NORTHWESTERN UNIVERSITY

An Activating Mutation of the NSD2 Histone Methyltransferase Drives Oncogenic Reprogramming in Acute Lymphocytic Leukemia

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Driskill Graduate Training Program in Life Sciences

By Alok Swaroop

EVANSTON, ILLINOIS

December 2018

© Copyright by Alok Swaroop 2018

All Rights Reserved

Abstract

An Activating Mutation of the NSD2 Histone Methyltransferase Drives Oncogenic Reprogramming in Acute Lymphocytic Leukemia

NSD2, a histone methyltransferase specific for methylation of histone 3 lysine 36 (H3K36), exhibits a glutamic acid to lysine mutation at residue 1099 (E1099K) in childhood acute lymphocytic leukemia (ALL). Cells harboring this mutation can become the predominant clone in relapsing disease. We studied the effects of this mutant enzyme *in silico*, *in vitro*, and *in vivo* using gene edited cell lines. The E1099K mutation altered enzyme/substrate binding and enhanced the rate of H3K36 methylation. As a result, cell lines harboring E1099K exhibit increased H3K36 dimethylation and reduced H3K27 trimethylation, particularly on nucleosomes containing histone H3.1. Mutant NSD2 cells exhibit reduced apoptosis and enhanced proliferation, clonogenicity, adhesion, and migration. In mouse xenografts, mutant NSD2 cells are more lethal and brain invasive than wildtype cells. Transcriptional profiling demonstrates that mutant NSD2 aberrantly activates factors commonly associated with neural and stromal lineages, in addition to signaling and adhesion genes. AKT signaling is also activated in NSD2 E1099K cells, and inhibiting this pathway mitigates the proliferative advantage provided by the mutation, suggesting a direction for therapeutic intervention.

Acknowledgments

I could not have completed my PhD and my dissertation without the help of my mentors, colleagues, collaborators, friends, and family. Both iterations of the Licht laboratory that I worked in were amazing yet so different. At Northwestern, Jon Oyer started the project that would eventually turn into my dissertation and gave me the crucial guidance I needed early on in my PhD. In addition to helping me with so many technical issues with my mice, Marinka Bulic was and is a wonderful friend who helped me through so much more. Mrinal Shah, Teresa Ezponda, Christine Will, Relja Popovic, Eliza Small, and Behnam Nabet were also excellent resources for me whenever I needed them. At the University of Florida, I found a place full of warm and friendly people. Richard Bennett and Olga Guryanova were fantastic resources for me, whose expertise, experience, and availability helped me push boundaries and grow creatively. The friendship and knowledge of Xiaoxiao Huang, Aditya Bele, Daphné Dupéré-Richer, Catalina Troche, Kartika Venugopal, Marta Kulis, Crissandra Lee Piper, Sayantan Maji, and Sharon Norton was so important to me. I could not have made it through the last year of my PhD without them. They lent me their time, experience, and even their homes. I could not be more grateful.

I have been lucky that I have found such amazing mentors at every stage. In high school I worked with Pamela Raymond and Ted Allison and got a wonderful introduction to the world of basic science research. In college, David Ginsburg and Jordan Shavit introduced me to the concept of the physician-scientist and inspired me to make this my own dream. Their advice and mentorship are still so valuable to me now. In Jonathan Licht, I found a PhD mentor who not

only taught me how to be a scientist, but teaches me how to be a mentor myself. In Florida he gave me more responsibility and reminded me why I first became interested in science. I would also like to acknowledge the guidance I received from John Crispino, Jindan Yu, and Ali Shilatifard on my thesis committee. I am grateful to have been funded through the Northwestern Medical Scientist Training Program (T32 GM008152) and Carcinogenesis Training Program (T32 CA009560) grants, as well as the NCI Ruth L. Kirschstein National Research Service Award (F30 CA203292).

I have to thank my family, as well. My dad was my first mentor, providing the initial spark for my interest in science, and he continues to feed the fire today. His excitement when talking about science is contagious, and is no doubt what drew me to the field when I was little. I will never stop relying on him for advice on being a better scientist and a better person. My mom provided a different spark, in that she has always served to be my primary motivator and advocate. I was always a dreamer, but my mom showed me how to make those dreams into reality. I would not be anywhere without her help. Though she is younger than me my sister was and is an inspiration. Her motivations have always been so pure, and her actions so unencumbered by external forces. She does what she feels is right, and she does it with all her heart. Her example is one I strive for. Finally, thank you to my wife. I don't know if graduate school has been more difficult for me or for her, but we made it through. The trip back after our wedding and honeymoon, when we took different connecting flights from the Atlanta airport was one of hardest things we've been through together. Throughout it all, she has provided me with such amazing love, comfort, and encouragement. Her confidence in me never wavered, and it was everything I needed and more. I can't thank her enough. She will also start a new chapter in her career now, and I am excited to see what comes next for her, me, and us.

Preface

Chapter 1 sections on NSD2 are adapted from a review I wrote with Richard Bennett, Catalina Troche, and Jonathan Licht, entitled "The Role of Nuclear Receptor – Binding SET Domain Family Histone Lysine Methyltransferases in Cancer" for *Cold Spring Harbor Perspectives in Medicine*.

Chapter 2 is adapted from an original research article in press in *Oncogene* entitled "An Activating Mutation of the NSD2 Histone Methyltransferase Drives Oncogenic Reprogramming in Acute Lymphocytic Leukemia," with the addition of ongoing work directly related to this project. Figures have been reformatted and detail added where appropriate. *In silico* modeling of NSD2 was performed by Wenbo Yu and Alex MacKerell (University of Maryland). Mass spectrometry was performed by Xiaoxiao Huang with Neil Kelleher (Northwestern University). CRISPR-edited cell lines were generated by Jon Oyer and Christine Will (Northwestern University). RNA-seq analysis was performed in collaboration with Alberto Riva (University of Florida). Figure captions indicate data and figures generated by my colleagues.

Chapters 3 and 4 contain sections adapted from the above manuscript with considerable modification and expansion. Methods provided and data generate by my colleagues are indicated as such.

Table of Contents

Abstract 3

Acknowledgments 4

Preface 6

List of Figures 9

Chapter 1: Introduction 10
- **1.1. Acute lymphoblastic leukemia** 10
- **1.2. Genetic origins of ALL** 10
- **1.3. E2A-PBX1 in ALL** 11
- **1.4. PI3K/Akt pathway activation in ALL** 12
- **1.5. BCR-ABL in leukemia** 13
- **1.6. TEL-AML1 in ALL** 14
- **1.7. NOTCH1 activation in ALL** 14
- **1.8. Utilizing CRISPR/Cas9 to understand the genetic basis of cancer** 15
- **1.9. Compaction and storage of DNA affects its function** 17
- **1.10. Histone proteins and variants** 18
- **1.11. Epigenetic modification of histone proteins** 20
- **1.12. Structure and function of the histone methyltransferase, NSD2** 22
- **1.13. NSD2 in development** 23
- **1.14. NSD2 in cancer** 25
- **1.15. The NSD2-activating point mutation, E1099K** 29
- **1.16. Study goals** 31

Chapter 2: Results 32
- **2.1. The E1099K mutation enhances NSD2 interaction with chromatin** 32
- **2.2. Gene editing to remove the E1099K mutation restores the balance of H3K36me2 to H3K27me3** 38
- **2.3. E1099K mutation promotes cancer-associated phenotypes** 48
- **2.4. NSD2-mutant cells behave aggressively in vivo** 53
- **2.5. Widespread gene activation in E1099K cells drives oncogenic programs** 57
- **2.6. E1099K cells are susceptible to inhibition of AKT** 68
- **2.7. Mechanism of AKT activation by NSD2^{E1099K}** 71
- **2.8. Identifying downstream effectors of NSD2 E1099K** 74
- **2.9. Developing a mouse model of NSD2 E1099K** 77

Chapter 3: Discussion 79
- **3.1. E1099K hyperactivates NSD2 function by increasing affinity for chromatin** 79
- **3.2. NSD2 E1099K Activates Oncogenic Programs** 82
- **3.3. Future directions** 84
- **3.4. Generating new cell lines** 85
- **3.4. Resistance to therapy** 86
- **3.5. Mechanism for AKT Pathway Activation** 87
- **3.6. Epigenetic Therapy** 88
- **3.7. DNA damage** 88
- **3.8. Mouse models** 89

Chapter 4: Methods........92
4.1. NSD2 Modeling........92
4.2. Cell culture........92
4.3. Plasmids, mutagenesis, and fluorescence recovery after photobleaching........93
4.4. CRISPR Generation of NSD2-Edited Cells........94
4.5. Immunoblot........94
4.6. NSD2-nucleosome association assay........95
4.6. Quantitation of histone modifications by targeted mass spectrometry........96
4.7. MS-based measurement and modeling of histone methylation kinetics........96
4.8. Lentivirus production and infection of leukemia cell lines........97
4.9. In Vivo Tumorigenicity........97
4.10. Proliferation........98
4.11. Colony formation........98
4.12. Annexin V staining........98
4.13. Propidium iodide staining........99
4.14. Adhesion........99
4.15. Migration........99
4.16. RNA extraction, cDNA synthesis, and RNA-sequencing........99
4.17. RNA-sequencing alignment and analysis........100
4.18. Antibody microarray........100
4.19. Conditioned media........101
4.20. Serum starvation and stimulation........101
4.21. Cloning of crRNAs into sgRNA expression vectors........101
4.22. Generation and management of NSD2^{E1099K} knockin and E2A-PBX1 mice........102
References........103
Vita........119

List of Figures

Figure 1.14. NSD2 is highly expressed in cancer 28
Figure 2.1.1. *in silico* modeling indicates E1099K mutation enhances NSD2 interaction with chromatin 34
Figure 2.1.2. E1099K increases NSD2 affinity for chromatin by Fluorescence Recovery After Photobleaching 36
Figure 2.2.1. CRISPR/Cas9 mediated editing of NSD2 E1099K 40
Figure 2.2.2. Gene editing to correct the E1099K mutation restores the balance of H3K36me2 to H3K27me3 42
Figure 2.2.3. M4K in RCH-ACV cells reveals increased rate of monomethylation by mutant NSD2 43
Figure 2.2.4. M4K in SEM cells reveals increased rate of monomethylation by mutant NSD2 45
Figure 2.2.5. Hyperactive NSD2 preferentially increases H3K36 dimethylation of histone H3.3 47
Figure 2.3.1. E1099K mutation promotes proliferation and clonogenicity 49
Figure 2.3.2. E1099K mutation reduces apoptosis and alters cell cycle progression 50
Figure 2.3.3. E1099K mutation increases adhesion and migration of ALL cells 52
Figure 2.4.1. NSD2-mutant cells exhibit increased lethality in mice 54
Figure 2.4.2. NSD2$^{WT/E1099K}$ cells demonstrate altered migration *in vivo* 55
Figure 2.5.1. Widespread gene activation in E1099K cells 59
Figure 2.5.2. Genes upregulated by E1099K show strong overlap between cell lines 60
Figure 2.5.3. E1099K upregulated genes are generally marked for repression 61
Figure 2.5.4. E1099K mutation activates genes commonly expressed in stromal cell lineage 62
Figure 2.5.5. Mutant NSD2 upregulates genes involved in many cancer associated pathways 64
Figure 2.5.6. Pathways upregulated in E1099K cells are common between cell lines 65
Figure 2.5.7. Gene expression data from RCH-ACV, SEM, and RPMI-8402 cell lines resembles patient data 67
Figure 2.6.1. Altered signaling between NSD2 wildtype and mutant cells 69
Figure 2.6.2. E1099K cells are susceptible to inhibition of AKT 70
Figure 2.7. A factor secreted by NSD2$^{WT/E1099K}$ cells activates AKT and increases proliferation 73
Figure 2.8. Design of a targeted CRISPR array screen to identify important downstream effectors of NSD2 E1099K 76
Figure 2.9. Splenic B cells in NSD2 E1099K knock-in mice show appropriate changes in histone methylation 78

Chapter 1: Introduction

1.1. Acute lymphoblastic leukemia

ALL is a genetically heterogeneous disease of immature lymphocytes that overgrow their niche and impede the production of normal cells in the bone marrow, infiltrating other organs and resulting in devastating consequences (Golub, 2007). ALL is the most common pediatric cancer, making up 26% of cases, and while significant strides have been made in its treatment, it remains the second most common cause of cancer-related death in children (Society, 2016). Pediatric ALL used to be nearly universally fatal, with a 5-year survival rate of only 57% (Society, 2014). Now, however, the 5-year survival is 90%. Still, relapse is not uncommon and the mechanisms of relapse are poorly understood (Mullighan, 2013).

1.2. Genetic origins of ALL

ALL is broadly classified by cell of origin, B or T cell (B-ALL or T-ALL), with B-ALL making up 85% of cases (Aster and DeAngelo, 2017). Leukemic B cells are arrested in the pre-B developmental stage and express early B cell markers such as CD19, while lacking mature B cell markers such as surface immunoglobulin (Aster and DeAngelo, 2017; Eswaran et al., 2015). Similarly, leukemic T and B cells are terminal deoxynucleotidyl transferase (TdT) positive, a marker of precursor lymphocytes. T leukemic blasts do not express surface T cell receptor, indicating their immature developmental stage (Juárez-Velázquez et al., 2013). Leukemic immature B cells are frequent carriers of chromosomal translocations that help classify this disease. Translocations result in the formation of fusion proteins that are typically not sufficient to initiate malignancy independently (Greaves and Wiemels, 2003). Still, they characterize the genomic profiles of the malignancies that do arise following the accumulation of new mutations. New advances in next-generation sequencing have allowed for in-depth analysis of the genetics of ALL before and after relapse beyond the major translocations, providing key insights into disease etiology and prognosis. ALL pathogenesis is frequently associated with disruption of transcription factors required to guide normal lymphocyte development such as PAX5, IKZF1, EBF1, and LEF1 (Zuckerman and Rowe, 2014). Aberrant activation of key signaling pathways, and loss of tumor suppressor genes required for cell cycle control such as CDKN2A, CDKN2B, RB1, and TP53 are also hallmarks of B-ALL (Mullighan, 2013). NGS data has revealed that epigenetic regulators are frequently mutated in ALL, as they are in many cancers (Beà et al., 2013; Kandoth et al., 2013; Lafave and Levine, 2013; Lawrence et al., 2014; Mullighan et al., 2011). The focus of this work is the histone methyltransferase, *NSD2*, which is mutated 5-10% of relapsed ALL cases, particularly in those harboring the translocations E2A-PBX1 and TEL-AML1 (Ding et al., 2017; Jaffe et al., 2013; Loh et al., 2013).

1.3. E2A-PBX1 in ALL

The E2A-PBX1 fusion oncoprotein is generated as a result of translocation t(1;19), and is found in 5% of pediatric ALL (Greaves and Wiemels, 2003). Generally, patients with this E2A-PBX1 ALL fare well, though there is a strong potential for central nervous system (CNS) infiltration, increasing morbidity and mortality due to both tumor effects and toxicity associated with treatment (Alsadeq and Schewe, 2017; Gaynes et al., 2017; Roberts and Mullighan, 2015). The activation domain of the E2A transcription factor is fused to the DNA-binding motif of homeobox protein, PBX1 (Pre-B cell leukemic homeobox1; (Hunger et al., 1991)). E2A plays a critical role in B cell development, but when fused to PBX1, it instead binds to and activates transcription of PBX1 targets, activating genetic programs regulating cell proliferation and cell cycle (Andersson et al., 2005; Aspland et al., 2001). Though E2A-PBX1 is found in B-ALL, when the fusion oncoprotein is introduced into bone marrow progenitors, it induces acute myeloid leukemia (AML) instead (Kamps and Baltimore, 1993). Enforced expression in the lymphoid compartment induces T-ALL (Dedera et al., 1993). A mouse model for E2A-PBX1 B-ALL was eventually generated by driving E2A-PBX1 expression in the lymphoid compartment while preventing the development of T cells by eliminating the expression of CD3ε (Bijl et al., 2005). Though these mice do recapitulate B-ALL, there is significant delay before onset of disease (average: 403 days).

1.4. PI3K/Akt pathway activation in ALL

In ALL, as in many other malignancies, the phosphoinositide 3-kinase/protein kinase B (PI3K/AKT) pathway is significantly disregulated (Kandoth et al., 2013; Neri et al., 2014; Silva et al., 2011). Translocations, mutations, and copy-number variations (CNV) affecting the pathway are detected at many levels. For example, mutation of the negative regulator phosphatase and tensin homolog (PTEN) is a common event in T-ALL that supports leukemia survival by maintaining PI3K/AKT activation (Silva et al., 2008). Activated PI3K converts phophatidylinositol-4,5-bisphosphate (PIP2) to phosphatidylinositol-3,4,5-trisphosphate (PIP3), which in turn recruits AKT to the plasma membrane (Cantley, 2002). Phosphorylation by 3-phosphoinositide-dependent protein kinase 1 (PDK1) then activates AKT, allowing it to regulate many different pathways. Additional phosphorylation by mTOR and DNA-dependent protein kinase (DNA-PK) can activate AKT even further and unlock more downstream processes.

Activated AKT regulates many different cellular functions that are important for oncogenesis, including metabolism, apoptosis, cell cycle, growth, angiogenesis, and metastasis (Altomare and Testa, 2005; Chang et al., 2003).

1.5. BCR-ABL in leukemia

The most well studied example of unrestrained RTK signaling in cancer is from the fusion protein, BCR-ABL, resulting from translocation of chromosomes 9 and 22, t(9;22) (q34;q11), commonly known as the Philadelphia chromosome (Salesse and Verfaillie, 2002). Most frequently observed in chronic myeloid leukemia (CML), BCR-ABL is also found at high frequency in adult ALL (Moorman et al., 2007). An inhibitor, imatinib (Gleevec), was developed to inhibit the tyrosine kinase, selectively killing cells harboring *BCR-ABL* (Druker et al., 1996). This drug proved to be revolutionary for patients with CML, driving remission in 95% of patients 6 years after beginning treatment (Gambacorti-Passerini et al., 2011). Eventually imatinib was approved for other malignancies characterized by BCR-ABL, including Philadelphia chromosome-positive (Ph+) childhood and adult ALL, and new generations of TKIs were developed to provide even more effective and durable treatment. Still, resistance to TKIs, often due to mutations in ABL itself, is not uncommon (Milojkovic and Apperley, 2009). Critically, BCR-ABL activates many signaling pathways, while PI3K-AKT activation is essential for transformation (Skorski et al., 1997). As such, PI3K-AKT can become constitutively activated in TKI resistant ALL and CML due to mutations in various components of the pathway, allowing cells to no longer be reliant on BCR-ABL (Quentmeier et al., 2011). Inhibition of PI3K and PI3K-mTOR together has been shown to be effective at reducing proliferation of these cells (Quentmeier et al., 2011; Tasian et al., 2017).

1.6. TEL-AML1 in ALL

The gene fusion *TEL-AML1* (*ETV6-RUNX1*) is the most common translocation in childhood B-ALL, making up 25% of cases, and also results in PI3K/AKT pathway activation (Romana et al., 1995; Shurtleff et al., 1995). This fusion results from a translocation between chromosomes 12 and 21, t(12;21)(p13;q22), bringing together transcription factors, AML protein 1 (AML1, RUNX1) and Ets variant 6 (TEL, ETV6). A major function of TEL-AML1 is to bind AML1 targets and form stable repressor complexes, leading to widespread repression of genes such as those involved in cellular activation, immune response, apoptosis, development, and differentiation (Fuka et al., 2011; Zelent et al., 2004). However, PI3K/AKT pathway regulated genes are activated in TEL-AML1 ALL (Andersson et al., 2005). Inhibition of the pathway increases apoptosis and sensitizes resistant cells to glucocorticoid therapy (Fuka et al., 2012). The mechanism of PI3K/AKT activation is not fully understood, but one path is downstream of erythropoietin signaling. Erythropoietin receptor (EPOR) is directly upregulated by TEL-AML1, conferring increased survival and decreased sensitivity to glucocorticoids primarily mediated by PI3K/AKT (Inthal et al., 2008). Again, suggesting PI3K/AKT pathway inhibition as a promising therapy for relapsed, resistant TEL-AML1 ALL.

1.7. NOTCH1 activation in ALL

PI3K/AKT pathway inhibition also appears promising for T-ALL. Activating mutations in *NOTCH1*, a key factor in T cell development, characterize 60% of T-ALL. Cleavage of the cell surface domain of NOTCH1 allows the intracellular domain to migrate to the nucleus and activate transcriptional programs that activate various cancer-associated programs. While anti-NOTCH1 therapies have shown some promise in pre-clinical models, resistance was noted in cells harboring PTEN mutations (Palomero et al., 2007; Real and Ferrando, 2009). Exogenous activation of AKT also led to resistance of NOTCH1 inhibition, while inhibition of AKT led to sensitization. Inactivating mutations in PTEN and activating mutations in PI3K components and AKT are common in T-ALL, but are not significantly enriched with NOTCH1 mutations, indicating that the PTEN/PI3K/AKT pathway is implicated in T-ALL pathogenesis beyond its relationship with NOTCH1 (Gutierrez et al., 2009). Indeed, these mutations also confer resistance to glucocorticoids, one of the mainstays in treatment of lymphoid malignancy (Piovan et al., 2013).

1.8. Utilizing CRISPR/Cas9 to understand the genetic basis of cancer

Cell lines are a valuable resource for understanding disease. However, comparisons between cell lines has always been complicated by significant genetic heterogeneity, particularly in cancer cells, which can harbor chromosomal translocations as well as hundreds of mutations. Recent advances in gene editing technology have developed a set of powerful tools for disrupting genes and introducing specific mutations that improve the data generated from cell lines and the cost-effectiveness and ease-of-use for many animal models. These technologies are based around the generation of double stranded breaks (DSBs) in DNA, which are repaired by endogenous DNA repair pathways. Non-homologous end joining (NHEJ) is inherently error-prone, and is exploited to induce non-specific insertions and deletions (indels) that can knockout a gene (Bibikova et al., 2002; Jasin, 1996). Homologous recombination (HR) is employed to introduce specific mutations or sequences by using an exogenously introduced DNA template for repair (Jasin, 1996).

The next generation of gene editing tools harnesses the bacterial adaptive immune system known as CRISPR (clustered regularly interspaced short palindromic repeats) to make specific, RNA-guided cuts to DNA. While predecessors such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) are effective at cutting DNA, the specificity of the nuclease is mediated by protein-DNA interactions, which are sometimes hard to predict with accuracy (Juillerat et al., 2014; Ramirez et al., 2008). CRISPR, with the endonuclease Cas9 (CRISPR associated protein 9), utilizes a CRISPR-RNA (crRNA) to confer specificity, and a trans-activating crRNA (tracrRNA) that hybridizes with the cRNA to aid in complex maturation with ribonuclease III and Cas9 (Brouns et al., 2008; Deltcheva et al., 2011). Also aiding in specificity is the protospacer-adjacent motif (PAM) at the 3' end of the target site, a DNA sequence that can vary between Cas proteins (Hsu et al., 2014). For example, the widely used, spCas9, from *Streptococcus pyogenes*, recognizes an –NGG motif (Zhang et al., 2014). Later, combined crRNAs and tracrRNAs, called single guide RNAs (sgRNAs) were also developed, but the principle was the same, scientists could direct gene editing by designing RNAs with complementary binding to their target locus of interest (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013). The applications of easy and efficient gene editing are numerous, but for the work detailed herein CRISPR was useful to isolate a single point mutation and study its effects.

1.9. Compaction and storage of DNA affects its function

DNA is one of the most compact and efficient forms of information storage known. Human nuclei hold 3 billion base pairs, yet can be as small as a few microns in diameter. The compaction of the DNA is not random, however, but is instead organized in many levels, each of which influences how the DNA is utilized. Epigenetics, in part, is the study of this process; heritable changes that do not affect the DNA sequence but influence the cell fate and patterns of gene expression. Contributing to this process are DNA methylation, histone modifications, chromatin remodeling, and even noncoding RNAs (Allis and Jenuwein, 2016). The work detailed in this thesis is focused on posttranslational modifications of histone tails that affect the structure of the chromatin, altering the accessibility of genes and whether or not they will be transcribed. Chromatin regulation gives genetically identical cells the ability to express dramatically different transcriptional programs. In addition, chromatin structure can affect DNA mutation rate and repair capacity (Allis and Jenuwein, 2016). The critical nature of these marks makes epigenetic regulators prominent targets for mutation and dysregulation in cancer (Berdasco and Esteller, 2010; Shah et al., 2014; Suvà et al., 2013). Indeed, with the advances in NGS allowing for sequencing of tumors on an unprecedented scale, we are beginning to realize just how frequently chromatin regulators are mutated. Pan-cancer analysis shows that chromatin regulators are among the most commonly mutated genes in cancer (Kandoth et al., 2013; Lawrence et al., 2014).

1.10. Histone proteins and variants

Facilitating the compaction of DNA are histone proteins. The nucleosome core particle consists of an octomer of the core histone proteins, H2A, H2B, H3, and H4, wrapped twice by approximately 146 base pairs of DNA. The exposed histone tails can be chemically modified, affecting transcription, DNA repair, accessibility, and replication (Chen and Dent, 2014). The positions of the nucleosomes themselves as well as the composition of the octomer can also have similar effects. Histone proteins, particularly H2A and H3, have variant forms that perform different functions relevant for both development and disease. Canonical histone proteins are primarily deposited during DNA synthesis, but variants can also be incorporated into the nucleosome outside of synthesis (Henikoff and Smith, 2015). The difference in sequence between a canonical and a variant histone is sometimes minimal, but it can alter the physical properties of the nucleosome and the way it interacts with other proteins, affecting chromatin compaction, DNA repair, and transcription (Ausio, 2006). Histone H3 has 5 known non-canonical mammalian variants. The canonical proteins, H3.1 and H3.2, only differ from H3.3 by 5 and 4 amino acids, respectively, while centromere protein A (CENP-A) and the primate-specific H3.X and H3.Y vary dramatically (Szenker et al., 2011). CENP-A has a specialized role, replacing canonical H3 at centromeres and is critical for centromere definition and function (Black et al., 2004; McKinley and Cheeseman, 2016).

Despite its similarities to H3.1/2, H3.3 performs a wide variety of functions. The proteins DAXX (death-domain associated protein) and ATRX (alpha-thalassaemia/mental retardation X-linked syndrome protein) form a complex that target and chaperone H3.3 to transcriptionally silenced regions such as telomeres and pericentromeric heterochromatin (Goldberg et al., 2010). The histone chaperone protein, HIRA (histone cell cycle regulator), replaces canonical H3 at promoters of actively transcribed genes, gene bodies, and enhancer elements (Goldberg et al., 2010). As such, it follows that knockout of H3.3 in mouse embryos leads to defects in chromosome segregation and chromatin over-condensation, highlighting the role of H3.3 in maintaining both centromere structure and open chromatin, while leading to developmental arrest in the morula stage (Jin et al., 2009; Lin et al., 2013). H3.3 is also important for aging post-mitotic cells, accumulating in liver, kidney, brain, and heart as development progresses (Tvardovskiy et al., 2017). Active histone modifications such as H3K4me1, H3K4me3, H3K9ac, H3K27ac, and H3K36me1/2 are commonly enriched on H3.3 relative to H3.1/2 (Kraushaar et al., 2013; Szenker et al., 2011); however, H3K36me2 actually increases in proportion with H3.3 in aging cells, indicating the interdependence of mark and histone variant (Tvardovskiy et al., 2017). Knockout of H3.3 leads to dramatic, global reductions of H3K36me2 (Lin et al., 2013). Additionally, H3.3K36 methylation appears to be critical for H3.3 function, as exogenous expression of mutant H3.3 that cannot be methylated at lysine 36 (H3.3K36R) does not restore proper chromosome segregation and condensation (Lin et al., 2013). The H3.3K36R mutant was found to incorporate poorly at stimulated genes (Sarai et al., 2013). Wildtype H3.3 is typically deposited at stimulated genes by HIRA-bound RNA polymerase II (RNA Pol II) with transcriptional elongation (Ray-Gallet et al., 2011). Interestingly, HIRA is initially recruited to stimulated genes by nuclear SET domain-containing protein 2 (NSD2, MMSET, WHSC1), a histone methyltransferase responsible for dimethylating H3K36, although it has been suggested that the methyltransferase activity is not required for H3.3 deposition (Sarai et al., 2013). Deposition of H3.3, possibly with the help of H3K36me2, leaves a long-term transcriptional mark that could result in faster gene activation with future stimulation.

1.11. Epigenetic modification of histone proteins

Chemical modifications of histones can recruit DNA repair machinery, scaffolding proteins, nucleosomal remodelers, and transcription factors. These marks are placed by epigenetic writers and removed by epigenetic erasers. Acetylation can alter chromatin accessibility by altering the physical properties of nucleosomes (Gorisch et al., 2005). However, most histone marks are interpreted by epigenetic readers – proteins that contain domains to recognize the histone code established by the writers and erasers and facilitate downstream effects (Yun et al., 2011). The diversity and utility of histone marks is enormous, as the number of potentially modified residues along with the variety of modifications allows for many combinations and provides another layer of complexity in the epigenetic code.

