). "Cancer and the theory of organisers". In: *Nature* 135, pp. 606–608. ISSN: 0028-0836. DOI: 10.1038/135606a0.

Walter, M J et al. (Mar. 2011). "Recurrent DNMT3A mutations in patients with myelodysplastic syndromes". In: *Leukemia* 25, p. 1153.

Wang, Jing et al. (Nov. 2011). "Novel Histone Demethylase LSD1 Inhibitors Selectively Target Cancer Cells with Pluripotent Stem Cell Properties". In: *Cancer Research* 71.23, 7238 LP –7249.

Wang, Tim et al. (2017). "Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras". In: *Cell* 168.5, 890–903.e15. ISSN: 00928674. DOI: 10.1016/j.cell.2017.01.013.

Wang, Xi et al. (Oct. 2009). "Oncogenesis Caused by Loss of the SNF5 Tumor Suppressor Is Dependent on Activity of BRG1, the ATPase of the SWI/SNF Chromatin Remodeling Complex". In: *Cancer Research* 69.20, 8094 LP –8101.

Wang, Yanli et al. (Oct. 2018). "The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions". In: *Genome Biology* 19.1, p. 151. ISSN: 1474-760X. DOI: 10.1186/s13059-018-1519-9.

Warburg, Otto (Feb. 1956). "On the Origin of Cancer Cells". In: *Science* 123.3191, 309 LP –314. DOI: 10.1126/science.123.3191.309.

Warren, Luigi et al. (Nov. 2006). "Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 103.47, pp. 17807–17812. ISSN: 0027-8424. DOI: 10.1073/pnas.0608512103.

Wei, Xiaomu et al. (Jan. 2018). "Germline mutations in lysine specific demethylase 1 (LSD1/KDM1A) confer susceptibility to multiple myeloma". In: *Cancer Research*.

Weintraub, H et al. (July 1989). "Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD". In: *Proceedings of the National Academy of Sciences* 86.14, 5434 LP –5438. DOI: 10.1073/pnas.86.14.5434.

Whyte, Warren A. et al. (Apr. 2013). "Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes". In: *Cell* 153.2, pp. 307–319. ISSN: 0092-8674. DOI: 10.1016/j.cell.2013.03.035.

Whyte, Warren A et al. (Feb. 2012). "Enhancer decommissioning by LSD1 during embryonic stem cell differentiation". In: *Nature* 482.7384, pp. 221–225. ISSN: 0028-0836.

Witkowski, Leora et al. (Mar. 2014). "Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type". In: *Nature Genetics* 46, p. 438.

Wolpert, L. (1994). *Positional information and pattern formation in development*. DOI: 10.1002/dvg.1020150607.

Wong, Alexander K C et al. (Nov. 2000). "BRG1, a Component of the SWI-SNF Complex, Is Mutated in Multiple Human Tumor Cell Lines". In: *Cancer Research* 60.21, 6171 LP –6177.

Xu, Eugenia Y, Karl A Zawadzki, and James R Broach (2006). "Single-Cell Observations Reveal Intermediate Transcriptional Silencing States". In: *Molecular Cell* 23.2, pp. 219–229. ISSN: 1097-2765. DOI: https://doi.org/10.1016/j.molcel.2006.05.035.

Yamaguchi, Tomoya et al. (2013). "NKX2-1/TTF-1: An Enigmatic Oncogene that Functions as a Double-Edged Sword for Cancer Cell Survival and Progression". In: *Cancer Cell* 23.6, pp. 718–723. ISSN: 15356108. DOI: 10.1016/j.ccr.2013.04.002.

Yang, Yee Hwa et al. (Feb. 2002). "Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation". In: *Nucleic Acids Research* 30.4, e15–e15. ISSN: 0305-1048.

Yuan, Bao-Zhu, Amy M Jefferson, Kimberly T Baldwin, et al. (Dec. 2003). "DLC-1 operates as a tumor suppressor gene in human non-small cell lung carcinomas". In: *Oncogene* 23, p. 1405.

Yuan, Bao-Zhu, Amy M Jefferson, Lyndell Millecchia, et al. (2007). "Morphological changes and nuclear translocation of DLC1 tumor suppressor protein precede apoptosis in human non-small cell lung carcinoma cells". In: *Experimental Cell Research* 313.18, pp. 3868–3880. ISSN: 0014-4827. DOI: https://doi.org/10.1016/j.yexcr.2007.08.009.

Zabidi, Muhammad A et al. (Dec. 2014). "Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation". In: *Nature* 518, p. 556.

Zagorski, Marcin et al. (June 2017). "Decoding of position in the developing neural tube from antiparallel morphogen gradients". eng. In: *Science (New York, N.Y.)* 356.6345, pp. 1379–1383. ISSN: 1095-9203. DOI: 10.1126/science.aam5887.

Zhang, H. Steven et al. (Mar. 2000). "Exit from G1 and S phase of the cell cycle is regulated by repressor complexes containing HDAC-Rb-hSWI/SNF and RB-hSWI/SNF". In: *Cell* 101.1, pp. 79–89. ISSN: 00928674. DOI: 10.1016/S0092-8674(00)80625-X.

Zhou, Xiaoling, Snorri S Thorgeirsson, and Nicholas C Popescu (Dec. 2003). "Restoration of DLC-1 gene expression induces apoptosis and inhibits both cell growth and tumorigenicity in human hepatocellular carcinoma cells". In: *Oncogene* 23, p. 1308.

Zhou, Xin et al. (Nov. 2011). "The Human Epigenome Browser at Washington University". In: *Nature Methods* 8, p. 989.

Zhu, Jiang et al. (Jan. 2013). "Genome-wide Chromatin State Transitions Associated with Developmental and Environmental Cues". In: *Cell* 152.3, pp. 642–654. ISSN: 0092-8674. DOI: 10.1016/j.cell.2012.12.033.

Zlatanova, Jordanka, Corrine Seebart, and Miroslav Tomschik (2008). "The linker-protein network: control of nucleosomal DNA accessibility". In: *Trends in Biochemical Sciences* 33.6, pp. 247–253. ISSN: 0968-0004. DOI: https://doi.org/10.1016/j.tibs.2008.04.001.

Zuber, Johannes et al. (Aug. 2011). "RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia". In: *Nature* 478, p. 524.