This epigenetic code has proven valuable for the identification of genomic elements. While inactive enhancers may carry no marks, primed enhancers are defined by monomethylation of H3 lysine 4 (H3K4me1) (Heinz et al., 2015). A poised enhancer is also marked by trimethylation of lysine 27 on histone 3 (H3K27me3), but acetylation of the same residue (H3K27ac) is associated with enhancer activation. H3K27me3 is also a repressive mark found at gene promoters (Boyer et al., 2006). Active promoters can be decorated with a wide array of marks, but are broadly identified by H3K4 trimethylation (Pokholok et al., 2005). Actively transcribed gene bodies, meanwhile, are trimethylated at H3 lysine 36 (H3K36me3) (Barski et al., 2007). The decreasing cost and improving technology of genome wide techniques such as chromatin immunoprecipitation followed by microarray or next-generation sequencing (ChIP-chip or ChIP-seq) has allowed for the annotation of the genome in various cell types and disease states (Mikkelsen et al., 2007).

Intriguingly, epigenetic marks are commonly misregulated in disease, particularly cancer (Popovic and Licht, 2012). The regulators are common targets for cancer-associated mutations, but mutations have also been detected in histone tails themselves, highlighting their fundamental importance. Mutations in the genes encoding histones H3.1 or H3.3 were detected in more than half of diffuse intrinsic pontine gliomas (DIPGs) and supratentorial glioblastoma multiforme (GBMs) (Schwartzentruber et al., 2012; Wu et al., 2012). The mutation results in a lysine to methionine change at residue 27 of H3 (H3K27M). The mutant histone binds and inhibits the function of the histone methyltransferase, EZH2 (enhancer of zeste homolog 2), which is responsible for methylating H3K27 (Lewis et al., 2013). Therefore, despite the mutant protein only making up 4% of the total H3 produced in affected cells, this was sufficient to alter the chromatin pattern of H3K27me3 and gene expression, promoting gliomagenesis (Chan et al., 2013). When not mutated, H3K27 can be methylated by the polycomb repressive complex 2 (PRC2), which consists of EED, EZH1, EZH2, and SUZ12 (Cao et al., 2002). Trimethylation of H3K27 recruits epigenetic readers in the polycomb repressive complex 1 (PRC1). PRC1 can then further modify chromatin by ubiquitinylating histone H2A lysine 119 (H2AK119ub) or by compacting chromatin directly (Eskeland et al., 2010; Margueron and Reinberg, 2011).

Another lysine to methionine mutation was detected in 90% of chondroblastomas; this time only affecting H3.3K36 (H3.3K36M) (Fang et al., 2016). Similarly, H3.3K36M had an outsized effect.

By inhibiting the methyltransferases NSD2 and SETD2 (SET domain containing 2), H3.3K36M altered global H3K36 methylation and gene expression, thereby promoting oncogenesis.

1.12. Structure and function of the histone methyltransferase, NSD2

NSD2 is a histone methyltransferase that catalyzes the mono- and dimethylation of H3K36 at active gene promoters (Martinez-Garcia et al., 2011; Morishita et al., 2014). NSD2 plays an important role in development and disease, as do HMT family members NSD1 and NSD3 (WHSC1L1). In fact, NSD2 was initially identified by its role in Wolf-Hirschhorn syndrome (WHS), a congenital disorder characterized by deletion of a region of the short arm of chromosome 4 (Stec et al., 1998). Concurrently, NSD2 was identified as the translocation partner with the immunoglobin locus in translocation t(4;14) multiple myeloma (Chesi et al., 1998). This 90-kb, 25-exon gene on chromosome 4p16.3 encodes two main isoforms, NSD2-short (MMSET-I) and NSD2-long (MMSET-II), and an intronic transcript that encodes response element II–binding protein (RE-IIBP). The 1365 amino acid full-length species is composed of two PWWP domains, a high-mobility group (HMG) DNA-binding domain, four PHD zinc fingers, and a SET domain, whereas the 647 amino acid short species lacks all but the initial PWWP and HMG domains (Stec et al., 1998). The first amino-terminal PWWP domain of NSD2 specifically binds to H3K36me2 to stabilize NSD2 at chromatin, and the catalytic SET domain of NSD2 propagates this gene-activating mark to adjacent nucleosomes (Kuo et al., 2011; Martinez-Garcia et al., 2011; Morishita et al., 2014; Sankaran et al., 2016). Similar to family member, NSD1, the NSD2 post-SET domain is attached to the catalytic SET domain via an autoinhibitory loop region and inhibition is relieved on nucleosome binding (Poulin et al., 2016a; Qiao et al., 2011).

There has been much controversy over the targets of NSD2 methylation, with H4K20, H3K4, H3K27, and H3K36 all being implicated, but new studies using nucleosomal substrates indicate that NSD2 specifically catalyzes H3K36 dimethylation (Kuo et al., 2011; Li et al., 2009; Poulin et al., 2016b). The PHD domains help target NSD2 to particular genetic loci (Huang et al., 2013). The catalytic activity of NSD2 is not always required for gene activation. Sarai et al. show that interferon- (IFN) and UV-responsive genes accumulate the epigenetic reader, BRD4 (bromodomain-containing protein 4) upon stimulation, which recruits NSD2 and the positive transcription elongation factor b, P-TEFb. NSD2 appears to aid in transcription elongation and recruits HIRA, leading to H3.3 deposition following elongating RNA Pol II as described above. Upon knockout of NSD2, deposition of H3.3 is lost and transcription of IFN and UV responsive genes is reduced, but reconstitution of the system with wildtype or catalytically inactive NSD2 rescues the phenotype (Sarai et al., 2013). In multiple myeloma with NSD2 overexpression, however, when NSD2 is knocked out gene expression can only be rescued by catalytically active NSD2, indicating that there are multiple functions of NSD2 and multiple pathways by which NSD2 can activate transcription (Kuo et al., 2011).

1.13. NSD2 in development

NSD2 is broadly expressed and its importance in development is highlighted by its involvement in Wolf–Hirschhorn malformation syndrome (WHS). The full syndrome is characterized by brain defects associated with developmental delay and epilepsy as well as craniofacial anomalies, growth delay, heart defects, and midline fusion abnormalities (Battaglia et al., 1993; Bergemann et al., 2005; Stec et al., 1998). Variable deletions in the short arm of chromosome 4 (4p16.3) are typical of WHS and NSD2 is the only gene in this region that is deleted in almost every case.

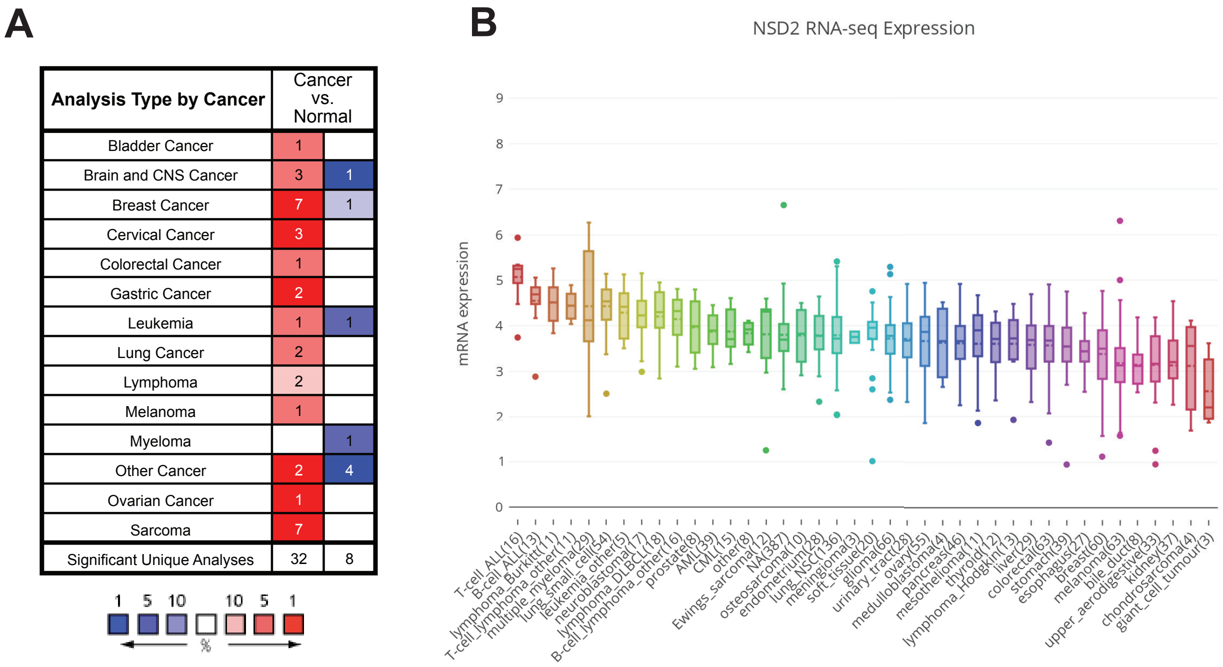
Partial or full hemizygosity of NSD2 appears to be necessary but not sufficient for the development of WHS, as the deletion of other genes nearby contributes to the constellation of abnormalities that make up the syndrome (Andersen et al., 2013; Bergemann et al., 2005). Mice with a homozygous NSD2 SET domain deletion do not survive past 10 days of age, and heterozygous mice develop significant developmental defects that imitate WHS (Nimura et al., 2009). ChIP experiments on embryonic stem (ES) cells of these mice revealed that NSD2 binds to several genes associated with development including Sall1, Sall4, and Nanog. Additionally, WHS patients can have antibody deficiencies, which first suggested a role for NSD2 in B-cell development (Hanley-Lopez et al., 1998). It has been reported that NSD2 recruits the DNA-damage responsive, p53-binding protein 1 (53BP1), which is critical for class switch recombination (CSR), and this was suspected to play a role in antibody deficiency found in WHS patients (Hajdu et al., 2011; Pei et al., 2011). Antibody class is a critical component of the immune system that allows antibodies to take on different functions. CSR selects the heavy chain constant region that encodes the antibody class by activating activation-induced cytidine deaminase (AID)(Stavnezer et al., 2008). AID induces mutations that lead to double strand breaks (DSBs). The resulting repair by recombination can select different heavy chain genes depending on the signals received, but requires 53BP1 to occur (Manis et al., 2004). Knock down of NSD2 results in a loss of 53BP1 recruitment to DSBs, thereby preventing effective CSR (Pei et al., 2013). While this might explain poor antibody production in WHS patients, a recent study showed that NSD2 plays a far more integral role in the hematopoietic cell development and function (Campos-Sanchez et al., 2017). Self-renewal of HSPCs (hematopoietic stem/progenitor cells) and their ability to reconstitute the immune system was severely compromised with NSD2 knockout. Many stages of B cells were affected profoundly, exhibiting reduced numbers and poor function. CSR was reduced in these cells, as were important B cell transcription factors. Most malignancies affected by NSD2 dysregulation are derived from B cells, so it follows that B cell lineage specification and proliferative capacity was compromised with NSD2 knockout.

1.14. NSD2 in cancer

As mentioned above, NSD2 was initially described as a gene rearranged and linked to regulatory sequences of the immunoglobulin heavy chain gene in t(4;14) multiple myeloma (Chesi et al., 1998; Stec et al., 1998); NSD2 and its translocations, amplifications, and mutations were subsequently identified in a wide spectrum of malignancies. In multiple myeloma, the t(4;14) translocation is present in 15%–20% of cases, resulting in overexpression of NSD2 and FGFR3 (Finelli et al., 1999; Keats et al., 2005). However, NSD2 is purported to be the primary oncogenic driver, as approximately 30% of cases harboring this translocation have normal expression of FGFR3 alongside NSD2 overexpression (Santra et al., 2003). Furthermore, knockdown of NSD2 expression in t(4;14) multiple myeloma cell lines reduces proliferation, cell-cycle progression, and DNA repair, while increasing apoptosis and adhesion (Brito et al., 2009; Huang et al., 2013; Lauring et al., 2008; Martinez-Garcia et al., 2011; Shah et al., 2016). These phenotypic changes are driven by redistribution of activating and repressive chromatin marks that, in turn, affect gene expression. Overexpression of NSD2 increases levels of H3K36me2 and decreases levels of H3K27me3, aberrantly activating oncogenic programs by inhibiting gene repression that occurs through EZH2-mediated dimethylation of H3K27 (Kuo et al., 2011; Martinez-Garcia et al., 2011; Popovic et al., 2014; Sankaran et al., 2016; Yuan et al., 2011; Zheng et al., 2012). Typically found at the 5’ end of genes, H3K36me2 from NSD2 overexpression spreads across gene bodies and into intergenic regions, potentially disrupting the function of genomic elements such as enhancers that rely on specific chromatin signatures. Abundant H3K36me2 restricts EZH2 to small islands of chromatin where it then hypermethylates H3K27me3 (Popovic et al., 2014; Zheng et al., 2012). The transcriptional disturbance that results from NSD2 overexpression primarily involves inappropriate activation of genes, but there are some genes that are inappropriately repressed because of these pockets of EZH2. This latter effect also appears to be important for the survival of cells overexpressing NSD2, as they are sensitive to inhibition of EZH2. One contribution to this sensitivity is that inhibition of EZH2 decreases c-MYC protein levels, a fundamentally upregulated gene in multiple myeloma (Popovic et al., 2014). In t(4;14) multiple myeloma cells EZH2 represses miR-126, a microRNA that targets the MYC transcript (Min et al., 2012). Therefore, EZH2 inhibitors derepress miR-126, allowing it to reduce c-MYC protein levels and slow proliferation. However, c-MYC is not the sole cause of high NSD2-mediated oncogenesis. Transcriptional profiling of t(4;14)+ multiple myeloma indicates that NSD2 regulates the expression of genes in apoptosis, DNA repair, cell-cycle control, and cell motility.

The role of NSD2 in DNA damage repair (DDR) has important implications in MM. Multiple myeloma is characterized by genetic instability and DNA damage, and indeed, the translocations and mutations resulting from erroneous repair are part of the pathogenesis of this disease (Gourzones-Dmitriev et al., 2013). Shah and colleagues demonstrated that cells with high NSD2 are more resistant to DNA damaging agents (Shah et al., 2016). At baseline these cells are uniquely affected by DNA damage, but in response to DNA damaging agents they are able to survive and repair better than cells with low NSD2. High-NSD2 cells also exhibited more efficient non-homologous end joining (NHEJ) and homologous recombination (HR). This was likely due to the ability of NSD2 to recruit repair proteins to DSBs as well as regulate transcription of repair genes such as RAD51 and 53BP1.

In addition to multiple myeloma, a recent report indicated that NSD2 is significantly upregulated B-ALL patients relative to normal B cells (Ding et al., 2017). Additionally, those with increased NSD2 expression were more likely to relapse earlier after treatment. An analysis of the Oncomine database and Broad Cancer Cell Line Encyclopedia shows that *NSD2* is overexpressed several different malignancies (Figure 1.14 A-B). Immunostaining revealed that NSD2 protein is overexpressed in gastric, colon, lung, and skin cancer as well as squamous cell carcinoma of the head and neck (Hudlebusch et al., 2011a; Saloura et al., 2015). It may play a major role in neuroblastoma and breast, bladder, and prostate tumors, where overexpression is associated with worse prognosis (Hudlebusch et al., 2011b). In many of these cancers, NSD2 expression is positively correlated with EZH2 expression, and as opposed to the situation in t(4;14) multiple myeloma, global levels of H3K27me3 and H3K36me2 are both increased (Asangani et al., 2013). In prostate cancer, EZH2 functions upstream of NSD2, repressing several microRNAs, including miR-203, miR-26, and miR-31, that target NSD2. However, it is the knockdown of NSD2 that abrogates the ability of prostate cancer cells to proliferate, form colonies, migrate, and invade (Ezponda et al., 2013). Furthermore, enforced expression of NSD2 in nontransformed prostate epithelial cells promotes migration and invasion and leads to epithelial–mesenchymal transition (EMT). NSD2 directly binds the TWIST1 locus and up-regulates expression of this EMT factor, which plays a key role in the aggressive biological behavior of advanced prostate cancer.



Analysis Type by Cancer	Cancer vs. Normal	
Bladder Cancer	1	
Brain and CNS Cancer	3	1
Breast Cancer	7	1
Cervical Cancer	3	
Colorectal Cancer	1	
Gastric Cancer	2	
Leukemia	1	1
Lung Cancer	2	
Lymphoma	2	
Melanoma	1	
Myeloma		1
Other Cancer	2	4
Ovarian Cancer	1	
Sarcoma	7	
Significant Unique Analyses	32	8

Figure 1.14. NSD2 is highly expressed in cancer

A, From Oncomine database showing number of datasets where NSD2 expression is significantly overexpressed (red) or underexpressed (blue) in cancer vs. normal. Color of box corresponds to best gene rank percentile. Adapted and updated from Kuo, 2011.

B, boxplot of NSD2 expression from CCLE RNA-seq dataset across cancer cell line lineages.

1.15. The NSD2-activating point mutation, E1099K

In some cases of ALL, NSD2 harbors a recurrent heterozygous point mutation in the SET domain (Oyer et al., 2014). The significance of the guanine to alanine substitution that results in a glutamic acid to lysine switch at amino acid 1099 (E1099K) of NSD2 was first noted following examination of the Broad Institute's Cancer Cell Line Encyclopedia (CCLE) that revealed numerous ALL cell lines with the mutation, which has since been identified in 5%–10% of relapsed pediatric ALL and 10% of chronic lymphocytic leukemia (CLL) (Barretina et al., 2012; Fabbri et al., 2011; Loh et al., 2013). Furthermore, this mutation, together with another NSD2 mutation, T1150A, was reported 12% of mantle cell lymphomas (MCL) (Beà et al., 2013). Rare mutations are found in glioblastoma, lung cancer, and MM. Even in ALL the mutation appears to be a rare subclone at diagnosis, present at 0.01 mean allele frequency (MAF), but it grows to 0.49 MAF at relapse (Ma et al., 2015). This suggests that NSD2^{E1099K} is involved in resistance to therapy and progression of disease rather than initiation. Cell lines harboring the E1099K mutation exhibited increased H3K36me2 and decreased H3K27me3 relative to NSD2^{WT} cells, similar to the epigenetic profile of t(4;14) MM (Oyer et al., 2014). Exogenous expression of the mutant protein in MM cells enhanced their colony formation capacity more than wildtype protein (Jaffe et al., 2013).

In summary, NSD2 plays a significant role in normal development and malignancy. Haploinsufficiency of NSD2 leads to severe developmental defects, such as cardiac lesions and midline abnormalities associated with WHS. In malignancy, NSD2 has prolific effects resulting from translocation, overexpression, and activating mutations of the gene. The t(4;14) translocation in MM has been extensively characterized and indicates that high levels of NSD2 drive oncogenic phenotypes by spreading H3K36me2 throughout the genome and altering gene-expression profiles. The relationship between NSD2 and EZH2 is crucial in myeloma as the reciprocal H3K27me3 mark deposited by EZH2 is reduced and restricted to certain regions by NSD2 overexpression. In other malignancies, such as prostate cancer, this relationship is different as EZH2 appears to function upstream of NSD2, although it is no less important. Inhibition or knockdown of NSD2 or EZH2 in both situations abrogates critical oncogenic pathways and phenotypes. However, there is still much to learn about NSD2 in these malignancies. Although NSD2 itself is already an attractive therapeutic candidate, thoroughly characterizing its binding partners will tell us how NSD2 functions both endogenously and in malignancy and provide more targets for de- signing therapies for NSD2-misregulated cancers.

The recurrent E1099K mutation that results in a hyperactive NSD2 is also not fully understood. Its global effects on chromatin mimic NSD2 overexpression, but it must be investigated further to identify its local effects on chromatin, gene expression, and oncogenicity. We then must understand the mechanism by which E1099K and other activating mutations alter the function of NSD2 to design directed therapies.

	NSD2 Overexpression	**NSD2 E1099K**
Cancer types	MM, prostate, gastric, colon, lung, skin, bladder, head and neck, neuroblastoma, ALL, CLL	Relapsed pediatric ALL, CLL, MCL **Rare**: MM, GBM, lung
Cause of dysregulation	**MM**: t(4;14) IgH promoter/enhancer **Prostate**: Loss of EZH2-mediated repression of miRNAs regulating NSD2	Mutation in SET domain
Histone	H3K36me2 □	Same pattern, but less

modifications	H3K27me3 □	dramatic
Genomic distribution of H3K36me2	Broad spread everywhere except EZH2/CTCF bound sites	Unknown
Biological effects	□□Proliferation, adhesion, clonogenicity, migration/invasion	Knockdown reduces ALL proliferation and MM clonogenicity
Pathways of interest	Cell cycle, DNA repair, cell death	DNA damage, B cell signaling Poorly characterized
Target genes	*TWIST1, JAM2, DSG2*	*TWIST1* (possible) Poorly characterized
Effect on EZH2 distribution	Redistributed to certain loci → gene repression	Unknown
DNA repair	Increased damage at baseline with increased repair following insult	Unknown
Enzymatic activity	Mass action: more protein yields increased mono- and dimethylation	Increased activity *in vitro* Mechanism unknown
Partner proteins	Chromatin organization, DNA repair, signaling, PTMs	Unknown

1.16. Study goals

The effects of chromatin regulators on development and malignancy are broad and widely recognized. The histone methyltransferase, NSD2, is recognized for its oncogenic properties in malignancies as varied as MM and prostate cancer. Until recently, however, it has been studied mostly in the context of overexpression. The E1099K mutation in NSD2 is still poorly characterized. With the help of my colleagues, in my dissertation work we attempt to address some of the questions highlighted above that remain unanswered regarding the role of NSD2^{E1099K} in cancer using an interdisciplinary approach that includes *in silico* protein modeling, genetically engineered cell lines, RNA-seq, mass spectrometry, and mouse models.

We find that the E1099K mutation enhances binding of NSD2 to chromatin, increasing catalysis of H3K36 monomethyation. Increased activity of mutant NSD2 activates a distinctively abnormal gene expression program that stimulates growth and other aggressive biological properties of ALL cells that may contribute to early relapse. We highlight several potential downstream effectors of mutant NSD2, including aberrantly activated signaling pathways, revealing new targets for therapy to counteract dysregulated NSD2.

Chapter 2: Results

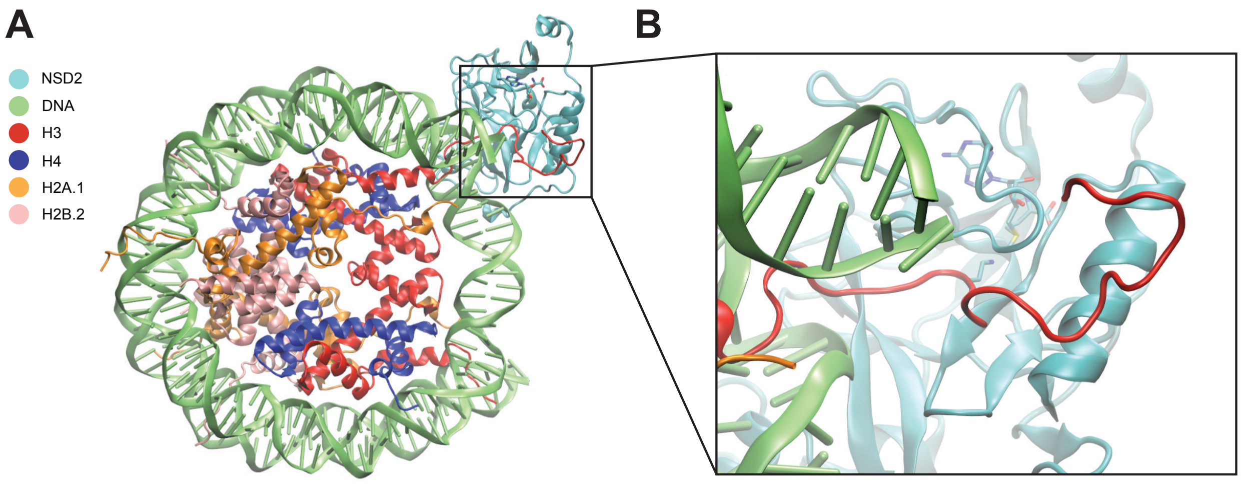
2.1. The E1099K mutation enhances NSD2 interaction with chromatin

To investigate how the E1099K mutation may alter the structure and function of NSD2, we used the structure of NSD1 (PDB ID:3OOI) (Qiao et al., 2011) and the nucleosome core particle (NCP, PDB ID: 1KX5) to construct a model in which NSD2 interacts with the H3 N-terminus and nucleosome DNA, allowing for a molecular understanding of the stabilization of the complex during H3K36 methylation (Figure 2.1.1A-B). Our model shows significant similarity to a recently published crystal structure of NSD2 that was solved with multiple point mutations, substantiating the results from our analysis (Tisi et al., 2016).

Molecular dynamics (MD) simulation predicted that the E1099K mutation makes the DNA-NSD2 interaction energy more favorable (Figure 2.1.1C, first row). The interaction of residue 1099 with other parts of the system was significantly augmented when glutamic acid is changed to lysine (Figure 2.1.1C, first row). The interaction energy of K1099 with the rest of the system was even more favorable when H3.3 was replaced with replication-associated histone variant 3.1

(Figure 2.1.1C, second row). H3.1 had a more stable interaction with DNA and NSD2 due to the presence of hydrophobic alanine instead of serine at residue 31 (Figure 2.1.1C, second row). The solvent-accessible surface area (SASA) of the NSD2 histone binding site surrounding H3K36 was predicted to be larger in the E1099K mutant allowing for more contact between H3K36 and the S-adenosylmethionine binding site required for methylation (Figure 2.1.1D). The SASA was further increased in the H3.1 model (Figure 2.1.1D). These results predict increased potential for histone methylation by mutant NSD2 relative to wildtype and suggest that nucleosomes containing H3.1 might be an even better target for methylation by NSD2 E1099K than those containing H3.3.

To determine whether mutant NSD2 has altered affinity for chromatin *in vivo*, fluorescence recovery after photobleaching (FRAP) was performed on HEK293 cells transfected with GFP-NSD2^{WT} or GFP-NSD2^{E1099K}. Following photobleaching of the nucleus, cells expressing NSD2^{E1099K} recovered fluorescence intensity at a slower rate than cells expressing NSD2^{WT} (Figure 2.1.2A). Furthermore, the amount of unbound NSD2^{E1099K}, as measured by mobile fraction, was less than that of NSD2^{WT} (Figure 2.1.2B). To further determine whether NSD2^{E1099K} exhibits increased binding to nucleosomes, direct measurement of this interaction was performed by quantifying wildtype and mutant NSD2 bound to nucleosomes in reactions containing increasing concentrations of salt (Figure 2.1.2C). While both wildtype NSD2 and NSD2^{E1099K} were bound to nucleosomes at 100mM NaCl, at 200mM NaCl wildtype NSD2 did not remain bound while mutant NSD2 maintained substantial interaction with nucleosomes (Figure 2.1.2C). Together these data indicate that mutant NSD2 has increased affinity for chromatin, which may underlie its enhanced activity.



C

	H3 interaction energy	DNA-NSD2 interaction energy	NSD2 potential energy	NSD2 residue 1099 interaction energy	H3 residue 31 interaction energy
E1099-H3.3 to K1099-H3.3	0.7% (17)	-2.1% (-231)	-2.6% (-193)	-105.5% (-259)	-17.5% (-12)
K1099-H3.1 to K1099-H3.3	6.7% (163)	0.4% (44)	1.1% (85)	-20% (-50)	11.3% (7)

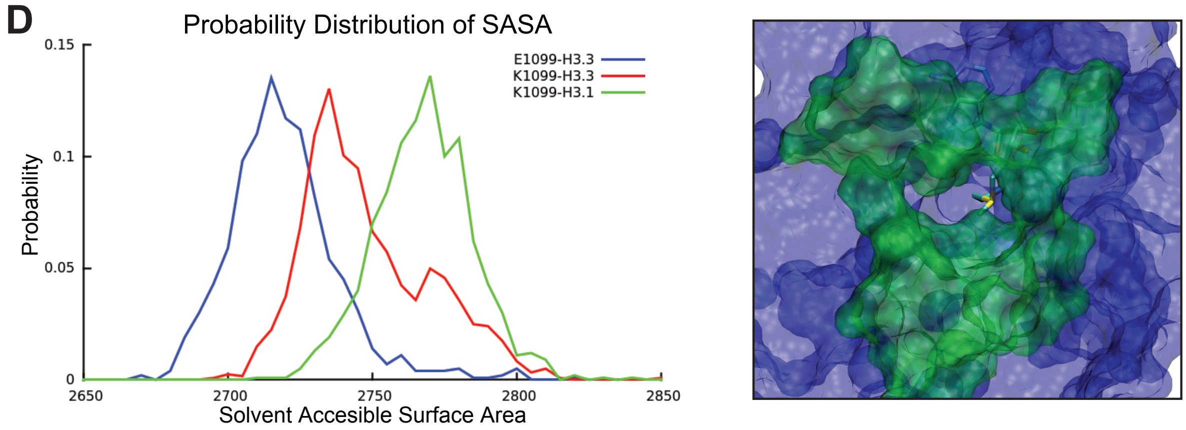


Figure 2.1.1. *in silico* modeling indicates E1099K mutation enhances NSD2 interaction with chromatin

A, in the NSD2-SAM-NCP model NSD2, DNA, Histone H3, H4, H2A.1 and H2B.2 are colored in cyan, green, red, blue, orange and pink, respectively. **B,** shows a closer view of the interaction interface between NSD2 and NCP. SAM and H3 K36 residue are shown in stick representation. **C,** the percentage difference of averaged interaction energies between three simulated systems were calculated from MD simulations. The energy difference is given in kcal/mol in parentheses such that negative values indicate more favorable interactions. **D,** probability distribution of the SASA (Å^2) of the NSD2 histone-binding site (residues M1119 to A1130 and N1178 to N1186) for the three systems. Right panel is a visual representation of NSD2 protein surface (blue) and binding region (green). SAM is shown in stick representation. *Data and figure generated by Wenbo Yu and Alex MacKerell.*

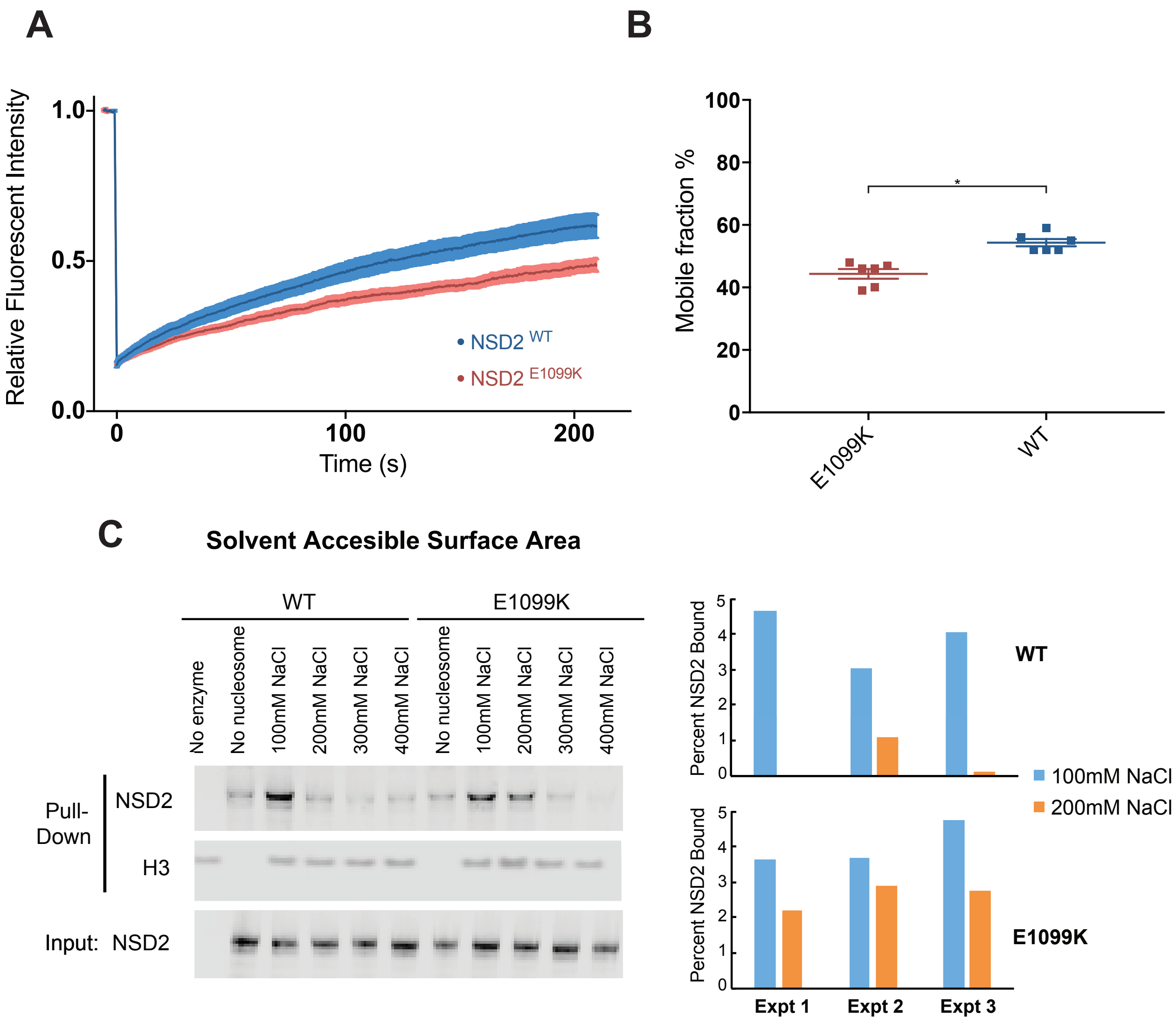


Figure 2.1.2. E1099K increases NSD2 affinity for chromatin by Fluorescence Recovery After Photobleaching

A, Representative FRAP curve comparing C terminus GFP-tagged wildtype and E1099K NSD2 transiently expressed in 293 cells. Plotted as average (n=11) ±SEM. **B,** average mobile fraction ±SEM of GFP-tagged wildtype and E1099K mutant NSD2 measured by FRAP (n=6). Two sample t-test with unequal variance, p=0.0006. **C,** H3 and NSD2 immunoblot of bound and input fractions from in vitro pulldown assay showing wildtype and mutant NSD2 bound to nucleosomes following incubation with increasing concentrations of NaCl. *C panel data and figure generated by Richard Bennett.*

2.2. Gene editing to remove the E1099K mutation restores the balance of H3K36me2 to H3K27me3

To elucidate the oncogenic effects of the E1099K mutation, we used CRISPR/Cas9 to generate isogenic cell lines differing only in the mutation status of NSD2. A single guide RNA (sgRNA) was designed to target the mutant allele of NSD2 (Figure 2.2.1.A-B). Three genotypes were isolated from two B-ALL cell lines (SEM and RCH-ACV) and one T-ALL cell line (RPMI-8402). Control, non-targeted, cell lines were heterozygous mutant ($NSD2^{WT/E1099K}$) cells that underwent transfection and subcloning procedures but did not undergo gene editing because they did not receive the sgRNA (Figure 2.2.1C, left). Two hemizygous wildtype ($NSD2^{WT/-}$) cell lines with frameshift mutations in the NSD2 E1099K allele were isolated from both SEM cells and RPMI-8402 cells (Figure 2.2.1C, right). $NSD2^{WT/-}$ RCH-ACV cells were isolated but difficult to maintain in culture and thus eventually lost. Finally, we isolated homozygous wildtype ($NSD2^{WT/WT}$) cells from RCH-ACV cell lines, which were presumably the result of interchromosomal gene conversion as no donor oligonucleotide was included in the gene editing procedure (Figure 2.2.1C, middle).

$NSD2^{WT/E1099K}$ cells had globally high H3K36me2 and low H3K27me3, while $NSD2^{WT/-}$ or $NSD2^{WT/WT}$ had low H3K36me2 and high H3K27me3 (Figure 2.2.2A). The increase in H3K27me3 was not due to changes in EZH2 expression (Figure 2.2.2B). The increased activity of NSD2 E1099K was further characterized by mass spectrometry-based measurements and modeling of histone methylation kinetics (M4K), which quantifies the rate histone modification *in vivo*. Relative to RCH-ACV NSD2$^{WT/WT}$ cells, NSD2$^{WT/E1099K}$ cells formed H3K36me1 at a six-fold higher rate on histones devoid of H3K27 modification and an eight-fold greater rate on histones bearing a monomethylated H3K27 (Figure 2.2.3A). By contrast, the rate of conversion of H3K36me1 to H3K36me2 was similar in NSD2$^{WT/E1099K}$ RCH-ACV cells and in cells where the mutation was removed (Figure 2.2.3A-B). In SEM NSD2$^{WT/E1099K}$ cells, the monomethylation rate was two-fold higher compared with NSD2$^{WT/-}$ cells on histones with unmodified H3K27 and three-fold increased when K27 was monomethylated (Figure 2.2.4A). The dimethylation rate was unchanged between mutant and wildtype cells. The overall result was a 2.5-fold increase in global levels of H3K36me2 and 5-fold decreased level of H3K27me3 in NSD2$^{WT/E1099K}$ cells as compared to NSD2$^{WT/WT}$ and NSD2$^{WT/-}$ (Figure 2.2.4A-B). Of note, upon elimination of the mutant NSD2 allele there was an accumulation of unmethylated H3K36 without an increase of H3K36me1, consistent with *in vitro* data indicating NSD2 catalyzes the mono- and dimethylation of H3K36 (Li et al., 2009). In analyzing histone variants H3.3 and H3.1, we detect high H3K36 dimethylation on histone H3.3 in NSD2$^{WT/E1099K}$ cells that remains elevated even upon elimination of the mutant allele (Figure 2.2.5A). In contrast, H3K36 dimethylation on histone H3.1 is high in NSD2$^{WT/E1099K}$ cells and significantly reduced in cells without the mutation (Figure 2.2.5B). This relationship also holds true for NSD2-overexpressing cells compared to NSD2-underexpressing cells derived from the MM cell line, KMS11 (Figure 2.2.5C). With normal expression of NSD2, H3.1K36 is not frequently dimethylated. However, in both ALL and MM, mutant NSD2 dimethylates K36 of H3.1 to levels resembling H3.3K36me2, thus facilitating the spread of H3K36me2 and inhibition of H3K27me3 across the epigenome.

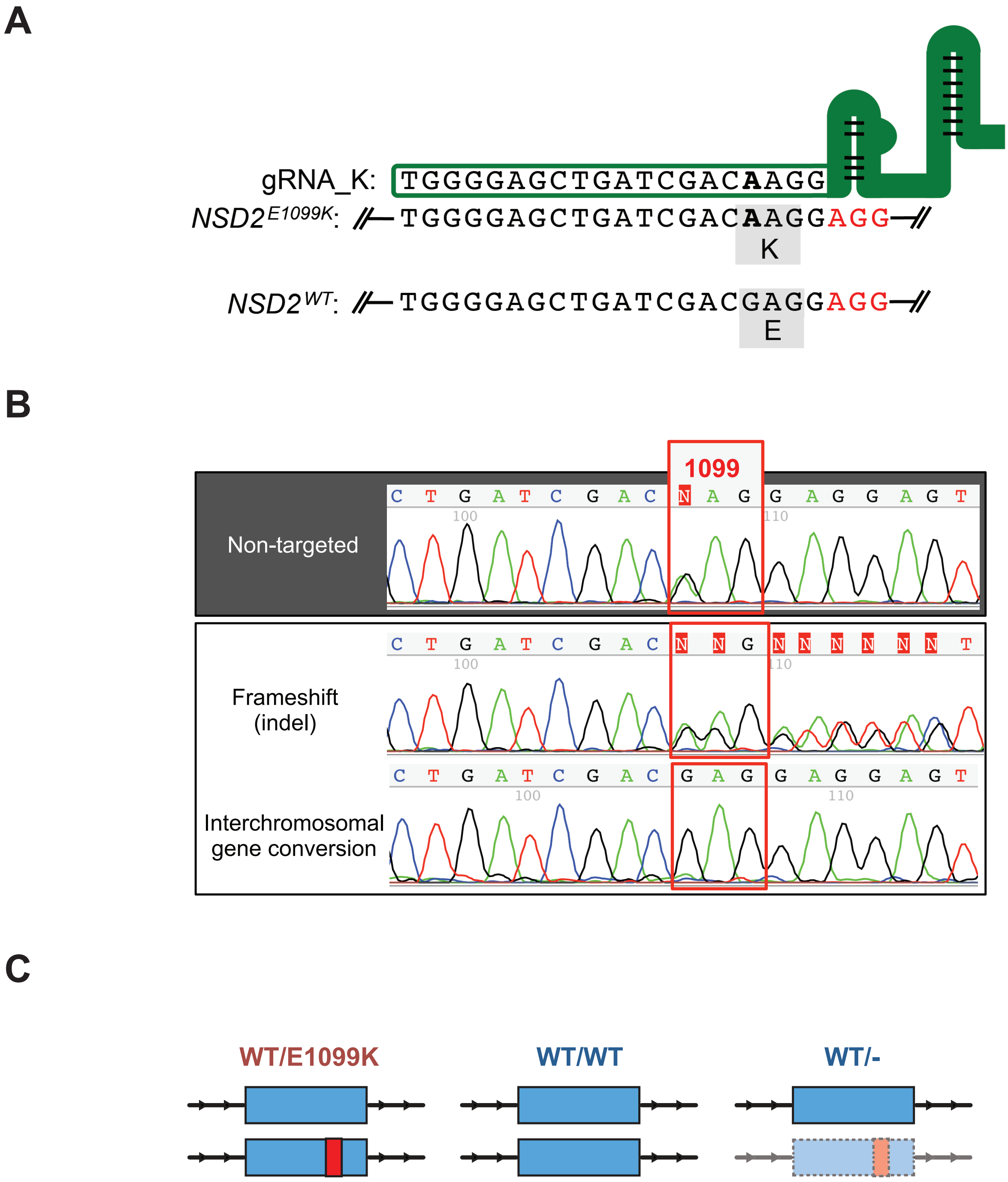

A
gRNA_K: TGGGGAGCTGATCGACAAGG
NSD2E1099K: TGGGGAGCTGATCGACAAGGAGG
K
NSD2WT: TGGGGAGCTGATCGACGAGGAGG
E
B
1099
Non-targeted
C T G A T C G A C N A G G A G G A G T
100
110
Frameshift (indel)
C T G A T C G A C N N G N N N N N N T
100
110
Interchromosomal gene conversion
C T G A T C G A C G A G G A G G A G T
100
110
C
WT/E1099K
WT/WT
WT/-

Figure 2.2.1. CRISPR/Cas9 mediated editing of NSD2 E1099K

A, diagram illustrates the design of the sgRNA targeting the mutant NSD2^{E1099K} allele while sparing the NSD2^{WT} allele. **B,** Sanger sequencing traces of RCH-ACV/SEM/RPMI-8402 NSD2 around residue 1099. Top row is the parental non-targeted cell line containing the heterozygous E1099K mutation. The second row illustrates a frameshift indel mutation at position 1099. The third row shows an interchromosomal gene conversion that results in NSD2$^{WT/WT}$ following the conversion of the K1099 allele to E1099. **C**, diagram illustrating the varieties of NSD2-mutant and wildtype subclones isolated from RCH-ACV, SEM, and RPMI-8402 cell lines. *Data and figure generated with Jon Oyer and Christine Will.*

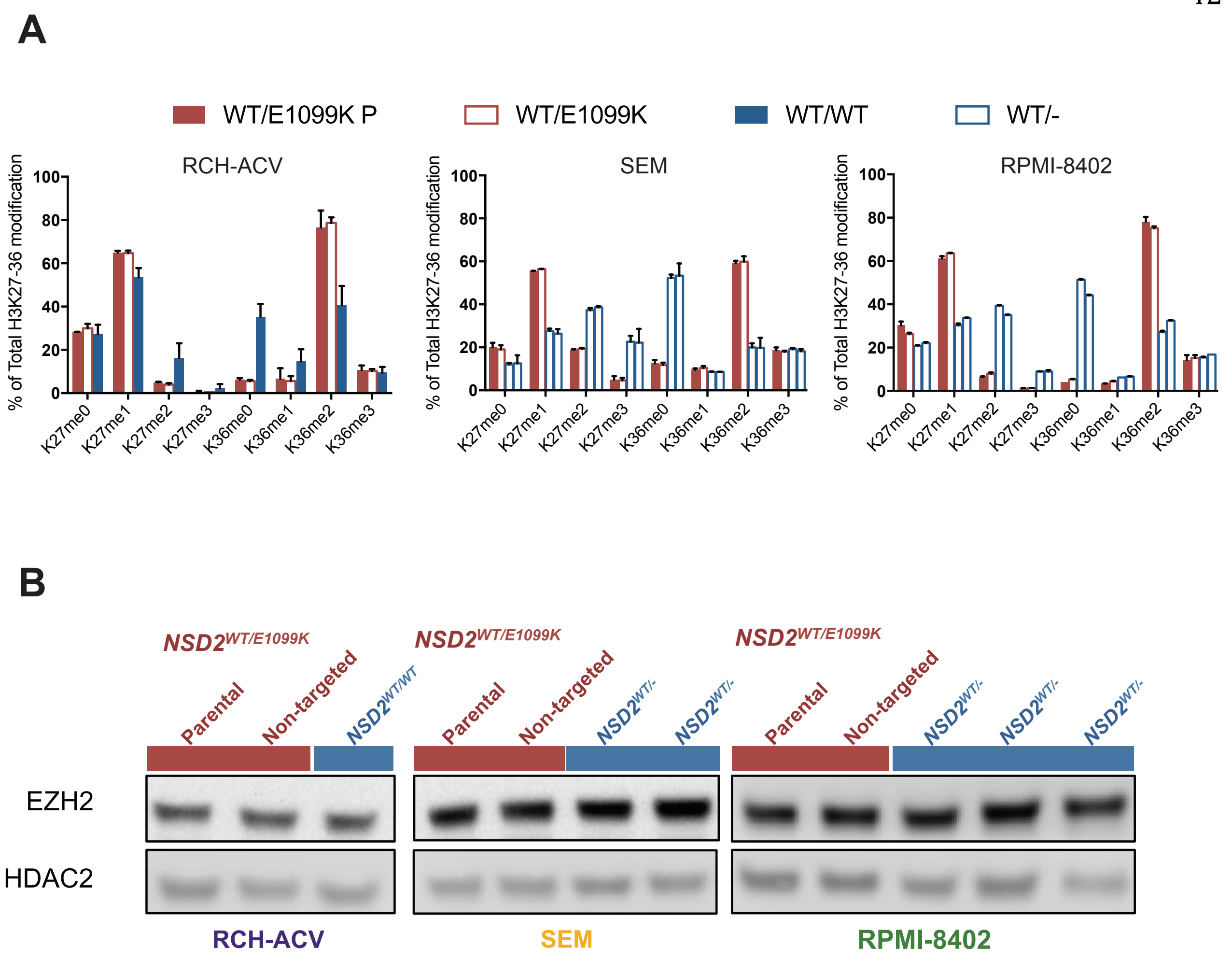
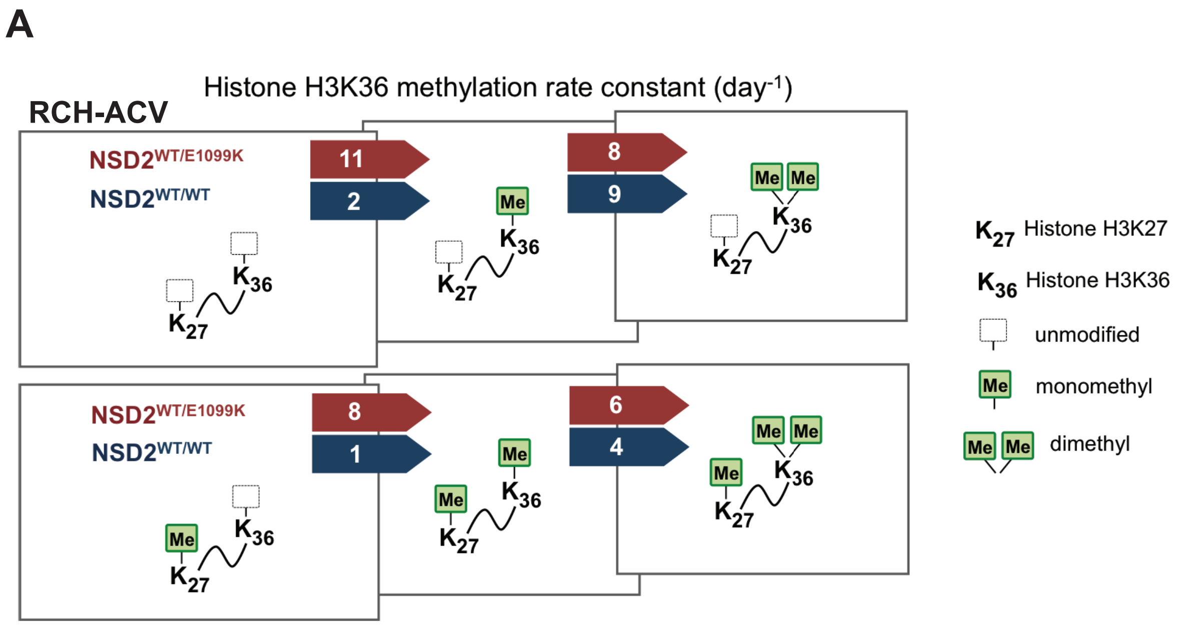


Figure 2.2.2. Gene editing to correct the E1099K mutation restores the balance of H3K36me2 to H3K27me3

A, mass spectrometry quantifying histone marks in CRISPR-edited and unedited cell lines (mean +SD, n=2 each). **B**, immunoblot for EZH2 and HDAC2 (loading control) was performed on nuclear lysates from parental RCH-ACV, SEM, and RPMI-8402 cells and their subclones. *Data generated by Xiaoxiao Huang (A) and Christine Will (B).*


A
Histone H3K36 methylation rate constant (day-1)
RCH-ACV
NSD2WT/E1099K
NSD2WT/WT
11
2
8
9
Me
Me
Me
K36
K27
NSD2WT/E1099K
NSD2WT/WT
8
1
6
4
Me
Me
Me
Me
Me
K36
K27
K27 Histone H3K27
K36 Histone H3K36
unmodified
Me monomethyl
Me Me dimethyl

B

K36 rate constant in RCHACV WT/E1099K						
day $^{-1}$	methylation			demethylation		
	me0->1	me1->2	me2->3	me1->0	me2->1	me3->2
K27me0	11 (0.7)	8 (0.2)	0 (0)	1 (1.3)	1 (0.1)	0 (0.1)
K27me1	8 (0.7)	6 (0.5)	0 (0)	6 (0.9)	0 (0)	0 (0.1)
K27me2	4 (0.9)	7 (1.2)	5 (3.9)	3 (0.7)	2 (2.1)	1 (0.6)
K27me3	1 (0.4)	0 (0.4)	ND	1 (0.7)	0 (0.5)	ND

K27 rate constant in RCHACV WT/E1099K						
day $^{-1}$	methylation			demethylation		
	me0->1	me1->2	me2->3	me1->0	me2->1	me3->2
K36me0	0 (0.3)	0 (0.6)	0 (0.3)	3 (0.9)	0 (0.4)	1 (1)
K36me1	10 (0.6)	7 (0.7)	3 (0.7)	0 (0.5)	0 (0.3)	0 (0.6)
K36me2	1 (0)	0 (0)	4 (2.7)	0 (0)	0 (0.9)	2 (0.9)
K36me3	1 (0)	0 (0.1)	ND	1 (0.1)	2 (0.3)	ND

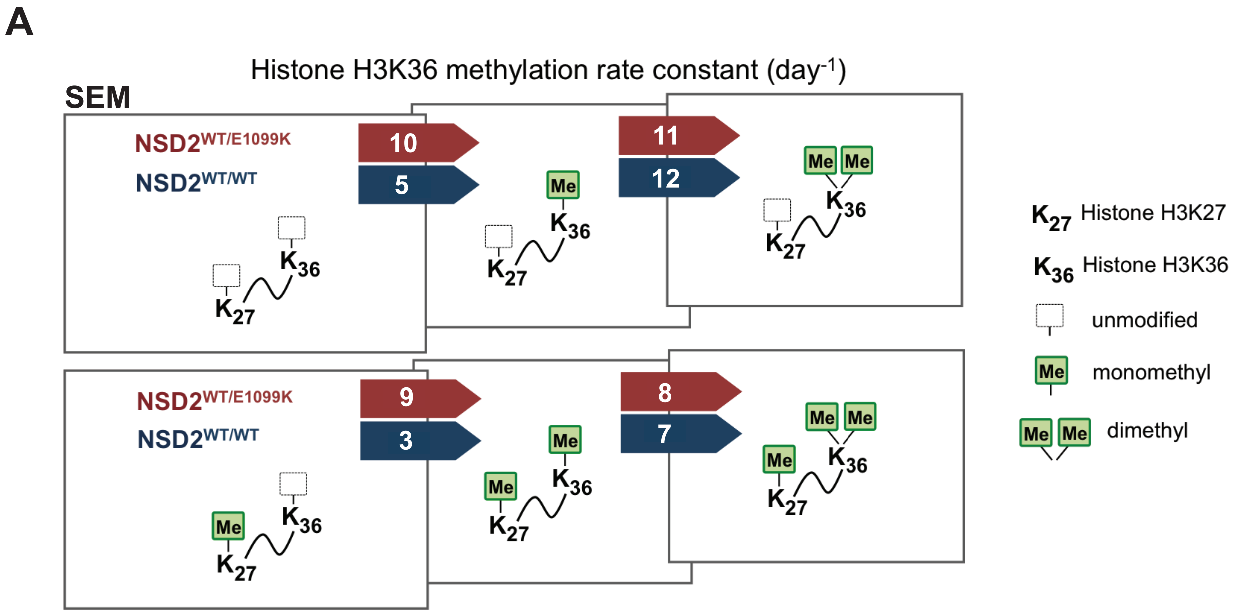
K36 rate constant in RCHACV WT/WT						
day $^{-1}$	methylation			demethylation		
	me0->1	me1->2	me2->3	me1->0	me2->1	me3->2
K27me0	2 (0.1)	9 (0.1)	0 (0)	0 (0.7)	1 (0.1)	0 (0.1)
K27me1	1 (0.1)	4 (0.1)	0 (0)	4 (0.4)	0 (0)	0 (0)
K27me2	0 (0)	0 (0.1)	6 (14.5)	0 (0.1)	4 (10)	1 (2.3)
K27me3	0 (0.9)	1 (1.7)	ND	1 (6)	2 (2)	ND

K27 rate constant in RCHACV WT/WT						
day $^{-1}$	methylation			demethylation		
	me0->1	me1->2	me2->3	me1->0	me2->1	me3->2
K36me0	1 (0)	1 (0)	0 (0)	1 (0.2)	0 (0)	0 (1)
K36me1	5 (0.2)	1 (0.1)	0 (0.1)	0 (0.3)	0 (0)	0 (0.5)
K36me2	1 (0)	0 (0)	6 (16)	0 (0)	0 (0.8)	2 (3.7)
K36me3	0 (0.1)	0 (0.1)	ND	0 (0.1)	4 (3.8)	ND

Figure 2.2.3. M4K in RCH-ACV cells reveals increased rate of monomethylation by mutant NSD2

A, diagram representing the comparison of histone H3K36 methylation *in vivo* effective rate constants between NSD2$^{WT/E1099K}$ and NSD2$^{WT/WT}$ RCH-ACV cells when neighboring K27 is unmodified (top) or monomethylated (bottom) as determined by M4K.

B, rate constants of methylation and demethylation on histones with and without preexisting H3K27 and H3K36 modifications in RCH-ACV, NSD2$^{WT/E1099K}$ and NSD2$^{WT/WT}$ cells determined by M4K. Out of 5000 modeling results, values within 10% of a minimum RMSD were averaged as the output rate constant. Numbers inside the parentheses are modeling errors (ND=not detected). *Data and figure generated by Xiaoxiao Huang.*


A
Histone H3K36 methylation rate constant (day-1)
SEM
NSD2WT/E1099K
NSD2WT/WT
10
5
11
12
9
3
8
7
Me
K36
K27
K27 Histone H3K27
K36 Histone H3K36
unmodified
monomethyl
dimethyl

B

K36 rate constant in SEM WT/E1099K

day $^{-1}$	methylation			demethylation		
	me0->1	me1->2	me2->3	me1->0	me2->1	me3->2
K27me0	10 (0.4)	11 (0.3)	1 (0)	0 (0.8)	1 (0.1)	1 (0.3)
K27me1	9 (0.3)	8 (0.2)	0 (0)	11 (0.4)	1 (0)	0 (0.1)
K27me2	1 (0)	1 (0.1)	0 (0.1)	0 (0.1)	0 (0)	0 (0.2)
K27me3	1 (0.1)	1 (0.2)	ND	0 (0.2)	0 (0.2)	ND

K27 rate constant in SEM WT/E1099K

day $^{-1}$	methylation			demethylation		
	me0->1	me1->2	me2->3	me1->0	me2->1	me3->2
K36me0	3 (0.6)	3 (0.1)	1 (0.1)	6 (0.5)	0 (0.1)	0 (0)
K36me1	11 (0.6)	4 (0.5)	0 (0.1)	0 (0.3)	0 (0.1)	0 (0)
K36me2	1 (0)	0 (0)	0 (0.1)	0 (0.1)	0 (0)	0 (0.1)
K36me3	1 (0)	0 (0)	ND	1 (0.1)	0 (0.1)	ND

K36 rate constant in SEM WT/-

day $^{-1}$	methylation			demethylation		
	me0->1	me1->2	me2->3	me1->0	me2->1	me3->2
K27me0	5 (0.7)	12 (0.3)	1 (0)	9 (0.9)	2 (0.3)	0 (0)
K27me1	3 (0.4)	7 (0.4)	0 (0)	0 (0.3)	0 (0.1)	1 (0)
K27me2	0 (0)	0 (0.1)	1 (0.2)	0 (0)	0 (0)	0 (0.1)
K27me3	0 (0)	0 (0.1)	ND	0 (0)	0 (0.6)	ND

K27 rate constant in SEM WT/-

day $^{-1}$	methylation			demethylation		
	me0->1	me1->2	me2->3	me1->0	me2->1	me3->2
K36me0	13 (0.3)	5 (0)	0 (0)	5 (0.4)	0 (0)	0 (0)
K36me1	5 (0.8)	3 (0.3)	0 (0.1)	1 (1.8)	0 (0.1)	0 (0.1)
K36me2	1 (0)	0 (0)	1 (1.1)	1 (0.1)	0 (0)	1 (1.8)
K36me3	1 (0)	0 (0)	ND	0 (0)	0 (0.1)	ND

Figure 2.2.4. M4K in SEM cells reveals increased rate of monomethylation by mutant NSD2

A, diagram representing the comparison of histone H3K36 methylation *in vivo* effective rate constants between NSD2$^{WT/E1099K}$ and NSD2$^{WT/WT}$ RCH-ACV cells when neighboring K27 is unmodified (top) or monomethylated (bottom) as determined by M4K.

B, rate constants of methylation and demethylation on histones with and without preexisting H3K27 and H3K36 modifications in SEM, NSD2$^{WT/E1099K}$ and NSD2$^{WT/-}$ cells determined by M4K. Out of 5000 modeling results, values within 10% of a minimum RMSD were averaged as the output rate constant. Numbers inside the parentheses are modeling errors (ND=not detected). *Data and figure generated by Xiaoxiao Huang.*

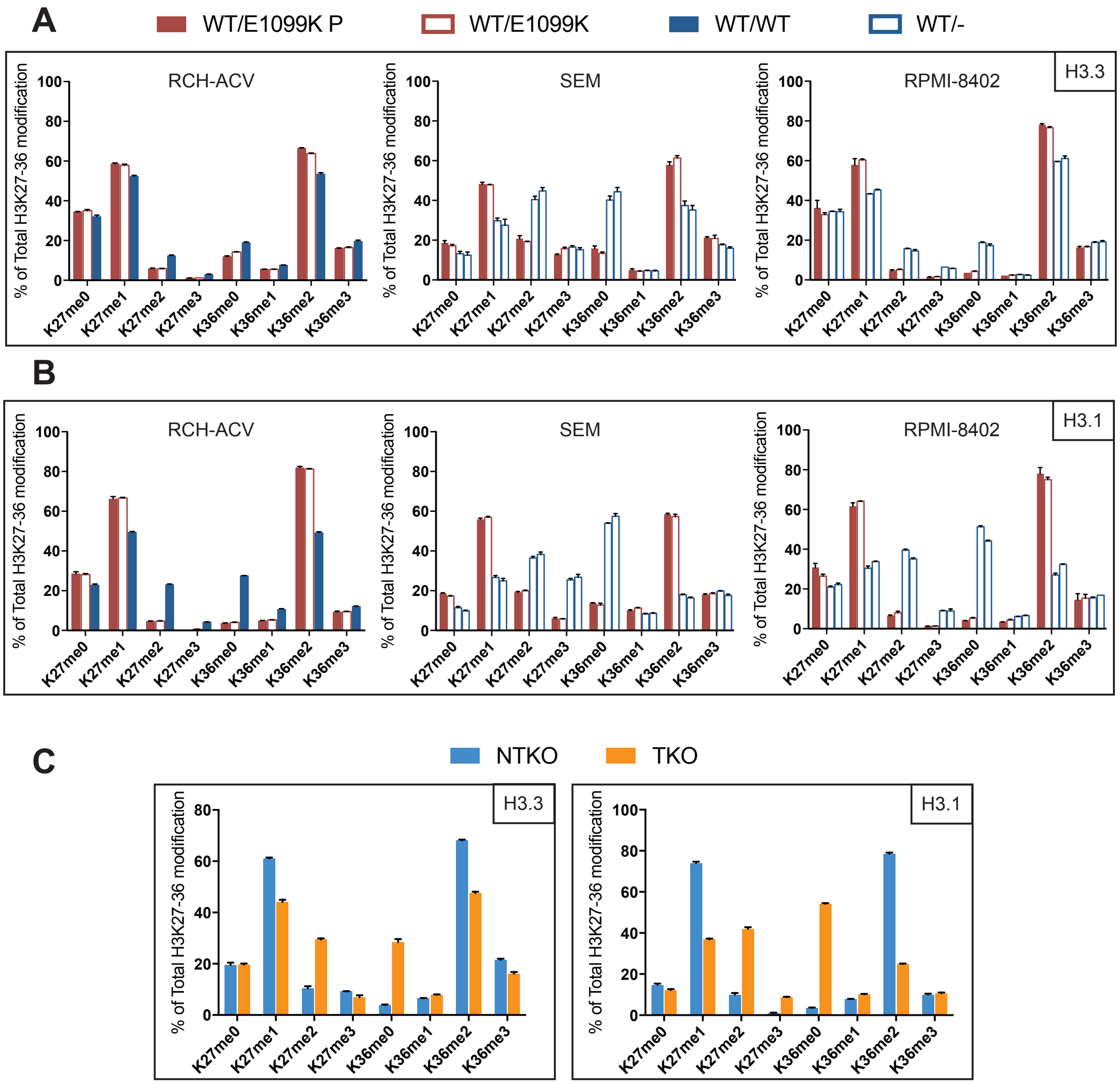


Figure 2.2.5. Hyperactive NSD2 preferentially increases H3K36 dimethylation of histone H3.3

Mass spectrometry quantifying the histone marks on **A,** H3.3 and **B,** H3.1 in RCH-ACV, SEM, and RPMI-8402 derived cells. **C,** quantification of histone marks on H3.3 and H3.1 in KMS11 derived TKO (NSD2-low) and NTKO (NSD2-high) cells. *Data generated by Xiaoxiao Huang.*

2.3. E1099K mutation promotes cancer-associated phenotypes

Next, we used the edited cell lines to explore the role of the E1099K mutation in ALL biology. Removal of the E1099K mutation reduced the proliferation of RCH-ACV, SEM, and RPMI-8402 cells. After 8 days in culture, RCH-ACV NSD2$^{WT/E1099K}$ cells were 8.5 times more numerous than cells without the mutation, while there were 1.4 times as many NSD2$^{WT/E1099K}$ SEM and 1.7 times as many RPMI-8402 cells than wildtype (Figure 2.3.1A). Similarly, there were 9-, 2-, and 3-fold more colonies formed by RCH-ACV, SEM, and RPMI-8402 NSD2$^{WT/E1099K}$ cells as compared to cells without the mutation (Figure 2.3.1B). Correlating with the decreased growth in culture, RCH-ACV and SEM cells in which NSD2^{E1099K} was removed demonstrated a modest increase of cells in G1, fewer cells in S and G2/M, and an increase in spontaneous apoptosis demonstrated by annexin V staining and a sub-G1 population (Figure 2.3.2). The decreased growth of NSD2$^{WT/-}$ RPMI-8402 cells was attributable to a marked, 2-fold increase in apoptosis (Figure 2.3.2A, D). Strikingly, the ability of the leukemia cells to adhere to stromal cells was significantly reduced upon removal of the E1099K mutation (Figure 2.3.3A), and RCH-ACV and SEM NSD2$^{WT/E1099K}$ cells migrated through Boyden chambers twice as effectively as cells without the mutation (Figure 2.3.3B). Thus, removal of the E1099K mutation resulted in the loss of many phenotypes associated with aggressive malignancy and disease progression.

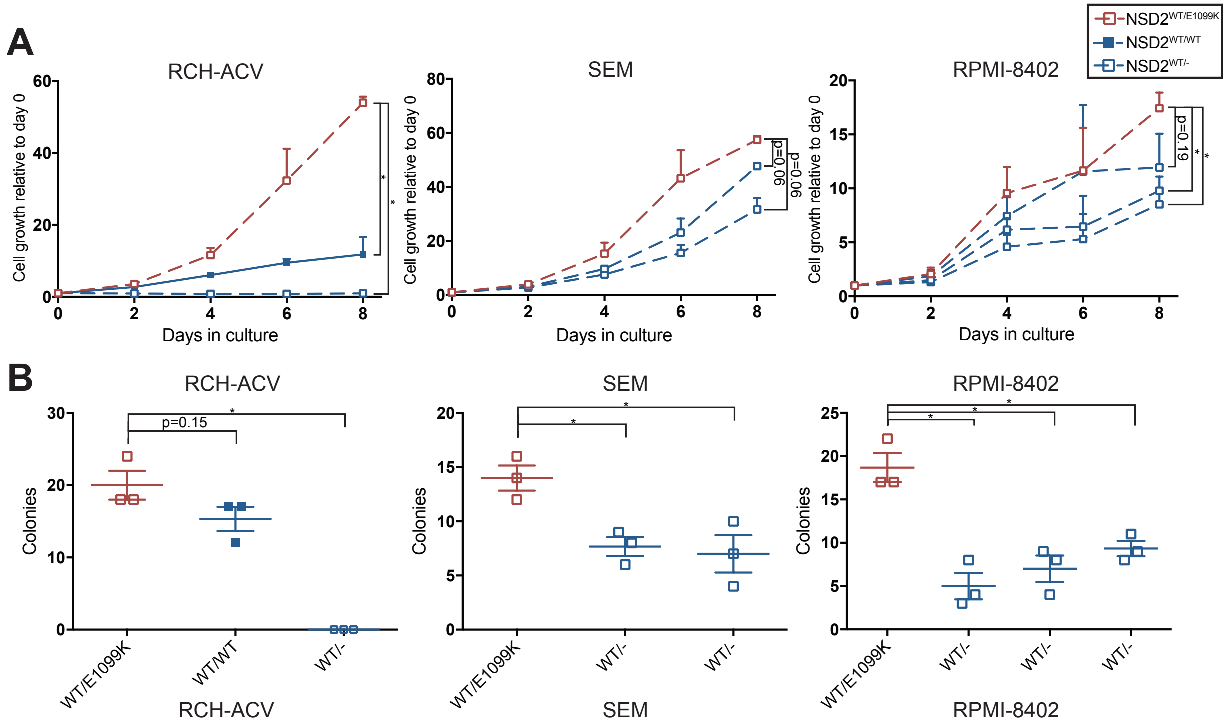


Figure 2.3.1. E1099K mutation promotes proliferation and clonogenicity

A, proliferation of RCH-ACV, SEM, and RPMI-8402 cells in liquid culture was monitored by Celltiter-Glo Luminescent Cell Viability Assay every two days (n=2-3; mean +SEM). Day 8 data analyzed by two-tailed t-test with equal variance. **B,** colonies formed in methylcellulose were counted 7 days after plating (n=3). Data are presented as mean ±SEM and analyzed by two sample t-test with unequal variance; * $p<0.05$. *Data generated with Christine Will.*

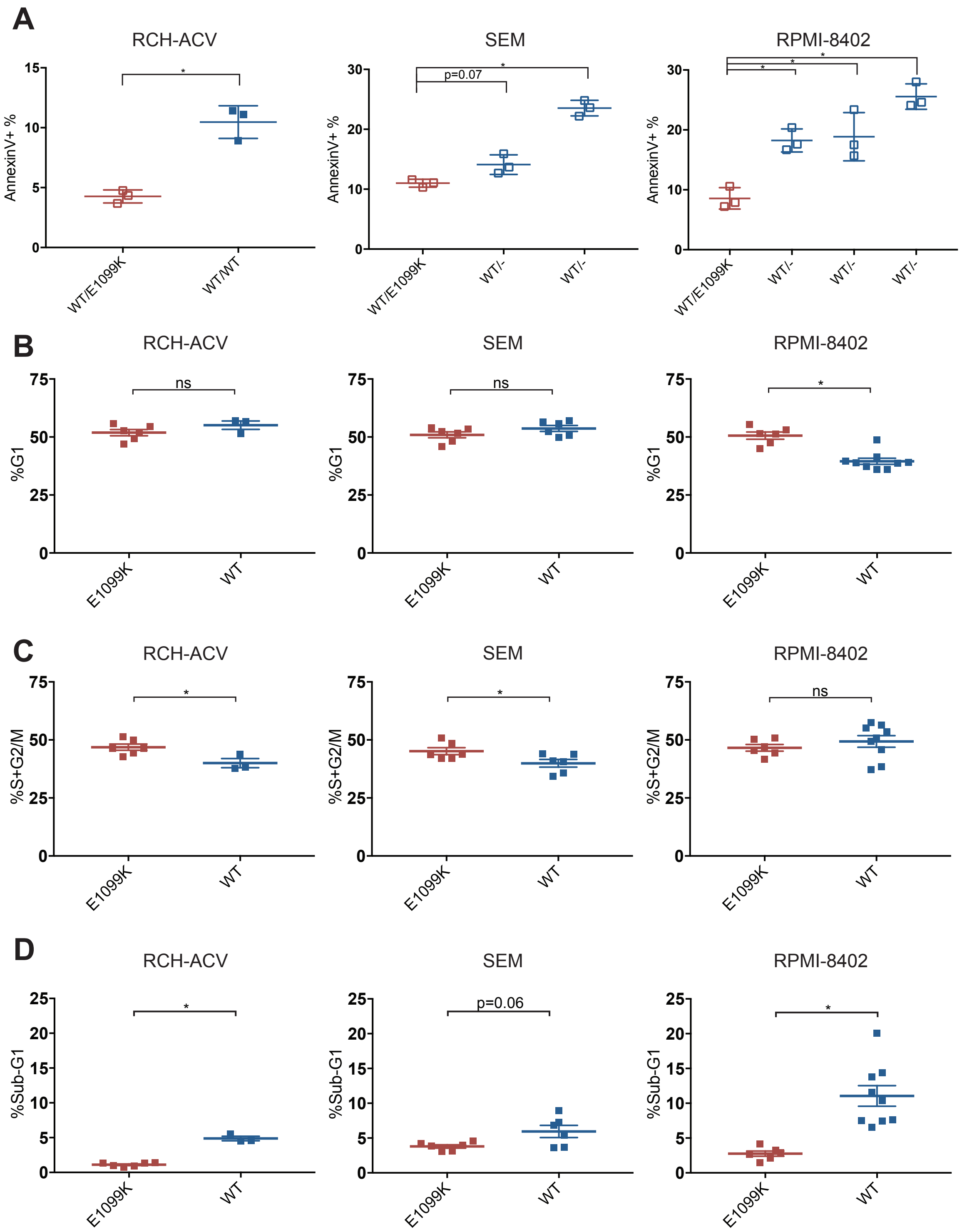


Figure 2.3.2. E1099K mutation reduces apoptosis and alters cell cycle progression

A, cells were stained for Annexin V using a FITC conjugated antibody and then quantified by flow cytometry (n=3, mean ±SD).

Cells were fixed and stained with propidium iodide, and cells in **B,** G1 **C,** S+G2/M, or **D,** sub-G1 phase were quantified by flow cytometry. For this experiment, cells harboring NSD2^{E1099K} were compared to those without the mutation. Each datapoint represents a biological replicate overlaid with mean ±SD. Data were analyzed by two sample t-test with unequal variance; * $p<0.05$. ns = nonsignificant.

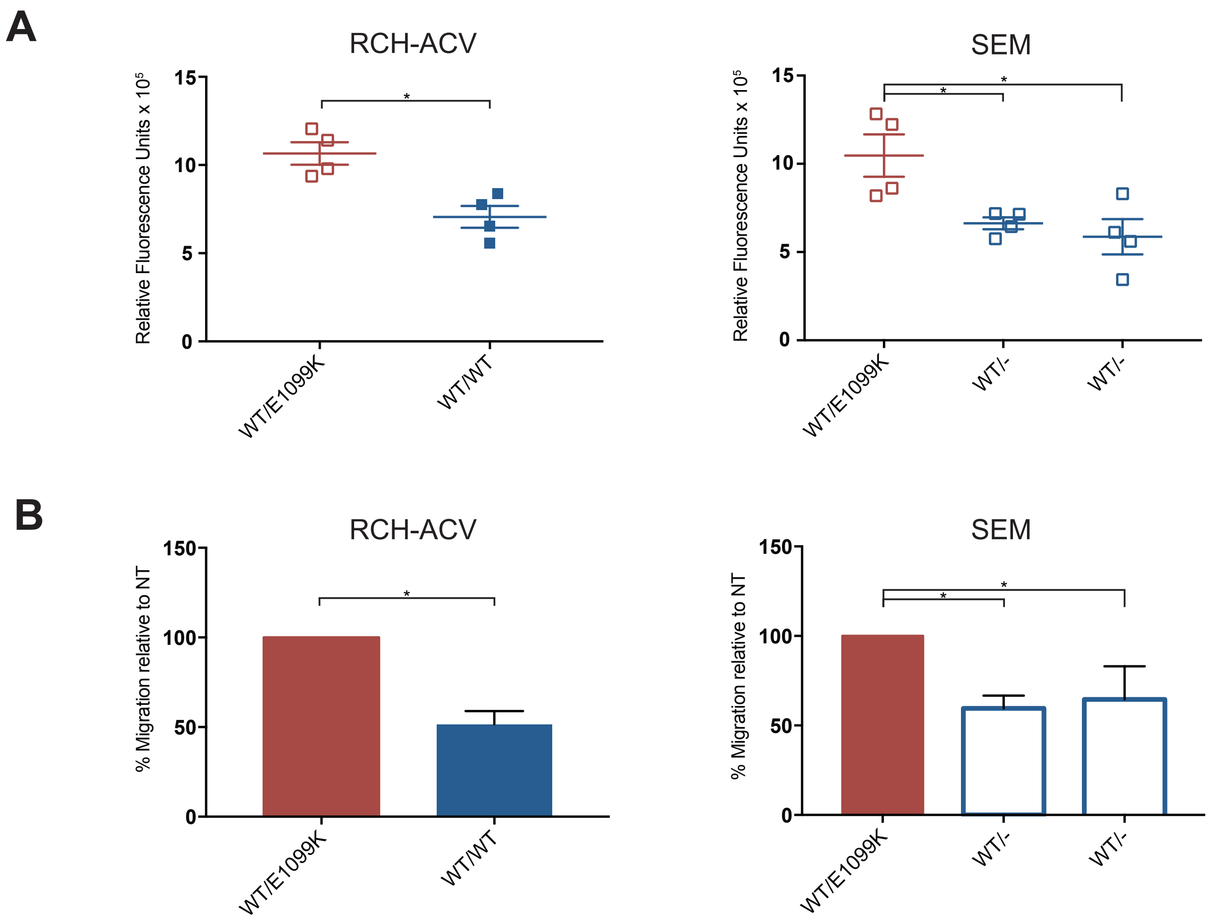


Figure 2.3.3. E1099K mutation increases adhesion and migration of ALL cells

A, cells were fluorescently labeled and incubated on a layer of HS-5 bone marrow stromal cells for 24 hours. Following washes, adherent cell numbers were estimated by relative fluorescence (n=4). **B,** cells were seeded in top of Boyden chamber containing no serum and allowed to migrate for 24 hours to serum-containing bottom chamber. Cells in bottom chamber were quantified by flow cytometry and normalized to non-targeted, NSD2$^{WT/E1099K}$ (RCH-ACV, n=3; SEM, n=4). Data are presented as mean ±SEM and analyzed by two sample t-test with unequal variance; * $p<0.05$.

2.4. NSD2-mutant cells behave aggressively in vivo

To examine the role of the E1099K mutation on tumor growth *in vivo*, luciferase-tagged mutant and CRISPR-edited ALL cell lines were injected by tail vein into NOD-SCID mice and monitored by *in vivo* imaging. Mice xenografted with RCH-ACV NSD2$^{WT/E1099K}$ cells had a median survival of 22 days, requiring sacrifice due to neurological impairment, while mice injected with NSD2$^{WT/WT}$ cells had a median survival of 27.5 days (Figure 2.4.1A, left panel). Similarly, mice injected with SEM NSD2$^{WT/E1099K}$ cells survived for 30 days compared to mice injected with two different clones of NSD2$^{WT/-}$ cells that had median survivals of 50 days and 59 days (Figure 2.4.1A, right panel). The total tumor burden detected by photon flux in mice injected with NSD2$^{WT/WT}$ or NSD2$^{WT/-}$ cells was 3-fold higher at time of death than those injected with NSD2$^{WT/E1099K}$ cells (Figure 2.4.1B), indicating that mice were able to tolerate wildtype cells better than mutant. Furthermore, while all cells were widely disseminated late in the disease course, NSD2$^{WT/E1099K}$ cells were detected in the head earlier and with more intensity than NSD2$^{WT/WT}$ or NSD2$^{WT/-}$ cells (Figure 2.4.2A-B), suggesting altered migration and response to brain microenvironment. Histological analysis of brain sections indicated leptomeningeal involvement of NSD2$^{WT/E1099K}$ and NSD2$^{WT/-}$ SEM cells, but only NSD2$^{WT/E1099K}$ cells invaded the brain parenchyma (Figure 2.4.2C). These observations indicate that the E1099K mutation promotes cancer phenotypes *in vitro* and dictates aggressive biological behavior *in vivo*.

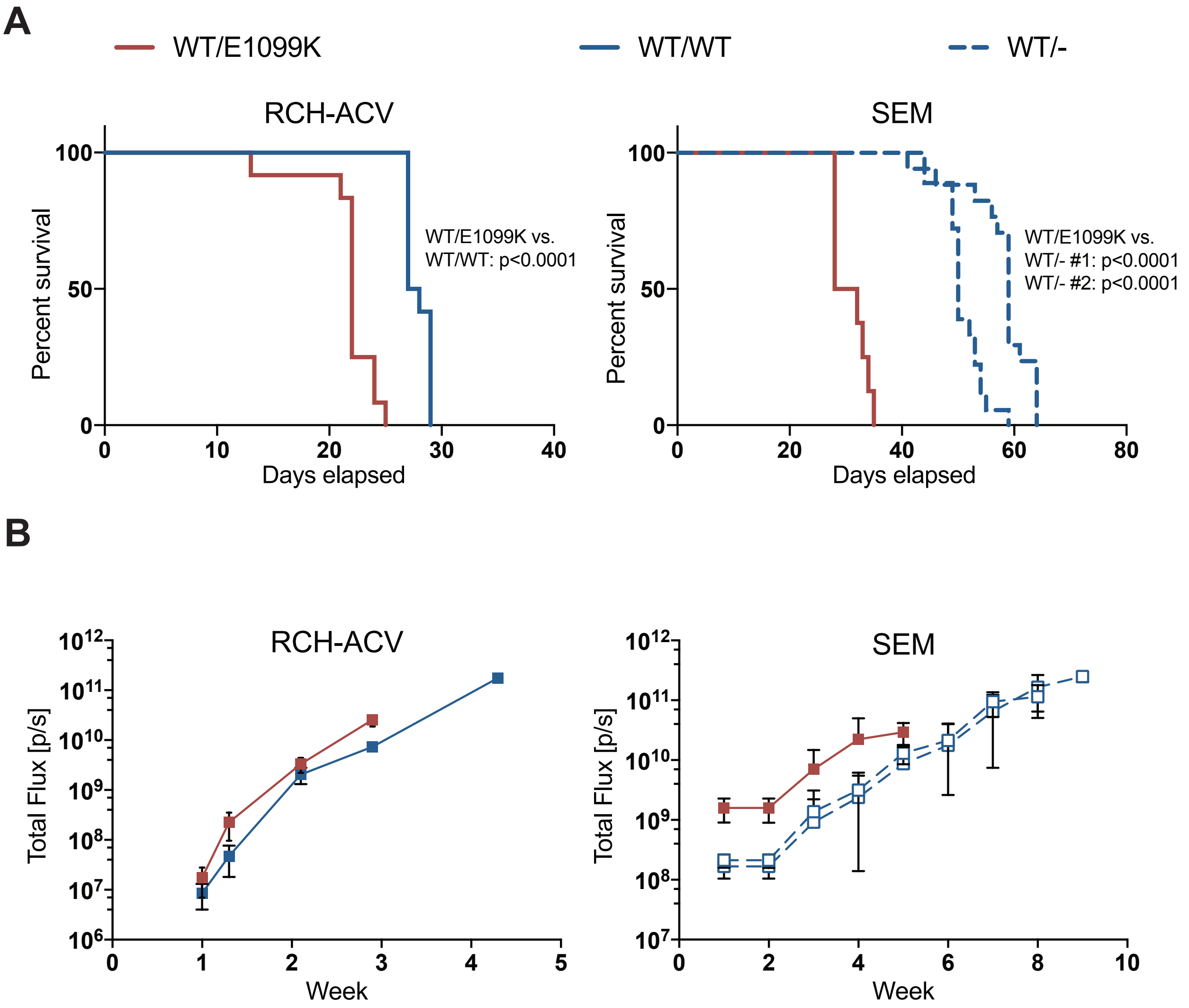


Figure 2.4.1. NSD2-mutant cells exhibit increased lethality in mice

A, survival curves following injection of NOD-SCID mice with 5x10^{6} RCH-ACV or SEM CRISPR-edited cells. Log-rank test was used for statistical analysis to compare NSD2$^{WT/E1099K}$ to NSD2$^{WT/WT}$ or NSD$^{WT/-}$ cells (RCH-ACV, n=24; SEM, n=51; $p<0.0001$). **B,** the total bioluminescence signal (radiance, photons/second) was quantified weekly after xenograft. Y-axis is logarithmic to highlight differences in data.

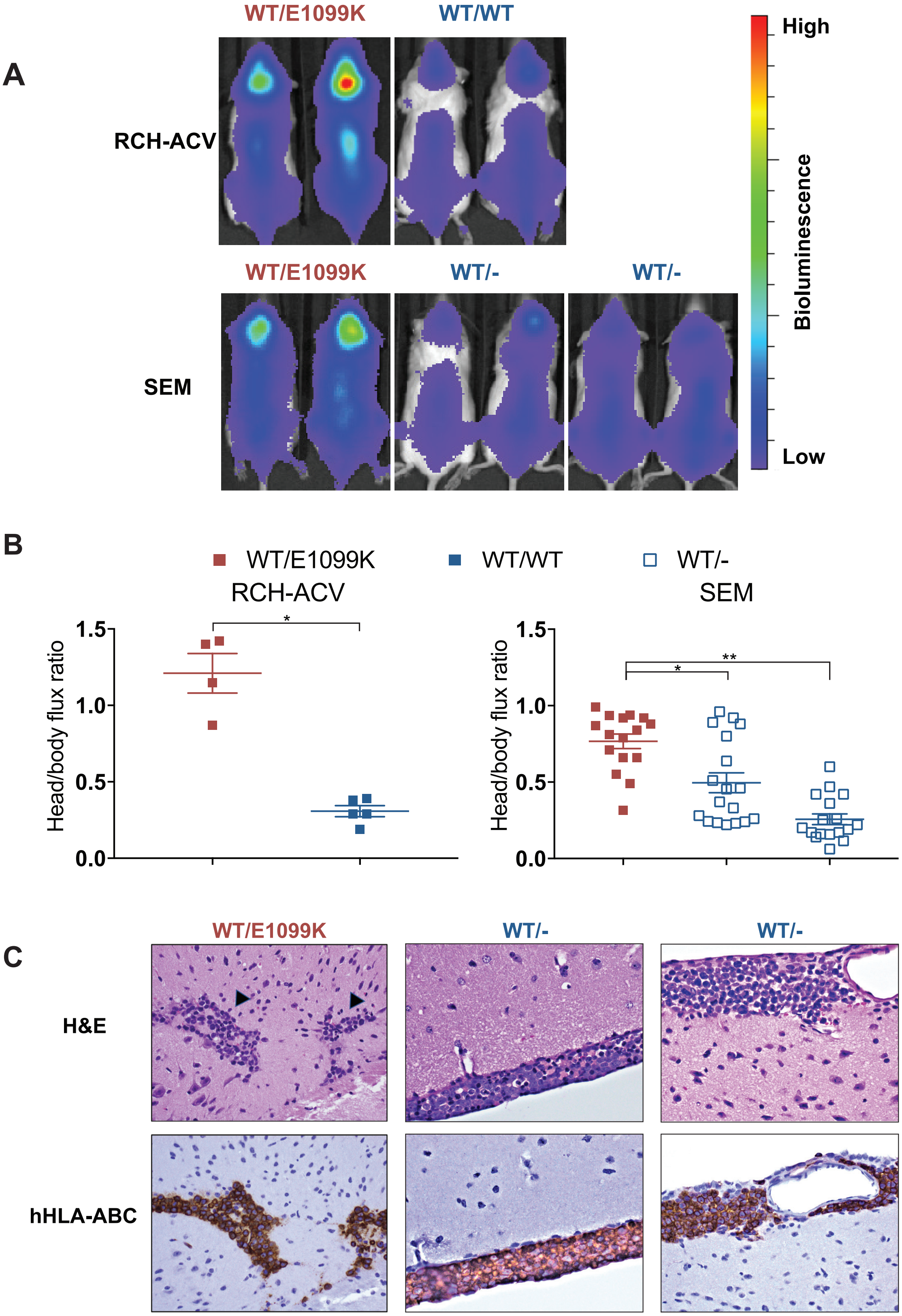

A
WT/E1099K
WT/WT
RCH-ACV
WT/E1099K
WT/-
WT/-
SEM
High
Bioluminescence
Low
B
WT/E1099K
WT/WT
WT/-
RCH-ACV
SEM
Head/body flux ratio
1.5
1.0
0.5
0.0
*
**
C
WT/E1099K
WT/-
WT/-
H&E
hHLA-ABC

Figure 2.4.2. NSD2$^{WT/E1099K}$ cells demonstrate altered migration *in vivo*

A, representative bioluminescence images from 3 weeks (RCH-ACV) or 4 weeks (SEM) post-injection. **B,** the ratio of bioluminescence signal between head and body was quantified at these time points for RCH-ACV- (n=9) and SEM-injected (n=41) mice. * $p<0.01$, ** $p<0.0001$; two sample t-test with unequal variance. **C**, representative brain sections from mice injected with SEM cells showing CNS infiltration by blasts with surface and cytoplasmic immunoreactivity for hHLA-ABC immunohistochemistry **(400x magnification)**. Black arrowheads indicate parenchymal infiltration.

2.5. Widespread gene activation in E1099K cells drives oncogenic programs

In order to characterize the pathways that drive aggressive phenotypes associated with the E1099K mutation, we performed global transcriptome analysis of CRISPR-edited cell lines. Most of the genes differentially expressed in NSD2$^{WT/E1099K}$ cells relative to wildtype cells were upregulated (Figure 2.5.1). Accordingly, there were 84 genes commonly over-expressed in NSD2$^{WT/E1099K}$ RCH-ACV, SEM, and RPMI-8402 cells, but only one gene commonly repressed (Figure 2.5.2), suggesting that the primary function of the E1099K mutation in NSD2 is to activate transcription. Gene set enrichment analysis (GSEA) using Enrichr revealed that E1099K upregulated genes overlapped with those regulated by repressive polycomb proteins EZH2 and CBX8, and displaying the H3K27me3 modification in mouse spleen and thymus (Figure 2.5.3). In fact, the E1099K upregulated genes tended to be under-expressed in normal B and T cells and over-expressed in stromal cells (Figure 2.5.4; www.immgen.org), indicating that E1099K activated a gene program distinctly ectopic to the hematopoietic lineage (Heng et al., 2008). GSEA using the Kyoto encyclopedia of genes and genomes (KEGG) showed that genes upregulated in NSD2$^{E1099K/WT}$ cells were part of oncogenic pathways including cell adhesion, cytokine-receptor interaction, and signaling (Figure 2.5.5). The overlap between Broad molecular signatures database (MSigDB) canonical pathways differentially regulated in NSD2$^{WT/E1099K}$ cells is shown in Figure 2.5.6. The 10 gene sets differentially upregulated in all three NSD2$^{WT/E1099K}$ cell lines are:

NABA_MATRISOME

REACTOME_DEVELOPMENTAL_BIOLOGY

REACTOME_POTASSIUM_CHANNELS

NABA_CORE_MATRISOME

REACTOME_AXON_GUIDANCE

REACTOME_NEURONAL_SYSTEM

NABA_MATRISOME_ASSOCIATED

REACTOME_NETRIN1_SIGNALING

KEGG_O_GLYCAN_BIOSYNTHESIS

REACTOME_L1CAM_INTERACTIONS

Interestingly, the major effect of NSD2^{E1099K} appears to be altering genes related to adhesion and migration.

In order to assess how relevant the data from our CRISPR-edited cell lines are to ALL patients, we compared it to datasets of genes upregulated in patient samples with NSD2 mutations (Subramanian 2007). Analysis of microarray data from E2A-PBX1 ALL patients, some of whom displayed the E1009K mutation, revealed 261 genes upregulated in NSD2$^{WT/E1099K}$ samples (fold change>2, $p<0.05$) (Zhang 2012, Jaffe). These genes also tended to be activated in all three NSD2$^{WT/E1099K}$ cell lines (Figure 2.5.7, row 1). From this same study, we compared ETV6-RUNX1 ALL samples and identified 144 E1099K upregulated genes. The ETV6-RUNX1 gene-set was enriched in RCH-ACV and RPMI-8402 NSD2$^{WT/E1099K}$ cells (Figure 2.5.7, row 2). Intriguingly, the data generated from the CRISPR-edited cell lines may be applicable beyond ALL. Genes differentially upregulated in mutant NSD2 MCL or t(4;14) MM were also enriched in NSD2$^{WT/E1099K}$ cell lines (Figure 2.5.7, row 3-4)(Beà et al., 2013; Broyl et al., 2010). The strong correlation between these gene expression data and patient data indicate our cell line models can recapitulate human disease and that similar pathways may be functioning downstream of dysregulated NSD2 in various malignancies.

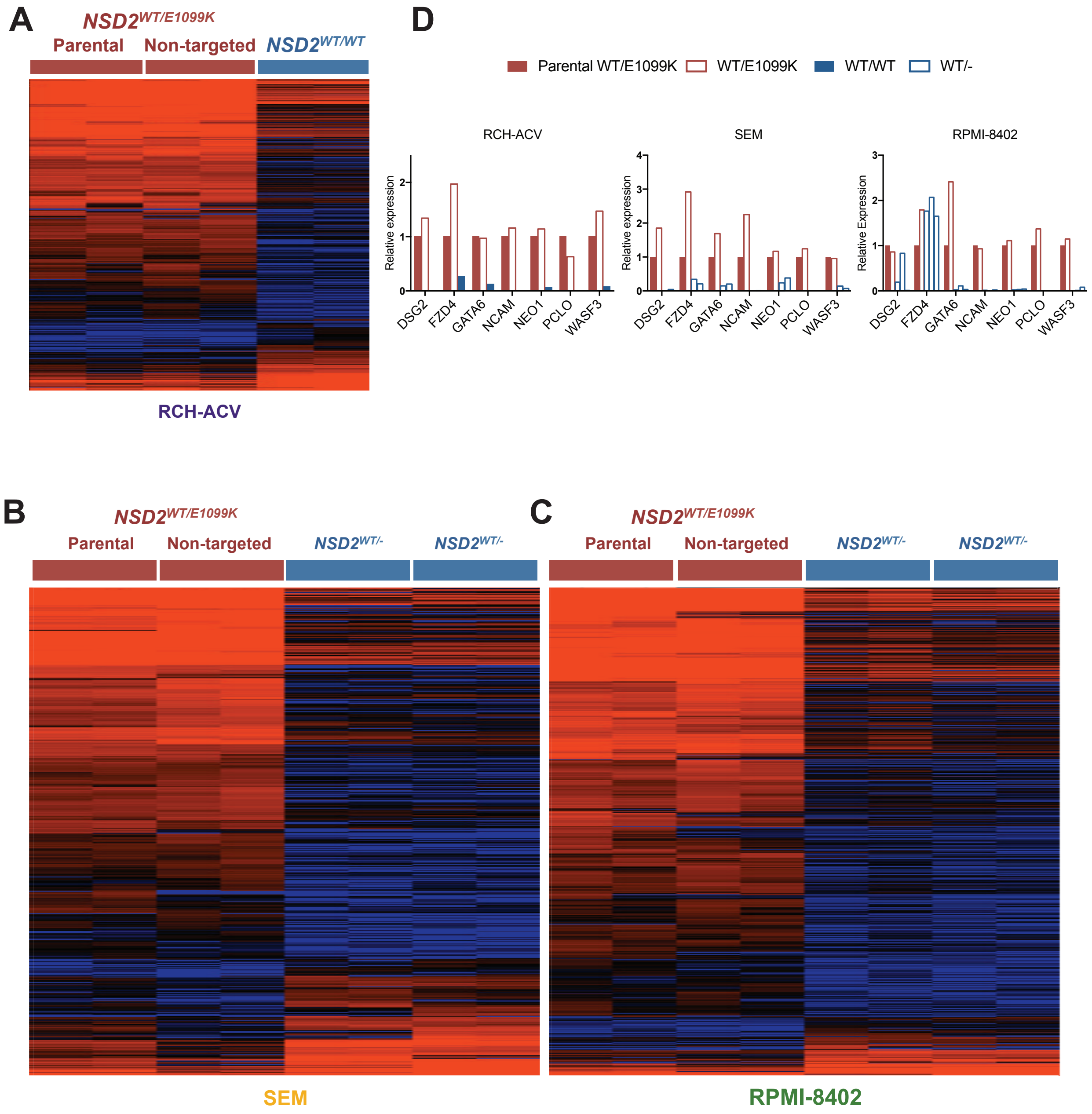
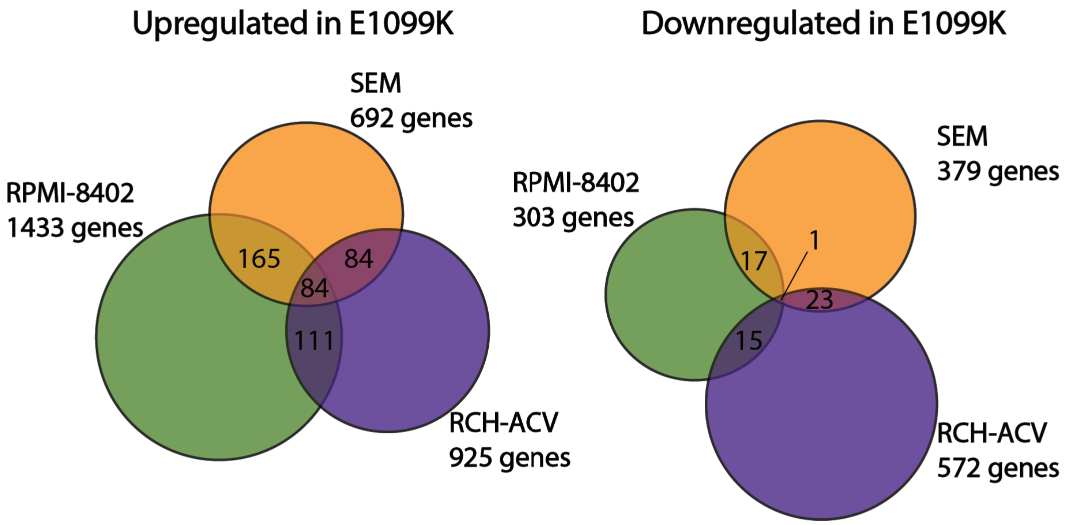


Figure 2.5.1. Widespread gene activation in E1099K cells

Heat map depicting differentially expressed genes between parental NSD2$^{WT/E1099K}$, non-targeted NSD2$^{WT/E1099K}$, and CRISPR-edited NSD2$^{WT/WT}$ or NSD2$^{WT/-}$ cell lines. **A,** RCH-ACV; **B,** SEM; **C,** RPMI-8402. Genes arranged by euclidean distance. **D,** Selected genes were tested by qPCR to validate RNA-seq results. *Data generated with Jon Oyer and Alberto Riva.*

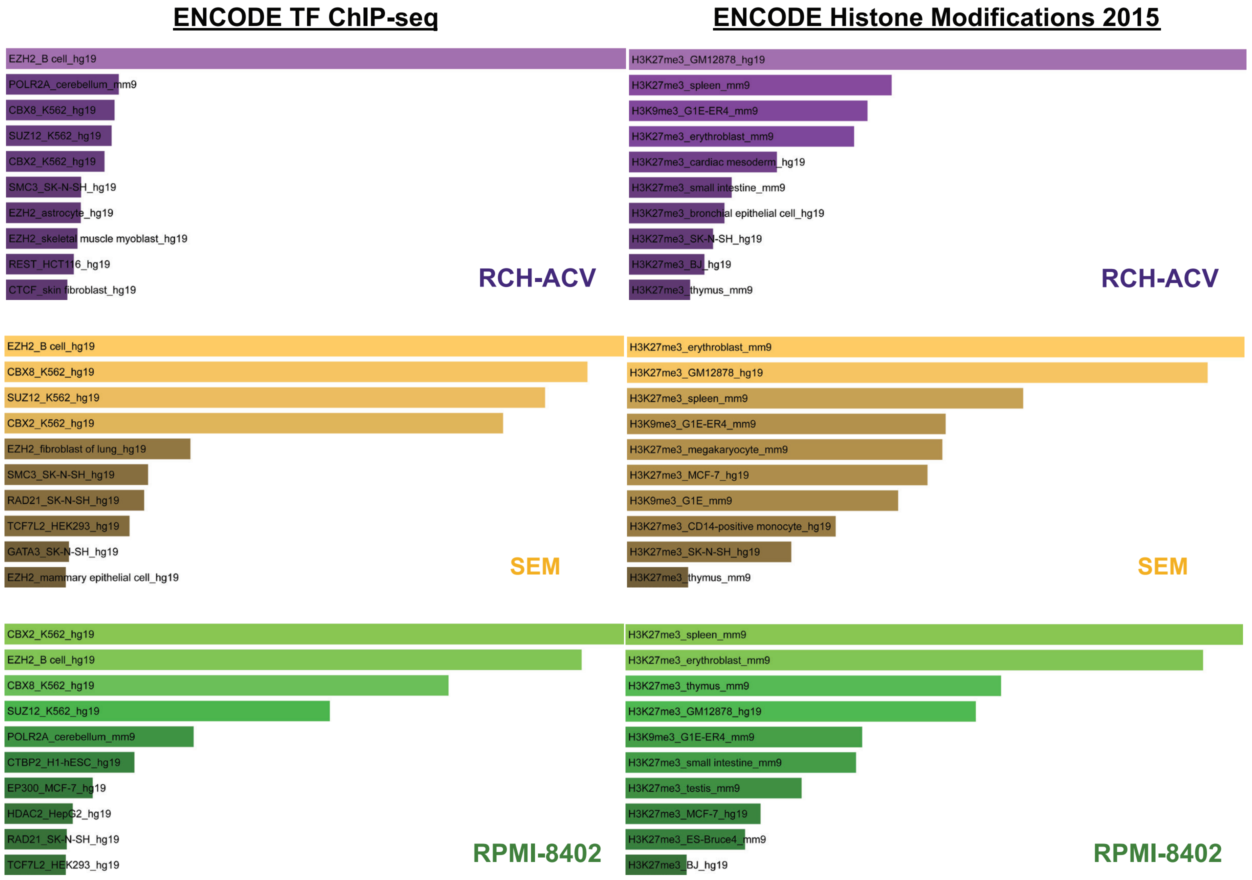


Intersection	Significance
SEM vs. RCH-ACV	0.001
SEM vs. RPMI-8402	0.001
RCH-ACV vs. RPMI-8402	0.001
ALL	0.001

Intersection	Significance
SEM vs. RCH-ACV	0.002
SEM vs. RPMI-8402	0.001
RCH-ACV vs. RPMI-8402	0.014
ALL	0.150

Figure 2.5.2. Genes upregulated by E1099K show strong overlap between cell lines

Venn diagram displays overlap between different cell lines in genes upregulated or downregulated in NSD2$^{WT/E1099K}$ versus NSD2$^{WT/WT}$ and/or NSD2$^{WT/-}$ cells (differentially regulated genes: fold-change>2, p<0.05).



ENCODE TF ChIP-seq

ENCODE Histone Modifications 2015

Figure 2.5.3. E1099K upregulated genes are generally marked for repression

Genes upregulated in $\text{NSD2}^{\text{WT/E1099K}}$ cells from RNA-seq analyzed for highly represented ENCODE transcription factor or histone modification gene sets by Enrichr. Length of bars corresponds to significance as measured by p and q combined score.

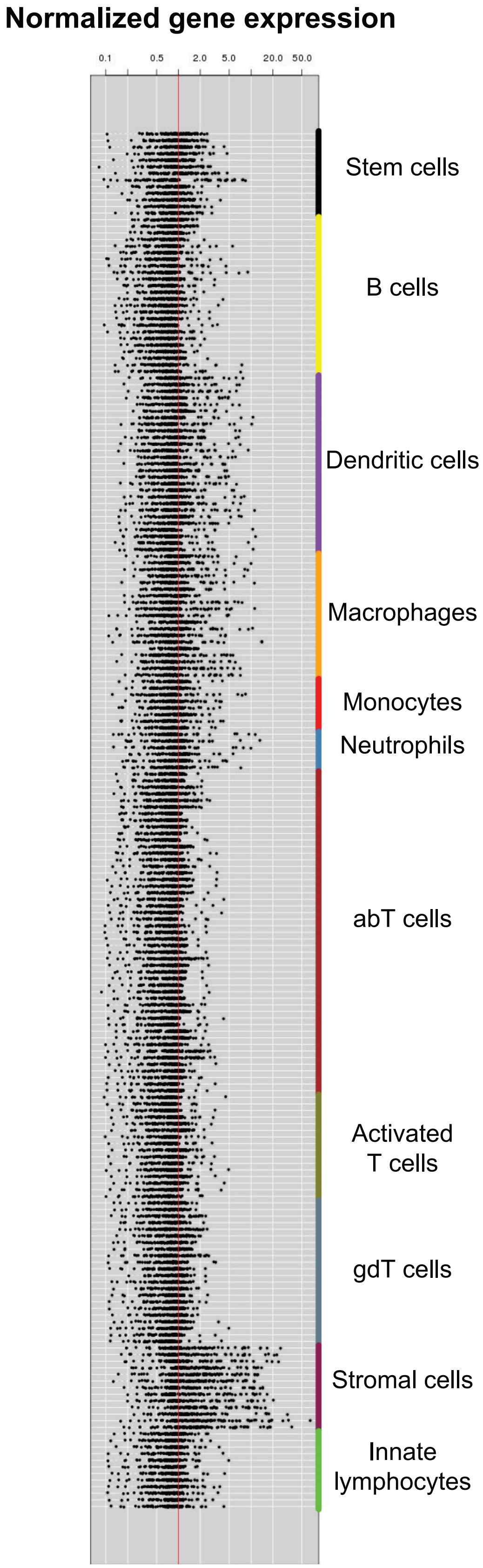


Figure 2.5.4. E1099K mutation activates genes commonly expressed in stromal

cell lineage

W plot (www.immgen.org) illustrates the relative expression of the 84 commonly upregulated genes between all NSD2$^{WT/E1099K}$ ALL cell lines in normal immune cells.

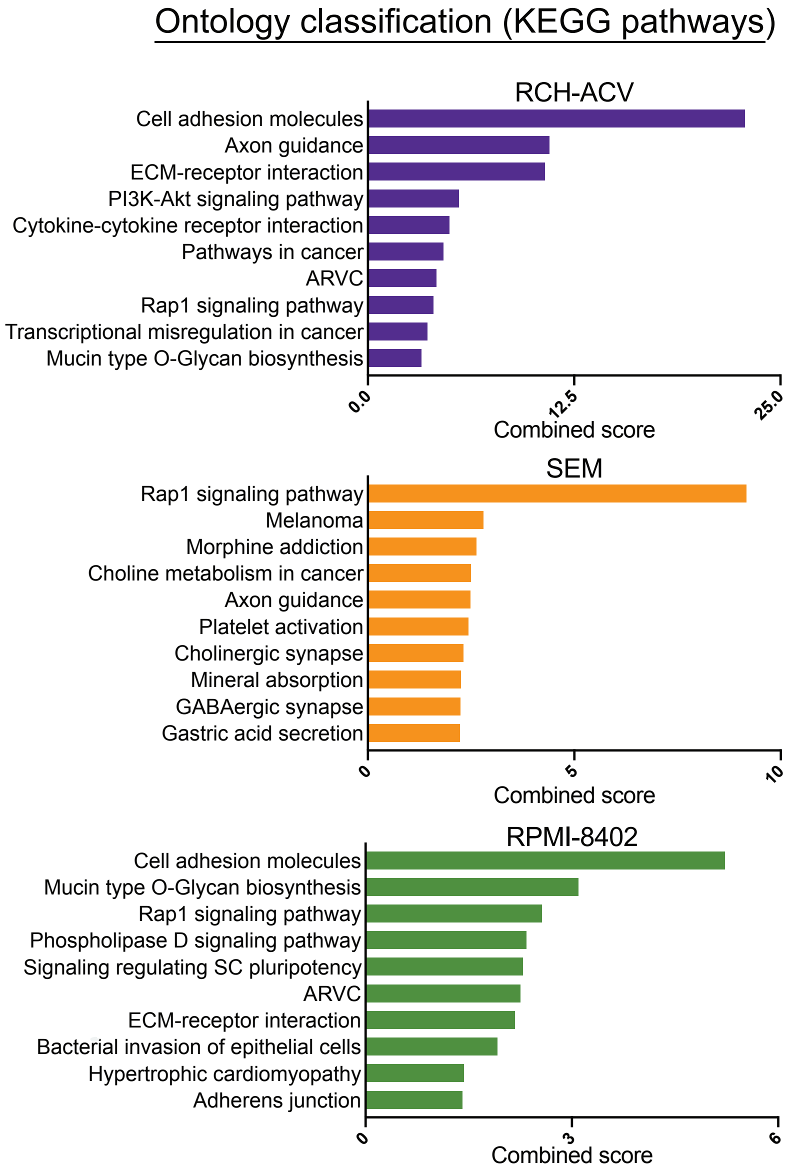
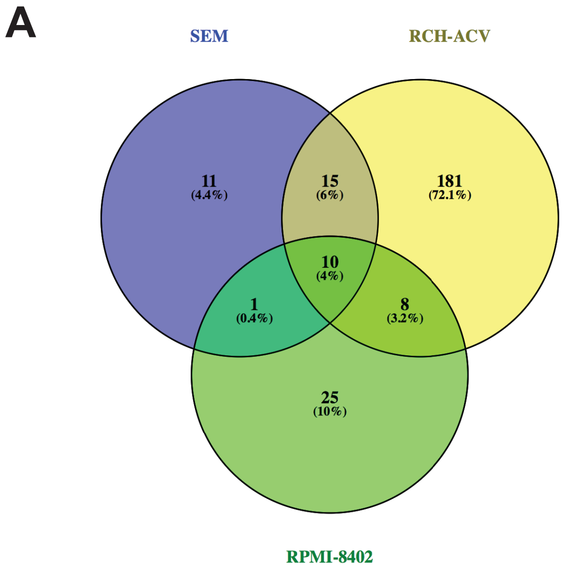


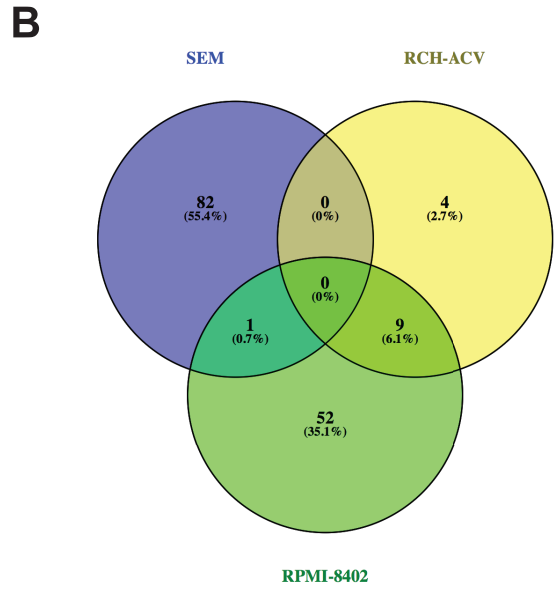
Figure 2.5.5. Mutant NSD2 upregulates genes involved in many cancer associated pathways

KEGG pathways represented in genes upregulated in NSD2$^{WT/E1099K}$ cells using gene ontology analysis (p and q combined score).



Upregulated gene sets

Intersection	Significance
SEM vs. RCH-ACV	0.001
SEM vs. RPMI-8402	0.001
RCH-ACV vs. RPMI-8402	0.001
ALL	0.001



Downregulated gene sets

Intersection	Significance
SEM vs. RCH-ACV	no overlap
SEM vs. RPMI-8402	0.985
RCH-ACV vs. RPMI-8402	0.001
ALL	no overlap

Figure 2.5.6. Pathways upregulated in E1099K cells are common between cell lines

Genes differentially regulated in NSD2$^{WT/E1099K}$ cells were analyzed by GSEA. The Venn diagrams display up- and downregulated gene sets in common between cell lines (differentially regulated MSigDB-Canonical Pathways, $p<0.05$).

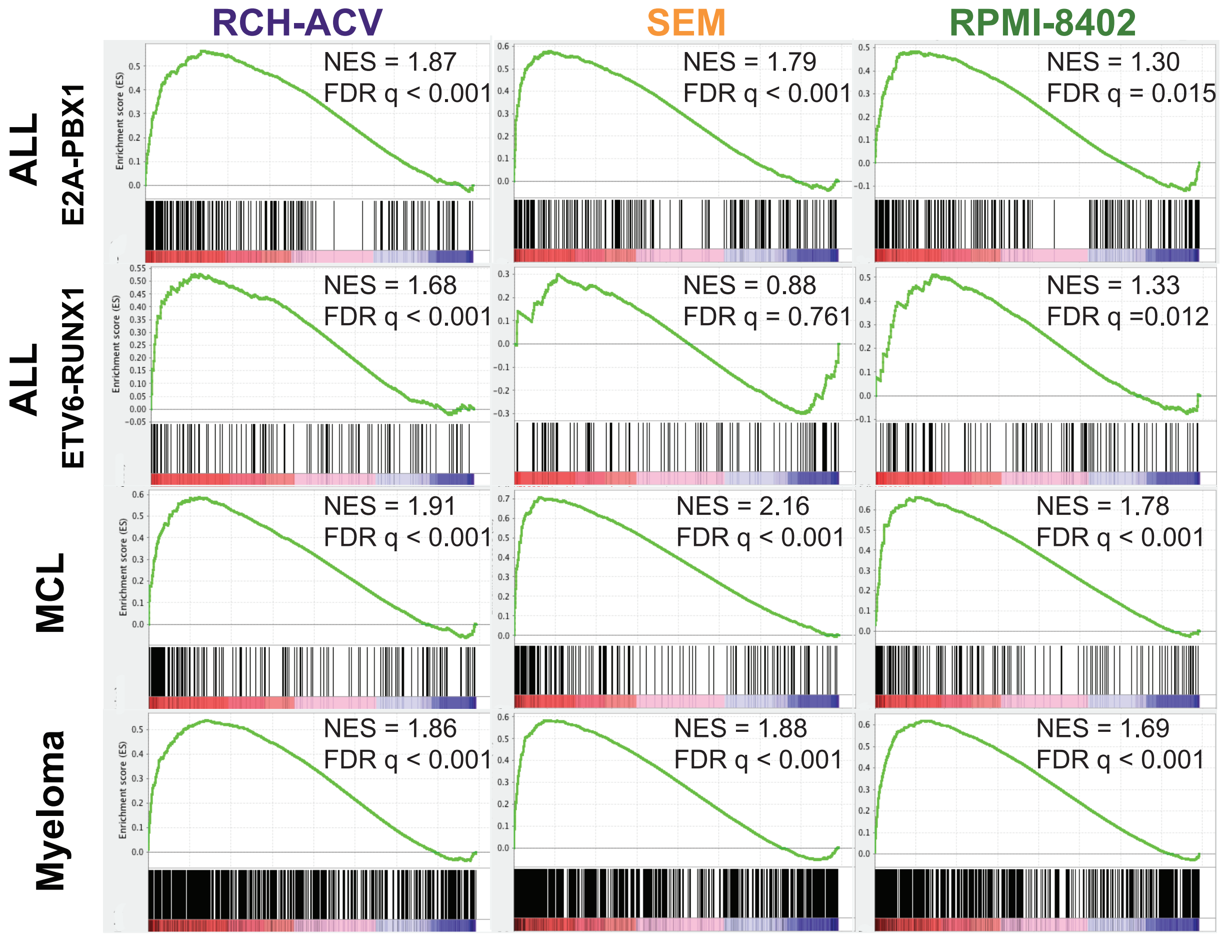


Figure 2.5.7. Gene expression data from RCH-ACV, SEM, and RPMI-8402 cell lines resembles patient data

Gene expression from NSD2 E1099K and wildtype RCH-ACV, SEM, and RPMI-8402 cell lines was analyzed by GSEA to compare it to genes upregulated in mutant NSD2 ALL, MCL, and MM patients.

2.6. E1099K cells are susceptible to inhibition of AKT

Analysis of transcriptional differences between wildtype and mutant NSD2 cells indicated that numerous signaling pathways, including PI3K/AKT, were activated by NSD2^{E1099K}. A reversed phase antibody microarray experiment using RCH-ACV cell lysates showed increased phosphorylation of several signaling proteins (Figure 2.6.1). Immunoblot revealed that AKT phosphorylation of serine 473 was higher in RCH-ACV NSD2$^{WT/E1099K}$ cells compared to wildtype, particularly following stimulation with insulin (Figure 2.6.2A). NSD2$^{WT/E1099K}$ RCH-ACV cells had increased sensitivity to AKT inhibitor, GSK690693, suggesting that interruption of PI3K/AKT pathways could be a viable therapy for NSD2-mutant ALL (Figure 2.6.2B).

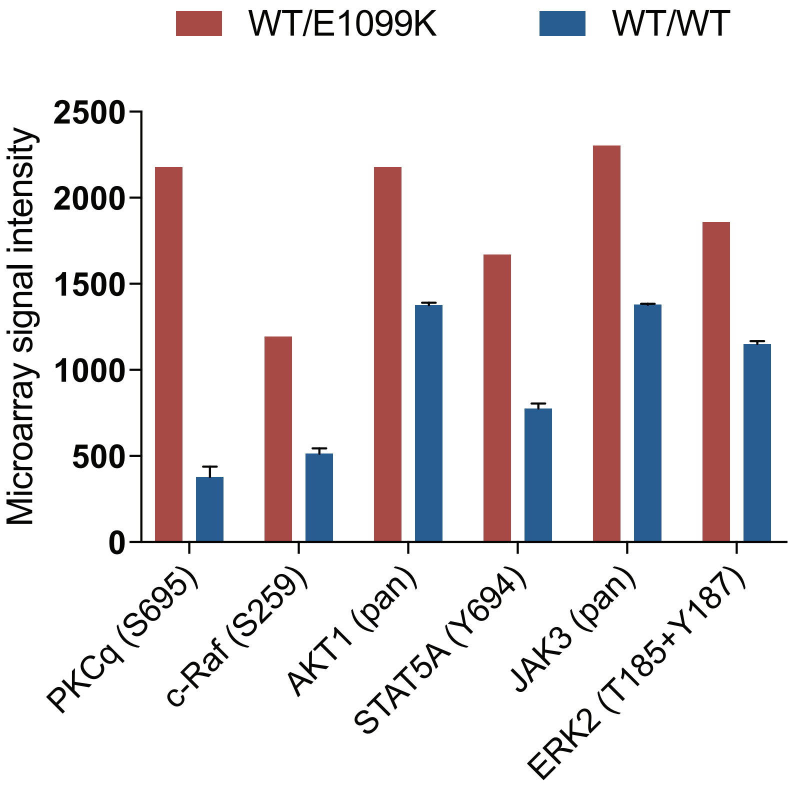
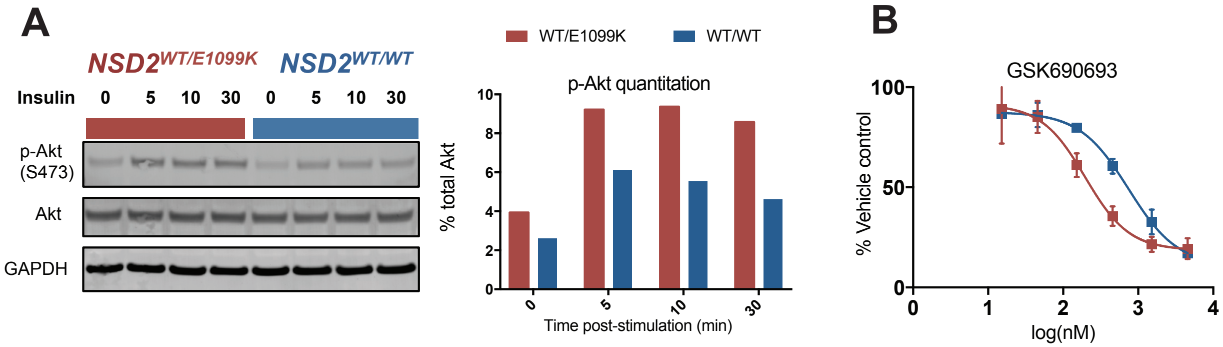


Figure 2.6.1. Altered signaling between NSD2 wildtype and mutant cells

Top candidates for differential phosphorylation between NSD2$^{WT/E1099K}$ and NSD2$^{WT/WT}$ cells selected from protein microarray of whole cell lysates from RCH-ACV edited cell lines (Kinexus).



	WT/E1099K	WT/WT
IC50	197 nM	744 nM

Figure 2.6.2. E1099K cells are susceptible to inhibition of AKT

A, AKT phosphorylation was ascertained by immunoblot with total AKT and GAPDH as loading controls. Dual fluorescence immunoblot was quantified with LI-COR Image Studio. **B,** proliferation of RCH-ACV cells treated with AKT inhibitor, GSK690693, for 72 hours displayed relative to DMSO control. Measured by Celltiter-Glo Luminescent Cell Viability Assay (mean ±SEM, n=3). Data are fit to a four-parameter logistic curve from which the IC-50 was calculated.

2.7. Mechanism of AKT activation by NSD2^{E1099K}

Though AKT is activated in mutant NSD2 cells, its mechanism of action is still unknown. RNA-seq data revealed numerous genes overexpressed in NSD2^{E1099K} cells that could augment signaling such as *DSG2*, *NEO1, GHR*, *LIFR*, *NCAM1*, and *FGF13*. KEGG pathway analysis showed that PI3K/AKT pathway genes were significantly upregulated by E1099K in all three cell lines (Table 2.7).

Cell Line	**E1099K upregulated genes in KEGG PI3K/AKT pathway**
RCH-ACV	PRKAA2; CSF3R; TNXB; FLT1; CSF1; ITGA2B; PDGFA; LAMC1; IL2RG; FGF5; GHR; HGF; FN1; LAMB1; PPP2R3A; GNG11; EPOR; COL4A2; PPP2R2C; LPAR6; COL6A2; KIT; ITGA8; ITGA7; ITGA6; SGK1; FGF13; EPHA2; ITGA9
SEM	EGF; IRS1; F2R; PRKCA; LAMB1; EFNA5; GNG11; THBS4; FGF5; GHR; GNG2; PDGFD; PDGFC; FGF13; TLR4; MET; FGFR1
RPMI-8402	TNXB; ITGB5; ITGB3; LAMC1; PIK3R5; GHR; GNG2; AKT3; PDGFC; SPP1; ITGB8; ITGAV; PDGFRB; NGFR; IL4R; ITGA3; ITGA2; FN1; LAMB1; PTK2; NFKB1; COL1A1; KITLG; LPAR6; IL7; IL2RB; SGK3; FGF13; SGK2; FGFR2; CREB5

Table 2.7. PI3K/AKT pathway genes upregulated by E1099K

DSG2 (desmoglein 2) is a particularly promising candidate because it was the gene most significantly overexpressed between all three mutant cell lines. It was also found to be overexpressed in t(4;14) MM and E2A-PBX1 ALL by our analysis and others' (Brito et al., 2009). DSG2 is known to activate multiple signaling pathways, including AKT, when overexpressed in keratinocytes, and its upregulation is common in many skin cancers (Brennan et al., 2007). While primarily known as a component of desmosomes, in squamous cell carcinoma DSG2 has been found to increase the production of extracellular vesicles (EVs) containing EGFR (epidermal growth factor receptor) that can activate AKT signaling in autocrine or paracrine fashion (Overmiller et al., 2017). To test for the presence of a secreted factor we applied media conditioned by NSD2$^{WT/E1099K}$ cells to wildtype and mutant cells and measured proliferation over 5 days. NSD2$^{WT/WT}$ cells grew 1.6-fold more in E1099K-conditioned media as compared to unconditioned media, while NSD2$^{WT/E1099K}$ cells exhibited negligible increase in proliferation (Figure 2.7A). Next, we generated media conditioned by both wildtype and mutant NSD2 cells pretreated with a broad matrix metalloproteinase inhibitor (MMPi, GM6001) and applied these to serum-starved cells (Figure 2.7B). Broad MMPi treatment has been shown to reduce EV production from DSG2 overexpressing cells by preventing cleavage of the DSG2 ectodomain (Overmiller et al., 2017). In early experiments, immunoblot of NSD2$^{WT/WT}$ cells treated with NSD2^{E1099K} conditioned media revealed phosphorylation of AKT increased to the same level as NSD2$^{WT/E1099K}$ cells (Figure 2.7C). This activation was abrogated when the media was conditioned by cells treated with an MMPi. AKT phosphorylation was decreased in response to media conditioned by NSD2$^{WT/WT}$ cells. With these experiments we have formed the foundation for understanding how the E1099K mutation activates signaling in ALL.

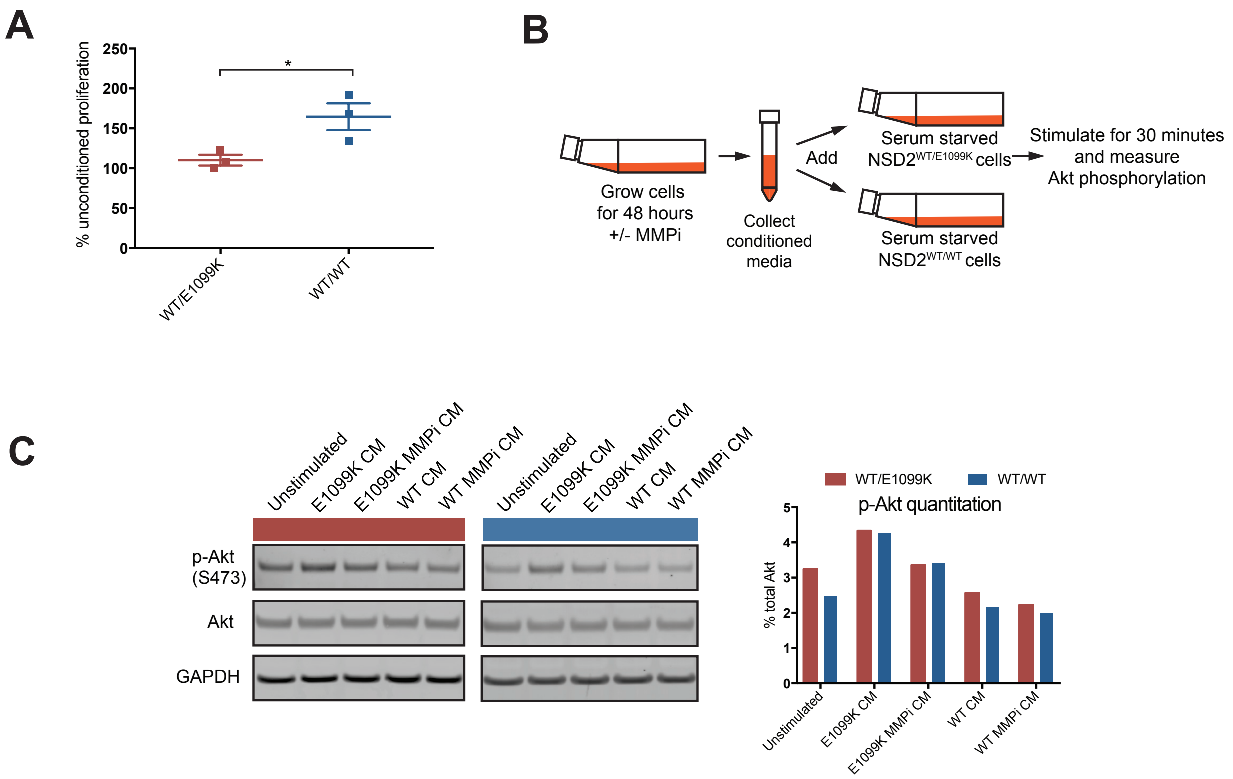


Figure 2.7. A factor secreted by NSD2$^{WT/E1099K}$ cells activates AKT and increases proliferation

A, proliferation of RCH-ACV cells in liquid culture supplemented with media conditioned by NSD2$^{WT/E1099K}$ cells. * $p<0.05$. Measured by Celltiter-Glo Luminescent Cell Viability Assay after 5 days (mean ±SEM, n=3).

B, diagram illustrating design of experiment to test the effect of conditioned media on RCH-ACV activation of PI3K/AKT pathway.

C, AKT phosphorylation measured by immunoblot following stimulation with media conditioned (CM) by NSD2$^{WT/E1099K}$ or NSD2$^{WT/WT}$ RCH-ACV cells treated with GM 6001 (MMPi) or vehicle control.

2.8. Identifying downstream effectors of NSD2 E1099K

In addition to PI3K/AKT activation, numerous other genes and pathways are upregulated by NSD2^{E1099K}. Some of these will prove to be more relevant than others in the promotion of ALL. In order to clarify which genes are most important, we have designed a CRISPR screen targeting genes most likely to be contributing to malignancy (Figure 2.8). In this arrayed screen, we have focused on the 11 genes most significantly upregulated by E1099K across all three edited ALL cell lines: DSG2, ELOVL7, FAT4, GATA6, GATM, MTUS1, NCAM1, PEG10, RASSF8, SGCE, and WASF3. Two sgRNAs per gene were generated to target functional domains, as these are more likely to generate true functional knockouts than those that target the 5’ end of the gene (Shi et al., 2015). The target sequences were cloned into lentiviral plasmids that express sgRNAs with a GFP marker (LRG) using a modified version of Gibson assembly (Gibson et al., 2009; Shi et al., 2015). The sgRNAs targeting one gene will be pooled and transduced into wildtype or E1099K NSD2 cells that stably express spCas9. Flow cytometry will be used to measure the proportion of GFP positive cells and assess how this proportion changes over time. Genes that are critical for proliferation will have sgRNAs that drop out of the population as they generate indels and eliminate gene function, resulting in a reduction in GFP+ cells. sgRNAs that dropout of the population specifically in NSD2$^{WT/E1099K}$ cells, but remain in NSD2$^{WT/WT}$ or NSD2$^{WT/-}$ cells are important for the oncogenic function of NSD2^{E1099K} and can be targeted for inhibition in mutant NSD2-driven malignancies. The benefit of doing an arrayed screen with a small number of genes is that it can be adapted easily to examine many phenotypes beyond proliferation. The assays used in this work to examine the biological consequences of NSD2^{E1099K} can be performed again with small modifications. Counting GFP+ colonies rather than all colonies in clonogenicity assays, for example. Adhesion to stromal cells can be quantified by a plate reader measuring GFP signal. GFP+ cells migrating through a Boyden chamber can be measured by flow cytometry.

In addition to this small, focused arrayed screen, we planned a wider screen to identify downstream effectors of NSD2^{E1099K} activity. We pooled individual shRNAs targeting each of the 84 genes commonly upregulated by E1099K from the Thermo Scientific The RNAi Consortium (TRC) Lentiviral shRNA library. We also selected 5 genes that would serve as positive controls, proffering a survival disadvantage for ALL cells transduced with their shRNAs, and 3 negative control genes that would neither help nor hinder growth of ALL cells when knocked down. The library consists of 4-5 shRNAs per gene, for a total of 373 targeting shRNAs. We will transduce NSD2$^{WT/E1099K}$, NSD2$^{WT/WT}$, and NSD2$^{WT/-}$ cells with virus produced from our targeted library at a multiplicity of infection (MOI) that ensures that a maximum of one shRNA will be present in each cell. Cells will be sequenced 2 days post-infection and then again 8 days post-infection to determine which shRNAs are present in the population.

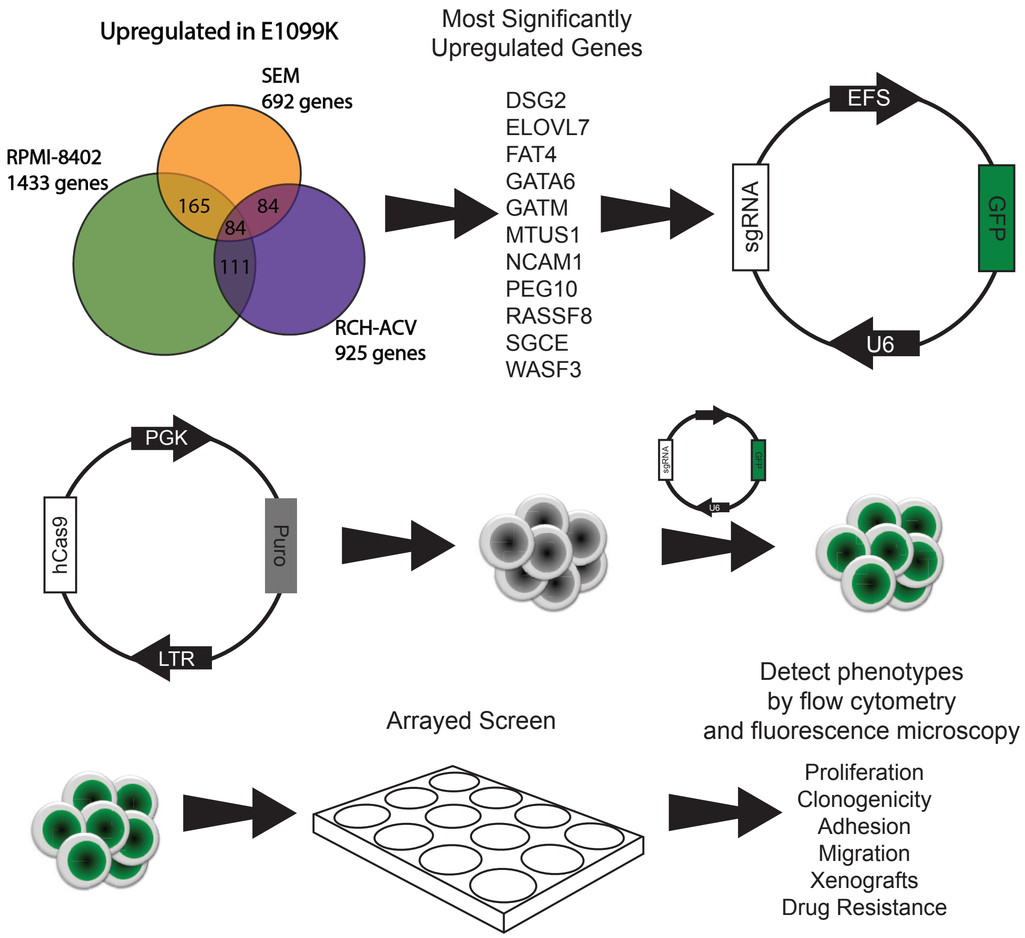

Upregulated in E1099K
SEM
692 genes
RPMI-8402
1433 genes
165
84
84
111
RCH-ACV
925 genes
Most Significantly
Upregulated Genes
DSG2
ELOVL7
FAT4
GATA6
GATM
MTUS1
NCAM1
PEG10
RASSF8
SGCE
WASF3
EFS
sgRNA
GFP
U6
PGK
hCas9
Puro
LTR
sgRNA
GFP
U6
Arrayed Screen
Detect phenotypes
by flow cytometry
and fluorescence microscopy
Proliferation
Clonogenicity
Adhesion
Migration
Xenografts
Drug Resistance

Figure 2.8. Design of a targeted CRISPR array screen to identify important downstream effectors of NSD2 E1099K

2.9. Developing a mouse model of NSD2 E1099K

Genetically engineered cell lines are extraordinarily valuable for understanding the role of NSD2^{E1099K} in malignancy, but they are artificial and encumbered by various mutations and translocations that can confound our analysis. In order to determine whether our cell line-derived findings could be recapitulated *in vivo*, we generated a mouse model of E1099K. This system will provide a new avenue for understanding the function of NSD2 in a model system that only consists of mutations and translocations that we engineer. In this mouse, human $NSD2^{E1099K}$ is placed under control of the Rosa26 promoter, but is preceded by a transcription terminator sequence flanked by LoxP sites. The $NSD2^{E1099K}$ mouse was crossed to Cd19cre mice (Jackson Laboratories) to excise the transcription terminator sequence and drive expression of mutant NSD2 in early B cells. Splenic B cells isolated from mice harboring one copy of $NSD2^{E1099K}$ and heterozygous for Cd19cre exhibited increased NSD2 and H3K36me2 (Figure 2.9). Mice with two copies of $NSD2^{E1099K}$ showed even higher levels of NSD2 and H3K36me2, and additionally showed a reduction in H3K27me3 relative to normal controls (Figure 2.9). The recapitulation of the E1099K epigenetic phenotype in mice suggests that these mice are functioning as expected and will be useful for further study.

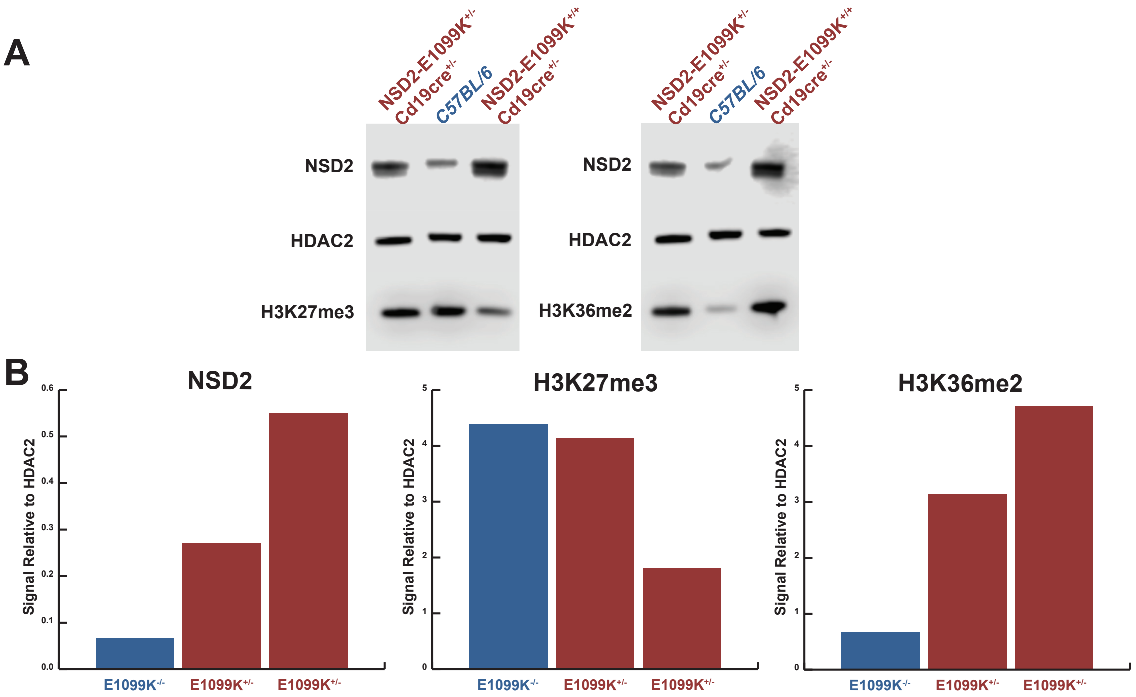


Figure 2.9. Splenic B cells in NSD2 E1099K knock-in mice show appropriate changes in histone methylation

A, immunoblot of splenic B cells isolated from C57BL/6 mice and mice with one or two copies of NSD2 E1099K knocked into the Rosa26 locus in the Cd19cre background.

B, bar graphs represents quantitation of immunoblots.

Chapter 3: Discussion

3.1. E1099K hyperactivates NSD2 function by increasing affinity for chromatin

Modern chemotherapy frequently results in the cure of pediatric ALL, but the pursuit of new modalities is crucial to solve the significant issue of relapse as well as to mitigate the side effects of aggressive chemotherapy. The discovery of the activating E1099K mutation in NSD2 presents a potential therapeutic opportunity, as EZH2, the other major histone methyltransferase with an oncogenic activating mutation, has already proven amenable to chemical inhibition (McCabe et al., 2012). The preliminary success of EZH2 inhibitors in clinical trials suggests that NSD2 inhibitors might also prove valuable in a clinical setting as the role of NSD2 in oncogenesis is established (Chen et al., 2013b). This work firmly establishes NSD2 as a target ripe for inhibition in ALL. Finding inhibitors for NSD2 and NSD2^{E1099K} is an ongoing effort by many labs, but progress is slow. Our model of the structure of NSD2 suggested that the E1099K mutation changes the binding of NSD2 to histones and DNA, providing insight that could be useful for the design of specific drugs that attenuate the action of the mutant protein, while sparing the function of normal NSD2. The predicted increased stability of the NSD2-nucleosome complex and the larger SASA of the NSD2 histone-binding pocket that occur due to the E1099K mutation provide a mechanistic basis for increased methyltransferase activity. This analysis was derived from a homology model of NSD2 built using NSD1, and it is important to note that since our collaborators first performed this work the crystal structure of NSD2 has been solved, albeit with 5 SET domain point mutations to aid in crystallization (Tisi et al., 2016). The strong similarity (root mean square deviation between backbone atoms = 1.9Å) of our model and the new crystal structure validates the results derived from our modeling of the E1099K mutation. In fact, the only significant difference in the structures is located at the N-terminus, which is far removed from the region of H3-NSD2 interaction that we are primarily studying in this work.

FRAP analysis also supports the *in silico* prediction of a more favorable binding energy between mutant NSD2 and the nucleosome. FRAP experiments performed in 2005 by Keats et al. first showed that the motility of NSD2 is significantly less than predicted by protein size alone, indicating that NSD2 binds chromatin tightly (Keats et al., 2005). This work finds that the E1099K mutation reduces NSD2 motility even further. Diffusion of $\text{NSD2}^{\text{E1099K}}$ in the 3.5-minute window of measurement is slower than NSD2^{WT}, as shown by the reduced slope of the fluorescence recovery curve. We also show that the mobile fraction, or the proportion of moving protein during the measurement window, is significantly less. The mobile fraction of wildtype NSD2 that we estimate from Keats' FRAP experiments in 2005 is similar to the mobile fraction we calculate from our wildtype NSD2 FRAP experiments, lending credence to our results.

Interestingly, our *in silico* analysis found superior stability and a larger SASA of the mutant NSD2-nucleosome complex containing canonical, replication-dependent histone H3.1 rather than replacement histone H3.3. H3.1 is associated with repressed genes and is commonly enriched in H3K27me3, while H3.3 is associated with active transcription and H3K36me2 (Hake and Allis, 2006; Szenker et al., 2011). Mass spectroscopic analysis showed that H3.3K36 is frequently dimethylated whether or not NSD2 is mutated. In cells expressing wildtype NSD2, however, a much smaller percentage of H3.1K36 is dimethylated. With the E1099K mutation, levels of dimethylation at H3.1K36 resemble H3.3K36. Increased avidity for mutant NSD2 to H3.1 likely allows the protein to bind a substrate for which it does not have a strong physiological affinity, therefore allowing it to spread aberrantly to H3.1 containing nucleosomes.

Mass spectroscopic analysis of NSD2-mutant and CRISPR-generated isogenic wildtype cells showed that NSD2 mutation caused a 2.5-fold increase in H3K36me2 levels. Strikingly, measurement of *in vivo* rates of histone methylation in NSD2-mutant versus wildtype cells by M4K showed that the rate of H3K36 monomethylation was accelerated, while H3K36 dimethylation was not. Previous data from isogenic KMS11 MM cells with high NSD2 expression compared to low NSD2 expression revealed a larger, three-fold increase in H3K36me2 levels and up a six-fold increased rate of H3K36 dimethylation in addition to 30-times increased rates of monomethylation (Zheng et al., 2012). These qualitative and quantitative differences suggest that the exact mechanism by which E1099K mutation and NSD2 overexpression drive malignancy may differ.

The PWWP domain of NSD2 detects H3K36me2 and improves binding to chromatin, aiding PHD domains that also bind chromatin, and giving NSD2 the capability to read and write the same mark (Huang et al., 2013; Sankaran et al., 2016). Sankaran and colleagues suggest that NSD2 binds H3K36me2 and propagates this mark along chromatin. Mutations in the PWWP domain reduce NSD2 affinity for chromatin and also reduce H3K36 dimethylation. The E1099K mutation aids the normal function of the PWWP domain by stabilizing the interaction of NSD2 with chromatin. Increased stability of NSD2^{E1099K} with H3.1 containing nucleosomes also promotes an atypical substrate. As such, the increased fraction of bound NSD2^{E1099K} detected by

FRAP is likely bound to H3.1 nucleosomes, propagating H3K36 dimethylation. In NSD2-high KMS11 MM cells NSD2 outcompetes and antagonizes PRC2, marking heterochromatic and intergenic regions rich in H3.1 with K36me2. Mutant NSD2 is likely functioning similarly, activating transcription of genes in that were previously designated for silencing. In the case of t(4;14) MM, sheer mass action of the overexpressed NSD2 may increase the interaction of NSD2 with chromatin. In contrast, the specific increase of the rate-limiting monomethylation step in NSD2^{E1099K} cells reflects more subtle gain in function that is dictated by enhanced binding of mutant NSD2 and chromatin. In both cases of NSD2 dysregulation H3.1K36 is increasingly dimethylated instead of H3.3K36. Physiologically, H3.3 increases in aging cells, while decreasing in rapidly proliferating cancer cells (Tvardovskiy et al., 2017). Upregulation of NSD2 activity is noted in many malignancies (Figure 1.14), and may be one mechanism for allowing cells to maintain transcriptional activity even in the context of high H3.1. Specific disruption of NSD2 binding to H3.1 may prove to be an effective strategy for targeting any cancer cells with increased NSD2 methyltransferase activity while sparing normal tissue.

3.2. NSD2 E1099K Activates Oncogenic Programs

This work presents the clearest picture of biological and transcriptional changes that occur due to the oncogenic point mutation NSD2^{E1099K}. Past work compared cell lines harboring the mutation versus those without, or reduced NSD2 expression indiscriminately with shRNAs or sgRNAs (Ding et al., 2017; Jaffe et al., 2013; Oyer et al., 2014). Here, we generated wildtype NSD2 isogenic cell lines derived from 3 different parent cell lines harboring the E1099K mutation. In these cells, the NSD2^{E1099K} mutation reduced apoptosis and enhanced proliferation, clonogenicity, adhesion, and migration, all hallmarks of cancer that are accentuated with leukemia progression

(Hanahan and Weinberg, 2011). NSD2 also enhances adhesion in MM, where knockdown of NSD2 results in impaired adhesion to stromal cells and extracellular matrix, concurrently reducing the ability of these cells to proliferate (Huang et al., 2013; Lauring et al., 2008). Pathway analysis of RNA-seq from NSD2$^{WT/E1099K}$ and NSD2 wildtype cells indicated that genes involved in cell adhesion were prominently upregulated by NSD2^{E1099K}. Adhesion proteins, including neural cell adhesion molecule 1 (*NCAM1, CD56*) and DSG2, often activate downstream signaling pathways (Brennan et al., 2007; Cavallaro and Christofori, 2004). NCAM1 is expressed on neurons, epithelial cells, muscle cells, and pancreatic β cells, but its role in neuronal development and plasticity is most well defined (Cavallaro and Christofori, 2004). In MM, NCAM1 and DSG2 are both upregulated by NSD2, and their expression is implicated in adhesion and proliferation (Brito et al., 2009; Damgaard et al., 2009). Disrupting the activity of these genes could have cascading effects on NSD2^{E1099K} ALL and MM beyond simply reducing adhesion.

The aggressive biology of NSD2$^{WT/E1099K}$ cells was evident *in vivo,* as mice xenografted with mutant cells died significantly faster than those grafted with wildtype cells. Enhanced proliferation and clonogenicity of mutant cells were likely not primary reasons for increased lethality, however. NSD2$^{WT/E1099K}$ cells also exhibited earlier invasion of the central nervous system, a common site of relapse in pediatric ALL (Wynn, 2010). Interestingly, the mutant cells also invade the brain parenchyma, a clinically rare but significant finding, which is likely the major reason for their increased lethality because general tumor burden was not predictive of death. Children with ALL get CNS disease in 10% of cases, but prior to widespread CNS prophylaxis, it was between 50% and 75% (Estey et al., 2008). While treatment can come with associated toxicity, the danger of CNS invasion is more concerning. Interestingly, expression of *NCAM1* in ALL correlates with increased likelihood of CNS invasion, and was suggested as a biomarker to identify patients most in need of aggressive CNS therapy (Ravandi et al., 2002). As mentioned, *NCAM1* was upregulated in all three of our E1099K cell lines along with other genes involved in neural processes. The activation of such a transcriptional program might contribute to the ability of mutant NSD2 cells to adhere and proliferate within the CNS microenvironment.

3.3. Future directions

Going forward we will further characterize the increased affinity of NSD2^{E1099K} for chromatin. Immunoblot for NSD2 will be performed following biochemical fractionation with increasing concentrations of salt on nuclei isolated from mutant and wildtype NSD2 cells (Mendez and Stillman, 2000). *In vitro* methylation assays with purified NSD2 will reveal whether nucleosomes containing H3.3 or H3.1 (Epicypher) are better substrates for the wildtype and mutant proteins.

ChIP-seq will be also useful to identify NSD2, H3.3, and H3.1 occupancy and local changes in histone methylation. How these the presence of these proteins and these PTMs correlate with changes in gene expression will be important to classify direct and indirect targets of NSD2^{E1099K}. It will be interesting to learn by ChIP-seq of NSD2 and H3K36me2 whether NSD2^{E1099K} spreads across the genome in the same fashion as it does in t(4;14) MM. In MM, NSD2 and H3K36me2 are no longer focal, but instead diffusely spread over genomic elements, obliterating most H3K27me3 marks 20 (Popovic et al., 2014). H3K27me3, meanwhile, is redistributed from its wide spread to small foci that are then hyper-repressed. These repressed loci are actually important for the pathogenesis of MM, making these cells sensitive to EZH2 inhibitors. If a similar pattern of histone methylation emerges in NSD2^{E1099K} cells, EZH2 inhibitors should also be tested here. However, the RNA-seq generated from our CRISPR-edited cells showed minimal overlap of NSD2^{E1099K} downregulated genes between cell lines. This finding makes it more likely that the activation of genes is the driver of oncogenesis in ALL.

3.4. Generating new cell lines

This work is based on isogenic cell lines generated from NSD2$^{WT/E1099K}$ where the E1099K mutation has been edited out by CRISPR. One criticism of this work is that there are no cell lines where the E1099K mutation has been knocked into a wildtype cell line. When this work began, our lab was not able to engineer specific mutations into the genome, but instead found success editing out mutations. Co-transfecting spCas9 plasmid with ssDNA oligos was inefficient in our cell lines and suffered from limitations inherent in the protocol. Namely, the transcription and translation of Cas9 did not coincide with transcription of the sgRNA. To solve these issues, new, commercially available spCas9 protein can be mixed with a crRNA and tracrRNA to form ribonucleoprotein (RNP) complexes. These RNPs can be directly transfected into cells and can cut DNA more effectively. In fact, these RNPs have already been very efficient at knocking in specific mutations and epitope tags into MM cell lines (Dupere-Richere, in preparation). We have designed sgRNAs and donor constructs, and we are working on generating cells expressing NSD2^{E1099K} from NSD2$^{WT/WT}$ parental cells. It will also be useful to expand research of the NSD2^{E1099K} mutation into other malignancies. MCL, in particular, would be of great interest due to the dearth of treatments currently available. The MCL cell lines we have profiled to this point, including Jeko, Mino, Maver, and Rec1 are all NSD2$^{WT/WT}$. Direct RNP transfection is being utilized to generate NSD2$^{WT/E1099K}$ cells. Ideally, clones with heterozygous knock-in of E1099K will be selected for further experimentation, though they might be difficult to find. One drawback of the CRISPR system's efficiency is that it tends to make biallelic DSBs, making it difficult to detect clones where only one editing event has occurred (Dow et al., 2015). It might be useful to characterize cells with homozygous E1099K mutations to determine whether there is a role for the wildtype NSD2 and mutant NSD2 working in tandem to drive oncogenesis, but generation of NSD2$^{WT/E1099K}$ cells will more valuable to recapitulate what is seen in cancer.

3.4. Resistance to therapy

An important aspect of NSD2^{E1099K} that was not addressed in this work was its effect on drug resistance. That NSD2^{E1099K} is a minor allele at diagnosis of ALL but grows to dominance after relapse is highly suggestive of a mutation that confers resistance to therapy (Ma et al., 2015). Additionally, t(4;14)+ MM, which is characterized by NSD2 overexpression, is known to be a poor responder to chemotherapy, suggesting that dysregulated NSD2 could contribute to drug resistance (Keats et al., 2003). DNA damaging agents are of particular interest, given the role of NSD2 in DNA damage repair, but glucocorticoids, and other mainstays of ALL therapy are being tested on the isogenic cell lines developed in this work. The differential activation of the PI3K/AKT pathway in NSD2$^{WT/E1099K}$ cells is particularly relevant because activation of PI3K/AKT confers resistance to glucocorticoids in T-ALL (Piovan et al., 2013). PI3K/AKT/mTORC inhibition has been effective against B-ALL in preclinical models, and a recent report showed that PI3K inhibition sensitized resistant cells, including RCH-ACV, to glucocorticoids (Badura et al., 2013; Kruth et al., 2017; Neri et al., 2014). However, glucocorticoid sensitization seemed to be confined to inhibition of PI3K/MAPK, with

AKT/mTOR inhibition not having the same effect (Kruth et al., 2017). In this work we show that RCH-ACV cells are sensitive to AKT inhibition, so inhibiting multiple levels of the PI3K/AKT pathway in combination with glucocorticoid treatment could lead to synergistic effects.

3.5. Mechanism for AKT Pathway Activation

Gaining a better understanding of the mechanism by which AKT is activated in NSD2^{E1099K} ALL will allow us to better target the pathway for inhibition. In addition to the experiments described above that have already begun, we plan to probe the characteristics of EVs secreted from ALL cells. Activation of AKT will also be measured following stimulation of cells with exosomes concentrated from conditioned media. Further characterization of EVs produced by NSD2^{E1099K} cells will be accomplished with the use of fluorescence nanoparticle tracking analysis (NTA). EVs generated by wildtype and mutant cells grown in exosome depleted media will be concentrated and subjected to NTA, allowing for quantitation and measurement of size (Dragovic et al., 2011). Immunoblot of EVs can also be performed to detect the presence of DSG2 ectodomains, EGFR, and EV surface markers.

Beginning with EGFR, immunoblots to detect activation of proteins up and downstream of AKT will be important to understand how AKT activation occurs and its role in NSD2^{E1099K} malignancy. Immunoblots will also be performed following small molecule inhibition of EGFR and shRNA knockdown or CRISPR knockout of DSG2 to test whether either protein is functioning upstream of AKT. It will be important to understand whether other phenotypes enhanced by NSD2^{E1099K} such as colony formation, adhesion, and migration are affected by

PI3K, AKT, and EGFR inhibition. Together these experiments will determine whether AKT activation is necessary for NSD2^{E1099K} to promote oncogenesis.

3.6. Epigenetic Therapy

This work firmly establishes NSD2 as a target ripe for inhibition in ALL due to its central role in regulating proliferation, clonogenicity, adhesion, migration, invasion, and apoptosis. Finding inhibitors for NSD2 and NSD2^{E1099K} is an ongoing effort by many labs, but progress is slow. While this work is continuing, epigenetic therapies that have already been developed are being tested on NSD2$^{WT/E1099K}$ ALL cells. EZH2 inhibitors have proven effective in NSD2-high MM cells. This effect is likely due to hyper-repression of important tumor suppressors by pockets of redistributed EZH2 (Popovic et al., 2014). Though there are very few genes downregulated in NSD2$^{WT/E1099K}$ cells relative to wildtype NSD2 cells, this strategy may still prove effective.

3.7. DNA damage

It remains to be seen whether DNA repair is affected in NSD2^{E1099K} cells as it is in t(4;14) MM, where cells with high levels of NSD2 exhibit increased DNA damage at baseline, but more efficient repair of damage after exposure to DNA damaging chemotherapy (Shah et al., 2016). These cells have high expression of DNA repair pathway genes that goes down with knockdown of NSD2 (Martinez-Garcia et al., 2011). However, it is unclear whether upregulation of these genes is sufficient to explain the difference in repair between NSD2-high and -low cells. NSD2-low cells are still able to recruit necessary components required to repair DNA. Shah et al. suggest that increased H3K36me2 and/or increased chromatin accessibility in NSD2-high cells, by virtue of H3K36me2 spread and H3K27me3 depletion, may be a contributing factor in increased repair (Fnu et al., 2011; Shah et al., 2016; Soria et al., 2012). Pathway analysis of genes upregulated in RCH-ACV, SEM, and RPMI-8402 E1099K+ cell lines relative to wildtype cells did not reveal significant upregulation of any DNA repair pathways.

3.8. Mouse models

Xenografts of CRISPR-edited cell lines were valuable to this work because they showed that NSD2$^{WT/E1099K}$ cells exhibit enhanced malignant properties in a more natural environment, as compared to a culture dish. They reveal additional phenotypes, such as enhanced CNS invasion. Interactions with stromal cells are also relevant for ALL. It will be necessary to test the efficacy of AKT inhibitors and chemotherapeutic agents on xenografted cells to assess their utility *in vivo*. This will also provide an opportunity to better quantify CNS infiltration of RCH-ACV, SEM, and RPMI-8402 NSD2$^{WT/E1099K}$ cell lines by flow cytometry and to determine whether PI3K/AKT inhibition alone or in combination with glucocorticoid therapy has the effect of reducing CNS disease. Immunohistochemical staining for CD56/NCAM1 should also be performed to determine whether this molecule plays a different role in mutant versus wildtype NSD2 CNS infiltrating leukemia.

It is also important to move beyond cell line xenografts towards more robust animal models of ALL. The early B cell NSD2^{E1099K} knockin mouse along with the E2A-PBX1/CD3-null mouse model of B-ALL described in this work are valuable tools for better understanding the effects of NSD2^{E1099K} and NSD2 mutant leukemia. As detailed above, NSD2 E1099K is usually detected in TEL-AML1 or E2A-PBX1 expressing ALL (Jaffe et al., 2013). There is strong evidence that chromosomal translocations often occur prenatally, while NSD2^{E1099K} is only present as a minor allele at diagnosis of ALL, and then takes over as the major clone at relapse (Ding et al., 2017; Greaves and Wiemels, 2003). Together, this indicates that translocations are important early events, and NSD2^{E1099K} likely relies on them and the accumulation of other mutations in order to initiate malignancy. Thus, the expression of NSD2^{E1099K} alone is unlikely to develop into cancer unless other oncogenes are introduced or tumor suppressors knocked out. To create mice expressing a driver oncoprotein along with mutant NSD2 in B cells, we are generating Cd19cre; NSD2^{E1099K}; E2A-PBX1; CD3-null mice. The complex genotype of this mouse makes it difficult to produce, but once the experimental and control genotypes are generated, mouse bone marrow can be isolated and transplanted into up to 20 wildtype, irradiated mice of the same genetic background (Beguelin et al., 2017; Pikman et al., 2006). Not only is this technique is more cost efficient due to the ability to stretch the power of expensive experimental mice, it also allows for *ex vivo* viral transduction of bone marrow to introduce new genes or mutations prior to transplantation. Viral mutagenesis followed by transplantation has been used for E2A-PBX1 mice to identify important oncogenic mutations, and also to reduce time to disease (Bijl et al., 2005). Viral transduction of marrow isolated from E2A-PBX1; CD3-null with wildtype and mutant NSD2 is another way that this experiment can be performed, and reduces the number of crosses required to obtain experimental mice. To examine cooperation of E1099K with TEL-AML1, bone marrow from NSD2^{E1099K} knockin mice can be transduced with TEL-AML1. TEL-AML1 transduction alone promotes pre-leukemic characteristics, but is not sufficient to generate leukemia in mice; perhaps expression of NSD2^{E1099K} would provide a second hit necessary to initiate disease (Jacoby et al., 2014; Morrow et al., 2004; Schindler et al., 2009). Cd19cre; NSD2^{E1099K} knockin mice will still be useful to analyze how basic biological functions of B cells, such as proliferation and differentiation, are affected by the mutation. Assessing chromatin changes by ChIP-seq and gene expression by RNA-seq will also useful to understand the strength of E1099K alone because these cells will be unbiased by any other oncogenes.

Chapter 4: Methods

4.1. NSD2 Modeling

SWISS-MODEL server (Schwede et al., 2003) was employed to construct the NSD2 homology model based upon NSD1 crystal structure. Hydrogen atoms were then added to this homology structure using the REDUCE software (Word et al., 1999). The entire system was minimized for 1,000 steepest descent (SD) steps followed by 500 adopted basis Newton-Raphson (ABNR) steps with harmonic restraints of 1 kcal/(mol Å). We overlaid the structure of the H3K9 peptide from the crystal structure of the homologous protein GLP-H3K9 complex. The H3-NSD2 complex was modeled in two steps using Langevin MD simulation in combination with the ZDOCK server (Pierce et al., 2014). Three 1.5 ns Langevin MD runs were conducted with the final systems using the CHARMM program (Brooks et al., 2009) with CHARMM C36 protein (Best et al., 2012; MacKerell et al., 1998) and nucleic acid (Hart et al., 2012) force field. The Generalized Born-Molecular Volume (GBMV) method (Lee, 2002) was used during the simulation to account for solvation effects. The Leapfrog integrator (Hockney, 1970) was used with a time step of 1.5 fs to integrate the equations of motion. The SHAKE algorithm (Ryckaert, 1977) was applied to constrain the length of covalent bonds involving hydrogen to their equilibrium values. *Methods provided and experiments performed by Wenbo Yu and Alex MacKerell.*

4.2. Cell culture

SEM cells (DSMZ, ACC-546) were cultured in IMDM/10% HI-FBS/1% penicillin-streptomycin. RPMI-8402 (ATCC, CRL-1994) and RCH-ACV cells (DSMZ, ACC-548) were cultured in

RPMI-1640/10% HI-FBS/1% penicillin-streptomycin. ALL cell lines and subclones were authenticated with STR DNA profiling by Genetica. HS-5 (ATCC, CRL-11882) and 293T cells were cultured in DMEM/10% HI-FBS/1% penicillin-streptomycin. All cells were grown in 37°C incubators with 5% CO_2. Routine mycoplasma testing was performed using MycoAlert Detection Kit (Lonza).

4.3. Plasmids, mutagenesis, and fluorescence recovery after photobleaching

NSD2^{E1099K} cDNA was amplified by PCR and inserted into Mfe-I digested pCAG-GFP (Addgene, #11150, (Matsuda and Cepko, 2004)) using HiFi DNA Assembly (NEB). Q5 site-directed mutagenesis (NEB) was used to convert K1099 to E1099. HEK 293T on glass bottom dishes (MatTek) were transfected with NSD2 constructs using FuGENE 6 (Promega). Cells in heated and CO_2 regulated chamber were imaged for FRAP using the 63x objective on a Nikon A1R+ confocal system. Three regions were measured using the NIS Elements Advanced Research software: photobleach, reference, and background. Background measurement was subtracted from all readings, and reference region was used to correct for unintentional photobleaching. 8-12 bright, uniform nuclei were imaged for each experiment, and measurements were plotted in GraphPad v6. For this work, 6 separate experiments were done on different days. Mobile fraction was calculated by the following equation described by Mueller and colleagues (Mueller et al., 2012):

$$\boldsymbol{M = \frac{\eta - F_0}{1 - F_0}}$$

In this equation, M is equal to mobile fraction, η is the final plateaued fluorescence intensity, and F_0 is the fluorescence intensity immediately following photobleaching. Imaging work was performed at the Northwestern University Center for Advanced Microscopy supported by NCI CCSG P30 CA060553.

4.4. CRISPR Generation of NSD2-Edited Cells

Direct synthesis of a double-stranded fragment (IDT gBlock) was used to generate a customized sgRNA expression construct composed of a U6 promoter, a 20bp sequence targeting the K1099 allele of *NSD2* (TGGGGAGCTGATCGACAAGG), and an sgRNA scaffold described previously (Chen et al., 2013a). This construct was amplified by PCR, column purified (Qiagen), and transfected along with Cas9-GFP (Addgene #42234, (Jinek et al., 2013)) using the Neon Transfection System (Life Technologies). Voltage: 1400V, pulse width: 20ms, 2 pulses. GFP-positive cells were sorted (BD FACSAria) approximately 20 hours post-transfection, plated at a density of 200/ml in methylcellulose (MethoCult H4100, Stem Cell Technologies), and single-cell derived colonies were isolated 10-14 days later. DNA extracted from cells (QuickExtract, Epicentre) was amplified by PCR and Sanger sequenced by GeneWiz to detect CRISPR-editing events. PCR products were cloned and sequenced to assess mutant allele frequency (PCR Cloning Kit, NEB).

4.5. Immunoblot

Nuclear extracts were prepared from cells using the Nuclear Complex Co-IP kit (Active Motif) and whole cell lysates using RIPA lysis buffer 2 (ADI-80-1284, Enzo). Proteins were separated using SDS-PAGE, transferred to PVDF membranes, and probed with antibodies to EZH2 (5246, Millipore), HDAC2 (05-814, Millipore), H3K36me2 (2901, Cell Signaling Tech.), H3K27me3

(9733, Cell Signaling Tech.; 07-449, Millipore), NSD2 (Martinez-Garcia et al., 2011), histone H4 (2935, Cell Signaling Tech.), phospho-Akt (9271, Cell Signaling Tech.), pan-Akt (2920, Cell Signaling Tech.), and GAPDH (47724, Santa Cruz Biotech.). Secondary antibodies used were HRP-conjugated anti-mouse IgG (95017-332, VWR) or donkey anti-rabbit IgG (95017-556, VWR) for chemiluminescence and Alexa-Fluor 790-conjugated goat anti-mouse IgG (A11357, ThermoFisher Scientific) or Alexa-Fluor 680-conjugated donkey anti-rabbit IgG (A10043, ThermoFisher Scientific) for fluorescence.

4.6. NSD2-nucleosome association assay

Recombinant NSD2 (Cat #HMT-21-122, lot #1963), E1099K-NSD2 (Cat #HMT-21-159, lot #1520) and biotinylated oligonucleosomes purified from HeLa cells (Cat #HMT-35-160, lot#1542) were purchased from Reaction Biology Corp. (Malvern, PA). Briefly, 0.25µM of recombinant enzymes and biotinylated nucleosome were incubated for 15 minutes at 30C in a 30ul reaction containing 50mM TrisCl pH 8.5, 5mM $MgCl_2$, 5uM SAM, 2µM TCEP, 1% BSA, and either 100mM, 200mM, 300mM, or 400mM NaCl in siliconized tubes. Following incubation, 10% of the reaction was kept as the input fraction. Reactions were diluted to 300ul in reaction buffer with 0.5% NP-40 and 10µl of streptavidin magnetic beads (Pierce Cat#88817). Reactions were incubated at 4C for 1 hour and nucleosome associated protein was isolated by magnetic bead capture. Complexes were washed three times in reaction buffer with 0.5% NP-40 and the appropriate salt concentration. Input and bound fractions were resolved by SDS-PAGE, and immunoblotting with antibodies to NSD2 (Abcam, Cat# ab75359) and H3 (Cell signaling Cat# 14269S) was performed to evaluate the fraction of NSD2 bound. Immunoblot was quantified by ImageJ densitometry analysis. *Methods provided and experiments performed by Richard Bennett.*

4.6. Quantitation of histone modifications by targeted mass spectrometry

Acid extracted histones from *in vitro* methylation assay or isolated from nuclei were digested with trypsin and chemically derivatized using propionic anhydride as described previously (Garcia et al., 2007). Histone peptides were analyzed by nanoLC-QqQ mass spectrometer (Dionex UltiMate 3000 and ThermoFisher Scientific TSQ Quantum) using selected reaction monitoring (SRM/MRM) method reported previously (Zheng et al., 2012). Data were analyzed using Skyline software (MacLean et al., 2010). The relative level of specific modified peptide was calculated by dividing its peak area against the total of all modified peptides sharing the same sequence. Individual H3K27 and H3K36 methylation levels were aggregated from combinatorial methylation levels measured from 16 combinatorial methylation species found in histone H3 peptide from residue 27 to 40 (KSAPATGGVKKPHR). *Methods provided and experiments performed by Xiaoxiao Huang.*

4.7. MS-based measurement and modeling of histone methylation kinetics

SEM and RCH-ACV cells were cultured as described above, then switched to media with [$^{13}C_6$]Arg and [$^{13}C_1$,$^{2}H_3$]Met (Cambridge Isotope Laboratories) and 10% dialyzed FBS (Sigma F0392). Cells were harvested at 0, 5, 10 and 25 hours after the switch. Histone molecules were extracted by acid extraction, derivatized, and digested with trypsin prior to LC-MS. Selected reaction monitoring (SRM) transitions were developed and data were analyzed using Skyline

(MacLean et al., 2010). The measured relative abundance of each species was used in a kinetic model (Zheng et al., 2014). *Methods provided and experiments performed by Xiaoxiao Huang.*

4.8. Lentivirus production and infection of leukemia cell lines

293T cells were transfected with luciferase vector (pFU-L2T; gift from Dr. Marcus Peter, Northwestern University) and lentiviral packaging plasmids (psPAX2 and VSVG) with FuGENE 6 (Promega). Viral supernatant was collected and passed through 0.45um filter at 48- and 72-hours post-transfection. Cells were infected with spinoculation (2000RPM for 90 minutes) and sorted by flow cytometry using the fluorescent Tomato protein 24-48 hours post-infection.

4.9. In Vivo Tumorigenicity

Luciferase-expressing ALL cell lines were resuspended in PBS. $5x10^6$ cells were injected into female NOD-SCID mice (Jackson Laboratory) between 12-18 weeks of age through the tail vein. To assess tumor growth, mice were subcutaneously injected with luciferin (Gold Biotechnology) in PBS, and were analyzed by bioluminescent imaging (IVIS Spectrum, Xenogen) 15 minutes post-injection. Mice were euthanized when physical examination, often done in collaboration with veterinary services, revealed excessive tumor burden. Experiments were conducted in pathogen-free conditions according to protocols approved by Northwestern University (#IS00000557) and University of Florida (#201509176) Institutional Animal Care and Use Committees. Organs were fixed in neutral buffered formalin then processed, embedded in paraffin, sectioned, and stained with hematoxylin and eosin or hHLA-ABC by the Molecular

Pathology Core at the College of Medicine, University of Florida. Slides were then analyzed in collaboration with Benjamin Durham (Memorial Sloan Kettering Cancer Center).

4.10. Proliferation

ALL cells were plated at 1000 cells/well in 96-well black cell culture plates (BD Falcon) in sextuplicate for each time point. For AKT inhibition, $2.5\text{x}10^4$ cells were plated and treated with GSK690693 (Selleckchem) or DMSO for 72 hours. CellTiter-Glo (Promega) detected using BMG plate-reader.

4.11. Colony formation

Methylcellulose-based medium (Methocult H4100, StemCell Technologies) was prepared using appropriate media for each cell line. Cells were resuspended in methylcellulose medium at 4000 per ml and 1ml media/well was added to a 12-well plate. After 7 days colonies were counted in 5 fields and averaged.

4.12. Annexin V staining

Early passage ALL cells were cultured for 4 days as detailed above, then stained using a FITC-Annexin V kit (BioLegend) according to manufacturer protocol, and analyzed on a BD Accuri C6 flow cytometer.

4.13. Propidium iodide staining

Early passage ALL cells were cultured for 4 days as detailed above, and then were fixed in 70% ethanol and stained with FxCycle PI/RNase Staining Solution (ThermoFisher Scientific) according to manufacturer protocol. Analysis was performed on a BD Accuri C6 flow cytometer.

4.14. Adhesion

HS-5 cells were plated at 1500 per well of a 96-well opaque black plate for 24 hours, coating the bottom of the well. ALL cells were stained with 10uM CellTracker Green CMFDA (Life Technologies) in 37C Opti-MEM (Life Technologies). Stained cells were gently washed twice with 37C PBS then plated on HS-5 cells at 5000 per well in triplicate. Cells were spun at 1200 rpm for 5 minutes and allowed to adhere for 24 hours at which point wells were gently washed twice with PBS. A plate reader (BMG) was used to measure the fluorescent signal.

4.15. Migration

ALL cells were resuspended in serum-free media and added to top of Boyden chamber (VWR 29442-118) at $2x10^5$ per well, while serum-containing media was added to bottom chamber. After 24 hours cells in bottom chamber were quantified by flow cytometry.

4.16. RNA extraction, cDNA synthesis, and RNA-sequencing

ALL cells were homogenized with Qiashredder columns (Qiagen), and RNA was extracted with the RNeasy Plus Mini kit (Qiagen). RNA was quantified by NanoDrop spectrophotometer (ThermoFisher Scientific). Library preparation and sequencing were performed by the

Epigenomics Core Facility of Weill Cornell Medicine using TruSeq RNA Library preparation kit (Illumina) and HiSeq 2500 system (Illumina). For RNA-seq validation, cDNA was synthesized with the iScript cDNA synthesis kit (BioRad). Quantitative real time PCR (qRT-PCR) was performed using SYBR green (Roche) on a Lightcycler 480 II (Roche).

4.17. RNA-sequencing alignment and analysis

Short reads were filtered and trimmed using Trimmomatic (v.0.36) and aligned to the hg38 reference genome using the STAR aligner (v.2.5.2a). Expression quantification and differential expression analysis were performed using RSEM (v.1.2.31). We used a log2(fold change) limit of 1 and p-value<5%. The analysis was performed on HiPErGator, a high-performance cluster operated by University of Florida Research Computing. GSEA was performed using the Enrichr web server (Chen et al., 2013b; Kuleshov et al., 2016) or Broad GSEA software (Subramanian et al., 2007). The significance of the intersections between sets of genes was determined by random sampling. 1000 Sets of the same size as the input ones containing random genes were generated and the fraction of cases in which the intersection was equal to or larger than the observed one was computed. This fraction represents the p value of the observed intersection.

4.18. Antibody microarray

Lysates of RCH-ACV NSD2$^{WT/E1099K}$ and NSD2$^{WT/WT}$ cells were prepared with Kinexus lysis buffer. Kinexus applied lysates KAM-900P Antibody Microarray and performed analysis.

4.19. Conditioned media

RCH-ACV NSD2$^{WT/E1099K}$ and NSD2$^{WT/WT}$ cells were plated at a density of 1x10^6 cells/mL in complete RPMI-1640, complete RPMI-1640 + DMSO, or complete RPMI-1640 + 20uM GM 6001 (Abcam) and allowed to condition media for 48 hours. Conditioned media was then filtered through 22um filters.

4.20. Serum starvation and stimulation

RCH-ACV NSD2$^{WT/E1099K}$ and NSD2$^{WT/WT}$ cells were plated at a density of 1x10^6 cells/mL in 1.2 mL RPMI-1640 without FBS for starvation overnight. In the morning cells were stimulated with 5ng/mL insulin (Sigma-Aldrich) or 0.6mL conditioned media for indicated time. Then lysates were prepared for immunoblot as described above.

4.21. Cloning of crRNAs into sgRNA expression vectors

LRG2.1T, provided by Christopher Vakoc (Cold Spring Harbor Laboratory), is a mammallian lentiviral vector with a GFP marker that expresses an sgRNA. It was digested with BsmBI (NEB) and gel purified (Qiagen). 20 base pair crRNAs targeting functional domains of genes of interest were designed using Desktop Genetics (deskgen.com). In order to clone the crRNA into the plasmid, plasmid-complementary sequence was added to each end of the target sequence. (5’-GTGGAAAGGACGAAACACCG-3’) was addended to the 5’ end of the target sequence and (5’-GTTTTAGAGCTAGAAATAGC-3’) to the 3’ end. The full, 60 base pair oligo was synthesized by IDT. The cloning reaction consisted of 2x NEBuilder HiFi DNA assembly master mix (NEB), 75ng BsmBI-digested LRG2.1T, 500nM target sequence oligo, in 15uL total volume.

This reaction was incubated at 50C for one hour and then transformed into high efficiency Stable Competent *E. coli* (NEB). 2-3 clones were selected for outgrowth and DNA was isolated using ZR Plasmid Miniprep (Zymo Research). In order to identify working plasmids, clones were sequenced by GeneWiz using the U6-forward primer (5'-GCATATACGATACAAGGCTGTTAG-3').

4.22. Generation and management of NSD2^{E1099K} knockin and E2A-PBX1 mice

NSD2^{E1099K} knockin mice were generated in collaboration with inGenious Targeting. A construct was designed to contain a neomycin cassette and transcription termination sequence surrounded by LoxP (LOX/Neo-stop_2XSV40-polyA/LOX) sites upstream of human *NSD2*^{E1099K} cDNA. Homology arms targeted this construct to the *Rosa26* locus. C57BL/6 embryonic stem cells with successfully integration were injected into Balb/c blastocysts. Resulting black mice were crossed to C57BL/6N. Mice were screened for the transgene by PCR, and those harboring the transgene were crossed back to C57BL/6 background. Cd19cre mice heterozygous for the Cre transgene (Stock No: 006785 - Jackson Laboratories) were crossed to NSD2^{E1099K} mice to delete the LOX/Neo-stop_2XSV40-polyA/LOX cassette and drive expression of mutant NSD2 in early B cells. Mice were further crossed to NSD2^{E1099K} mice in order to obtain experimental mice with two copies of mutant NSD2. E2A-PBX1 and CD3ε$^{-/-}$ mice were generously provided by Dr. Janetta Bijl (Université de Montréal).

References

Allis, C. D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat Rev Genet *17*, 487-500.

Alsadeq, A., and Schewe, D. M. (2017). Acute lymphoblastic leukemia of the central nervous system: on the role of PBX1. Haematologica *102*, 611-613.

Altomare, D. A., and Testa, J. R. (2005). Perturbations of the AKT signaling pathway in human cancer. Oncogene *24*, 7455-7464.

Andersen, E. F., Carey, J. C., Earl, D. L., Corzo, D., Suttie, M., Hammond, P., and South, S. T. (2013). Deletions involving genes WHSC1 and LETM1 may be necessary, but are not sufficient to cause Wolf-Hirschhorn Syndrome. European journal of human genetics : EJHG *22*, 464-470.

Andersson, A., Olofsson, T., Lindgren, D., Nilsson, B., Ritz, C., Eden, P., Lassen, C., Rade, J., Fontes, M., Morse, H.*, et al.* (2005). Molecular signatures in childhood acute leukemia and their correlations to expression patterns in normal hematopoietic subpopulations. Proc Natl Acad Sci U S A *102*, 19069-19074.

Asangani, I. a., Ateeq, B., Cao, Q., Dodson, L., Pandhi, M., Kunju, L. P., Mehra, R., Lonigro, R. J., Siddiqui, J., Palanisamy, N.*, et al.* (2013). Characterization of the EZH2-MMSET histone methyltransferase regulatory axis in cancer. Molecular cell *49*, 80-93.

Aspland, S. E., Bendall, H. H., and Murre, C. (2001). The role of E2A-PBX1 in leukemogenesis. Oncogene *20*, 5708-5717.

Aster, J. C., and DeAngelo, D. J. (2017). Acute Leukemias. In Pathophysiology of Blood Disorders, 2e, J.C. Aster, and H.F. Bunn, eds. (New York, NY: McGraw-Hill Education).

Ausio, J. (2006). Histone variants--the structure behind the function. Brief Funct Genomic Proteomic *5*, 228-243.

Badura, S., Tesanovic, T., Pfeifer, H., Wystub, S., Nijmeijer, B. A., Liebermann, M., Falkenburg, J. H., Ruthardt, M., and Ottmann, O. G. (2013). Differential effects of selective inhibitors targeting the PI3K/AKT/mTOR pathway in acute lymphoblastic leukemia. PLoS One *8*, e80070.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. V., Sonkin, D.*, et al.* (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603-607.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Battaglia, A., Carey, J. C., and South, S. T. (1993). Wolf-Hirschhorn Syndrome. In GeneReviews((R)), M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J.H. Bean, H.C. Mefford, K. Stephens, A. Amemiya, and N. Ledbetter, eds. (Seattle (WA)).

Beà, S., Valdés-Mas, R., Navarro, A., Salaverria, I., Martín-Garcia, D., Jares, P., Giné, E., Pinyol, M., Royo, C., Nadeu, F.*, et al.* (2013). Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. Proceedings of the National Academy of Sciences of the United States of America *110*, 18250-18255.

Beguelin, W., Rivas, M. A., Calvo Fernandez, M. T., Teater, M., Purwada, A., Redmond, D., Shen, H., Challman, M. F., Elemento, O., Singh, A., and Melnick, A. M. (2017). EZH2 enables germinal centre formation through epigenetic silencing of CDKN1A and an Rb-E2F1 feedback loop. Nat Commun *8*, 877.

Berdasco, M., and Esteller, M. (2010). Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Developmental cell *19*, 698-711.

Bergemann, A. D., Cole, F., and Hirschhorn, K. (2005). The etiology of Wolf-Hirschhorn syndrome. Trends in genetics : TIG *21*, 188-195.

Best, R. B., Zhu, X., Shim, J., Lopes, P. E., Mittal, J., Feig, M., and Mackerell, A. D., Jr. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. J Chem Theory Comput *8*, 3257-3273.

Bibikova, M., Golic, M., Golic, K. G., and Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics *161*, 1169-1175.

Bijl, J., Sauvageau, M., Thompson, A., and Sauvageau, G. (2005). High incidence of proviral integrations in the Hoxa locus in a new model of E2a-PBX1-induced B-cell leukemia. Genes & development *19*, 224-233.

Black, B. E., Foltz, D. R., Chakravarthy, S., Luger, K., Woods, V. L., Jr., and Cleveland, D. W. (2004). Structural determinants for generating centromeric chromatin. Nature *430*, 578-582.

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K.*, et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349-353.

Brennan, D., Hu, Y., Joubeh, S., Choi, Y. W., Whitaker-Menezes, D., O'Brien, T., Uitto, J., Rodeck, U., and Mahoney, M. G. (2007). Suprabasal Dsg2 expression in transgenic mouse skin confers a hyperproliferative and apoptosis-resistant phenotype to keratinocytes. J Cell Sci *120*, 758-771.

Brito, J. L. R., Walker, B., Jenner, M., Dickens, N. J., Brown, N. J. M., Ross, F. M., Avramidou, A., Irving, J. a. E., Gonzalez, D., Davies, F. E., and Morgan, G. J. (2009). MMSET deregulation affects cell cycle progression and adhesion regulons in t(4;14) myeloma plasma cells. Haematologica *94*, 78-86.

Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., *et al.* (2009). CHARMM: the biomolecular simulation program. J Comput Chem *30*, 1545-1614.

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., Dickman, M. J., Makarova, K. S., Koonin, E. V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.

Broyl, A., Hose, D., Lokhorst, H., de Knegt, Y., Peeters, J., Jauch, A., Bertsch, U., Buijs, A., Stevens-Kroef, M., Beverloo, H. B., *et al.* (2010). Gene expression profiling for molecular classification of multiple myeloma in newly diagnosed patients. Blood *116*, 2543-2553.

Campos-Sanchez, E., Deleyto-Seldas, N., Dominguez, V., Carrillo-de-Santa-Pau, E., Ura, K., Rocha, P. P., Kim, J., Aljoufi, A., Esteve-Codina, A., Dabad, M., *et al.* (2017). Wolf-Hirschhorn Syndrome Candidate 1 Is Necessary for Correct Hematopoietic and B Cell Development. Cell Rep *19*, 1586-1601.

Cantley, L. C. (2002). The phosphoinositide 3-kinase pathway. Science *296*, 1655-1657.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039-1043.

Cavallaro, U., and Christofori, G. (2004). Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. Nat Rev Cancer *4*, 118-132.

Chan, K. M., Fang, D., Gan, H., Hashizume, R., Yu, C., Schroeder, M., Gupta, N., Mueller, S., James, C. D., Jenkins, R., *et al.* (2013). The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. Genes Dev *27*, 985-990.

Chang, F., Lee, J. T., Navolanic, P. M., Steelman, L. S., Shelton, J. G., Blalock, W. L., Franklin, R. A., and McCubrey, J. A. (2003). Involvement of PI3K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: a target for cancer chemotherapy. Leukemia *17*, 590-603.

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., Park, J., Blackburn, E. H., Weissman, J. S., Qi, L. S., and Huang, B. (2013a). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479-1491.

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., and Ma'ayan, A. (2013b). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Chen, T., and Dent, S. Y. R. (2014). Chromatin modifiers and remodellers: regulators of cellular differentiation. Nature reviews Genetics *15*, 93-106.

Chesi, M., Nardini, E., Lim, R. S., Smith, K. D., Kuehl, W. M., and Bergsagel, P. L. (1998). The t(4;14) translocation in myeloma dysregulates both FGFR3 and a novel gene, MMSET, resulting in IgH/MMSET hybrid transcripts. Blood *92*, 3025-3034.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. a., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science (New York, NY) *339*, 819-823.

Damgaard, T., Knudsen, L. M., Dahl, I. M. S., Gimsing, P., Lodahl, M., and Rasmussen, T. (2009). Regulation of the CD56 promoter and its association with proliferation, anti-apoptosis and clinical factors in multiple myeloma. Leukemia & lymphoma *50*, 236-246.

Dedera, D. A., Waller, E. K., LeBrun, D. P., Sen-Majumdar, A., Stevens, M. E., Barsh, G. S., and Cleary, M. L. (1993). Chimeric homeobox gene E2A-PBX1 induces proliferation, apoptosis, and malignant lymphomas in transgenic mice. Cell *74*, 833-843.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602-607.

Ding, L. W., Sun, Q. Y., Tan, K. T., Chien, W., Thippeswamy, A. M., Eng Juh Yeoh, A., Kawamata, N., Nagata, Y., Xiao, J. F., Loh, X. Y.*, et al.* (2017). Mutational Landscape of Pediatric Acute Lymphoblastic Leukemia. Cancer Res *77*, 390-400.

Dow, L. E., Fisher, J., O'Rourke, K. P., Muley, A., Kastenhuber, E. R., Livshits, G., Tschaharganeh, D. F., Socci, N. D., and Lowe, S. W. (2015). Inducible in vivo genome editing with CRISPR-Cas9. Nat Biotechnol *33*, 390-394.

Dragovic, R. A., Gardiner, C., Brooks, A. S., Tannetta, D. S., Ferguson, D. J., Hole, P., Carr, B., Redman, C. W., Harris, A. L., Dobson, P. J.*, et al.* (2011). Sizing and phenotyping of cellular vesicles using Nanoparticle Tracking Analysis. Nanomedicine *7*, 780-788.

Druker, B. J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G. M., Fanning, S., Zimmermann, J., and Lydon, N. B. (1996). Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr–Abl positive cells. Nature Medicine *2*, 561-566.

Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A. I., Wutz, A., and Bickmore, W. A. (2010). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. Mol Cell *38*, 452-464.

Estey, E. H., Faderl, S. H., and Kantarjian, H. (2008). Hematologic malignancies : acute leukemias, (Berlin ; New York: Springer).

Eswaran, J., Sinclair, P., Heidenreich, O., Irving, J., Russell, L. J., Hall, A., Calado, D. P., Harrison, C. J., and Vormoor, J. (2015). The pre-B-cell receptor checkpoint in acute lymphoblastic leukaemia. Leukemia *29*, 1623-1631.

Ezponda, T., Popovic, R., Shah, M. Y., Martinez-Garcia, E., Zheng, Y., Min, D.-J., Will, C., Neri, a., Kelleher, N. L., Yu, J., and Licht, J. D. (2013). The histone methyltransferase MMSET/WHSC1 activates TWIST1 to promote an epithelial-mesenchymal transition and invasive properties of prostate cancer. Oncogene *32*, 2882-2890.

Fabbri, G., Rasi, S., Rossi, D., Trifonov, V., Khiabanian, H., Ma, J., Grunn, A., Fangazio, M., Capello, D., Monti, S.*, et al.* (2011). Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. J Exp Med *208*, 1389-1401.

Fang, D., Gan, H., Lee, J. H., Han, J., Wang, Z., Riester, S. M., Jin, L., Chen, J., Zhou, H., Wang, J.*, et al.* (2016). The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. Science *352*, 1344-1348.

Finelli, P., Fabris, S., Zagano, S., Baldini, L., Intini, D., Nobili, L., Lombardi, L., Maiolo, a. T., and Neri, a. (1999). Detection of t(4;14)(p16.3;q32) chromosomal translocation in multiple myeloma by double-color fluorescent in situ hybridization. Blood *94*, 724-732.

Fnu, S., Williamson, E. a., De Haro, L. P., Brenneman, M., Wray, J., Shaheen, M., Radhakrishnan, K., Lee, S.-H., Nickoloff, J. a., and Hromas, R. (2011). Methylation of histone H3 lysine 36 enhances DNA repair by nonhomologous end-joining. Proceedings of the National Academy of Sciences of the United States of America *108*, 540-545.

Fuka, G., Kantner, H. P., Grausenburger, R., Inthal, A., Bauer, E., Krapf, G., Kaindl, U., Kauer, M., Dworzak, M. N., Stoiber, D.*, et al.* (2012). Silencing of ETV6/RUNX1 abrogates PI3K/AKT/mTOR signaling and impairs reconstitution of leukemia in xenografts. Leukemia *26*, 927-933.

Fuka, G., Kauer, M., Kofler, R., Haas, O. A., and Panzer-Grumayer, R. (2011). The leukemia-specific fusion gene ETV6/RUNX1 perturbs distinct key biological functions primarily by gene repression. PLoS One *6*, e26348.

Gambacorti-Passerini, C., Antolini, L., Mahon, F. X., Guilhot, F., Deininger, M., Fava, C., Nagler, A., Della Casa, C. M., Morra, E., Abruzzese, E.*, et al.* (2011). Multicenter independent assessment of outcomes in chronic myeloid leukemia patients treated with imatinib. J Natl Cancer Inst *103*, 553-561.

Garcia, B. A., Mollah, S., Ueberheide, B. M., Busby, S. A., Muratore, T. L., Shabanowitz, J., and Hunt, D. F. (2007). Chemical derivatization of histones for facilitated analysis by mass spectrometry. Nat Protoc *2*, 933-938.

Gaynes, J. S., Jonart, L. M., Zamora, E. A., Naumann, J. A., Gossai, N. P., and Gordon, P. M. (2017). The central nervous system microenvironment influences the leukemia transcriptome and enhances leukemia chemo-resistance. Haematologica *102*, e136-e139.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods *6*, 343-345.

Goldberg, A. D., Banaszynski, L. A., Noh, K. M., Lewis, P. W., Elsaesser, S. J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X.*, et al.* (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell *140*, 678-691.

Golub, T. R. (2007). Genomics: global views of leukaemia. Nature *446*, 739-740.

Gorisch, S. M., Wachsmuth, M., Toth, K. F., Lichter, P., and Rippe, K. (2005). Histone acetylation increases chromatin accessibility. J Cell Sci *118*, 5825-5834.

Gourzones-Dmitriev, C., Kassambara, A., Sahota, S., Reme, T., Moreaux, J., Bourquard, P., Hose, D., Pasero, P., Constantinou, A., and Klein, B. (2013). DNA repair pathways in human multiple myeloma: role in oncogenesis and potential targets for treatment. Cell Cycle *12*, 2760-2773.

Greaves, M. F., and Wiemels, J. (2003). Origins of chromosome translocations in childhood leukaemia. Nat Rev Cancer *3*, 639-649.

Gutierrez, A., Sanda, T., Grebliunaite, R., Carracedo, A., Salmena, L., Ahn, Y., Dahlberg, S., Neuberg, D., Moreau, L. A., Winter, S. S.*, et al.* (2009). High frequency of PTEN, PI3K, and AKT abnormalities in T-cell acute lymphoblastic leukemia. Blood *114*, 647-650.

Hajdu, I., Ciccia, A., Lewis, S. M., and Elledge, S. J. (2011). Wolf-Hirschhorn syndrome candidate 1 is involved in the cellular response to DNA damage. Proceedings of the National Academy of Sciences of the United States of America *108*, 13130-13134.

Hake, S. B., and Allis, C. D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". Proc Natl Acad Sci U S A *103*, 6428-6435.

Hanahan, D., and Weinberg, R. a. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Hanley-Lopez, J., Estabrooks, L. L., and Stiehm, R. (1998). Antibody deficiency in Wolf-Hirschhorn syndrome. J Pediatr *133*, 141-143.

Hart, K., Foloppe, N., Baker, C. M., Denning, E. J., Nilsson, L., and Mackerell, A. D., Jr. (2012). Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. J Chem Theory Comput *8*, 348-362.

Heinz, S., Romanoski, C. E., Benner, C., and Glass, C. K. (2015). The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol *16*, 144-154.

Heng, T. S., Painter, M. W., and Immunological Genome Project, C. (2008). The Immunological Genome Project: networks of gene expression in immune cells. Nat Immunol *9*, 1091-1094.

Henikoff, S., and Smith, M. M. (2015). Histone variants and epigenetics. Cold Spring Harb Perspect Biol *7*, a019364.

Hockney, R. W. (1970). Methods in Computational Physics, (New York: Academic Press).

Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.

Huang, Z., Wu, H., Chuai, S., Xu, F., Yan, F., Englund, N., Wang, Z., Zhang, H., Fang, M., Wang, Y.*, et al.* (2013). NSD2 Is recruited through Its PHD domain to oncogenic gene loci to drive multiple myeloma. Cancer Research *73*, 6277-6288.

Hudlebusch, H. R., Santoni-Rugiu, E., Simon, R., Ralfkiær, E., Rossing, H. H., Johansen, J. V., Jørgensen, M., Sauter, G., and Helin, K. (2011a). The histone methyltransferase and putative oncoprotein MMSET is overexpressed in a large variety of human tumors. Clinical Cancer Research *17*, 2919-2933.

Hudlebusch, H. R., Skotte, J., Santoni-Rugiu, E., Zimling, Z. G., Lees, M. J., Simon, R., Sauter, G., Rota, R., De Ioris, M. A., Quarto, M.*, et al.* (2011b). MMSET is highly expressed and associated with aggressiveness in neuroblastoma. Cancer Research *71*, 4226-4235.

Hunger, S. P., Galili, N., Carroll, A. J., Crist, W. M., Link, M. P., and Cleary, M. L. (1991). The t(1;19)(q23;p13) results in consistent fusion of E2A and PBX1 coding sequences in acute lymphoblastic leukemias. Blood *77*, 687-693.

Inthal, A., Krapf, G., Beck, D., Joas, R., Kauer, M. O., Orel, L., Fuka, G., Mann, G., and Panzer-Grumayer, E. R. (2008). Role of the erythropoietin receptor in ETV6/RUNX1-positive acute lymphoblastic leukemia. Clin Cancer Res *14*, 7196-7204.

Jacoby, E., Chien, C. D., and Fry, T. J. (2014). Murine models of acute leukemia: important tools in current pediatric leukemia research. Frontiers in oncology *4*, 95.

Jaffe, J. D., Wang, Y., Chan, H. M., Zhang, J., Huether, R., Kryukov, G. V., Bhang, H.-e. C., Taylor, J. E., Hu, M., Englund, N. P.*, et al.* (2013). Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. Nature genetics *45*, 1386-1391.

Jasin, M. (1996). Genetic manipulation of genomes with rare-cutting endonucleases. Trends Genet *12*, 224-228.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941-945.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. Elife *2*, e00471.

Juárez-Velázquez, M. R., Salas-Labadía, C., Reyes-León, A., Navarrete-Meneses, M. P., Fuentes-Pananá, E. M., and Pérez-Vera, P. (2013). Genetic Markers in the Prognosis of Childhood Acute Lymphoblastic Leukemia. In Clinical Epidemiology of Acute Lymphoblastic Leukemia - From the Molecules to the Clinic, J.M. Mejia-Arangure, ed. (Rijeka: InTech), p. Ch. 09.

Juillerat, A., Dubois, G., Valton, J., Thomas, S., Stella, S., Marechal, A., Langevin, S., Benomari, N., Bertonati, C., Silva, G. H.*, et al.* (2014). Comprehensive analysis of the specificity of transcription activator-like effector nucleases. Nucleic Acids Res *42*, 5390-5402.

Kamps, M. P., and Baltimore, D. (1993). E2A-Pbx1, the t(1;19) translocation protein of human pre-B-cell acute lymphocytic leukemia, causes acute myeloid leukemia in mice. Mol Cell Biol *13*, 351-357.

Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A.*, et al.* (2013). Mutational landscape and significance across 12 major cancer types. Nature *502*, 333-339.

Keats, J. J., Maxwell, C. a., Taylor, B. J., Hendzel, M. J., Chesi, M., Bergsagel, P. L., Larratt, L. M., Mant, M. J., Reiman, T., Belch, A. R., and Pilarski, L. M. (2005). Overexpression of transcripts originating from the MMSET locus characterizes all t(4;14)(p16;q32)-positive multiple myeloma patients. Blood *105*, 4060-4069.

Keats, J. J., Reiman, T., Maxwell, C. A., Taylor, B. J., Larratt, L. M., Mant, M. J., Belch, A. R., and Pilarski, L. M. (2003). In multiple myeloma , t (4 ; 14)(p16 ; q32) is an adverse prognostic factor irrespective of FGFR3 expression. *101*, 1520-1529.

Kraushaar, D. C., Jin, W., Maunakea, A., Abraham, B., Ha, M., and Zhao, K. (2013). Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. Genome Biol *14*, R121.

Kruth, K. A., Fang, M., Shelton, D. N., Abu-Halawa, O., Mahling, R., Yang, H., Weissman, J. S., Loh, M. L., Muschen, M., Tasian, S. K.*, et al.* (2017). Suppression of B-cell development genes is key to glucocorticoid efficacy in treatment of acute lymphoblastic leukemia. Blood *129*, 3000-3008.

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A.*, et al.* (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res *44*, W90-97.

Kuo, A. J., Cheung, P., Chen, K., Zee, B. M., Kioi, M., Lauring, J., Xi, Y., Park, B. H., Shi, X., Garcia, B. A.*, et al.* (2011). NSD2 Links Dimethylation of Histone H3 at Lysine 36 to Oncogenic Programming. Molecular Cell *44*, 609-620.

Lafave, L. M., and Levine, R. L. (2013). Mining the epigenetic landscape in ALL. Nature genetics *45*, 1269-1270.

Lauring, J., Abukhdeir, A. M., Konishi, H., Garay, J. P., Gustin, J. P., Wang, Q., Arceci, R. J., Matsui, W., and Park, B. H. (2008). The multiple myeloma associated MMSET gene contributes to cellular adhesion, clonogenic growth, and tumorigenicity. Blood *111*, 856-864.

Lawrence, M. S., Stojanov, P., Mermel, C. H., Robinson, J. T., Garraway, L. A., Golub, T. R., Meyerson, M., Gabriel, S. B., Lander, E. S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495-501.

Lee, M. S. S., F. R. Jr.; Brooks III, C. L. (2002). Novel generalized Born methods. J Chem Phys *116*.

Lewis, P. W., Muller, M. M., Koletsky, M. S., Cordero, F., Lin, S., Banaszynski, L. A., Garcia, B. A., Muir, T. W., Becher, O. J., and Allis, C. D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science *340*, 857-861.

Li, Y., Trojer, P., Xu, C.-F., Cheung, P., Kuo, A., Drury, W. J., Qiao, Q., Neubert, T. a., Xu, R.-M., Gozani, O., and Reinberg, D. (2009). The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. The Journal of biological chemistry *284*, 34283-34295.

Lin, C. J., Conti, M., and Ramalho-Santos, M. (2013). Histone variant H3.3 maintains a decondensed chromatin state essential for mouse preimplantation development. Development *140*, 3624-3634.

Loh, M. L., Ma, X., Rusch, M., Wu, G., Harvey, R. C., Wheeler, D. A., Hampton, O. A., Carroll, W. L., Chen, I.-M., Gerhard, D. S*., et al.* (2013). Comparison Of Mutational Profiles Of Diagnosis and Relapsed Pediatric B-Acute Lymphoblastic Leukemia: A Report From The COG ALL Target Project. Blood *122*, 824.

Ma, X., Edmonson, M., Yergeau, D., Muzny, D. M., Hampton, O. a., Rusch, M., Song, G., Easton, J., Harvey, R. C., Wheeler, D. a*., et al.* (2015). Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. Nature communications *6*, 6604.

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S*., et al.* (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B *102*, 3586-3616.

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics *26*, 966-968.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Manis, J. P., Morales, J. C., Xia, Z., Kutok, J. L., Alt, F. W., and Carpenter, P. B. (2004). 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. Nat Immunol *5*, 481-487.

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

Martinez-Garcia, E., Popovic, R., Min, D. J., Sweet, S. M. M., Thomas, P. M., Zamdborg, L., Heffner, A., Will, C., Lamy, L., Staudt, L. M.*, et al.* (2011). The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t(4;14) multiple myeloma cells. Blood *117*, 211-220.

Matsuda, T., and Cepko, C. L. (2004). Electroporation and RNA interference in the rodent retina in vivo and in vitro. Proc Natl Acad Sci U S A *101*, 16-22.

McCabe, M. T., Ott, H. M., Ganji, G., Korenchuk, S., Thompson, C., Van Aller, G. S., Liu, Y., Graves, A. P., Della Pietra, A., 3rd, Diaz, E.*, et al.* (2012). EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. Nature *492*, 108-112.

McKinley, K. L., and Cheeseman, I. M. (2016). The molecular basis for centromere identity and function. Nat Rev Mol Cell Biol *17*, 16-29.

Mendez, J., and Stillman, B. (2000). Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. Mol Cell Biol *20*, 8602-8612.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Milojkovic, D., and Apperley, J. (2009). Mechanisms of Resistance to Imatinib and Second-Generation Tyrosine Inhibitors in Chronic Myeloid Leukemia. Clin Cancer Res *15*, 7519-7527.

Min, D.-J., Ezponda, T., Kim, M. K., Will, C. M., Martinez-Garcia, E., Popovic, R., Basrur, V., Elenitoba-Johnson, K. S., and Licht, J. D. (2012). MMSET stimulates myeloma cell growth through microRNA-mediated modulation of c-MYC. Leukemia *27*, 686-694.

Moorman, A. V., Harrison, C. J., Buck, G. A., Richards, S. M., Secker-Walker, L. M., Martineau, M., Vance, G. H., Cherry, A. M., Higgins, R. R., Fielding, A. K.*, et al.* (2007). Karyotype is an independent prognostic factor in adult acute lymphoblastic leukemia (ALL): analysis of cytogenetic data from patients treated on the Medical Research Council (MRC) UKALLXII/Eastern Cooperative Oncology Group (ECOG) 2993 trial. Blood *109*, 3189-3197.

Morishita, M., Mevius, D., and di Luccio, E. (2014). In vitro histone lysine methylation by NSD1, NSD2/MMSET/WHSC1 and NSD3/WHSC1L. BMC structural biology *14*, 25.

Morrow, M., Horton, S., Kioussis, D., Brady, H. J., and Williams, O. (2004). TEL-AML1 promotes development of specific hematopoietic lineages consistent with preleukemic activity. Blood *103*, 3890-3896.

Mueller, F., Karpova, T. S., Mazza, D., and McNally, J. G. (2012). Monitoring dynamic binding of chromatin proteins in vivo by fluorescence recovery after photobleaching. Methods Mol Biol *833*, 153-176.

Mullighan, C. G. (2013). Genome sequencing of lymphoid malignancies. Blood *122*, 3899-3907.

Mullighan, C. G., Zhang, J., Kasper, L. H., Lerach, S., Payne-Turner, D., Phillips, L. a., Heatley, S. L., Holmfeldt, L., Collins-Underwood, J. R., Ma, J.*, et al.* (2011). CREBBP mutations in relapsed acute lymphoblastic leukaemia. Nature *471*, 235-239.

Neri, L. M., Cani, A., Martelli, A. M., Simioni, C., Junghanss, C., Tabellini, G., Ricci, F., Tazzari, P. L., Pagliaro, P., McCubrey, J. A., and Capitani, S. (2014). Targeting the PI3K/Akt/mTOR signaling pathway in B-precursor acute lymphoblastic leukemia and its therapeutic potential. Leukemia *28*, 739-748.

Nimura, K., Ura, K., Shiratori, H., Ikawa, M., Okabe, M., Schwartz, R. J., and Kaneda, Y. (2009). A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf–Hirschhorn syndrome. Nature *460*, 287-291.

Overmiller, A. M., Pierluissi, J. A., Wermuth, P. J., Sauma, S., Martinez-Outschoorn, U., Tuluc, M., Luginbuhl, A., Curry, J., Harshyne, L. A., Wahl, J. K., 3rd*, et al.* (2017). Desmoglein 2 modulates extracellular vesicle release from squamous cell carcinoma keratinocytes. FASEB J *31*, 3412-3424.

Oyer, J. a., Huang, X., Zheng, Y., Shim, J., Ezponda, T., Carpenter, Z., Allegretta, M., Okot-Kotber, C. I., Patel, J. P., Melnick, a.*, et al.* (2014). Point mutation E1099K in MMSET/NSD2 enhances its methyltranferase activity and leads to altered global chromatin methylation in lymphoid malignancies. Leukemia *28*, 198-201.

Palomero, T., Sulis, M. L., Cortina, M., Real, P. J., Barnes, K., Ciofani, M., Caparros, E., Buteau, J., Brown, K., Perkins, S. L.*, et al.* (2007). Mutational loss of PTEN induces resistance to NOTCH1 inhibition in T-cell leukemia. Nat Med *13*, 1203-1210.

Pei, H., Wu, X., Liu, T., Yu, K., Jelinek, D. F., and Lou, Z. (2013). The Histone Methyltransferase MMSET Regulates Class Switch Recombination. The Journal of Immunology *190*, 756-763.

Pei, H., Zhang, L., Luo, K., Qin, Y., Chesi, M., Fei, F., Bergsagel, P. L., Wang, L., You, Z., and Lou, Z. (2011). MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. Nature *470*, 124-128.

Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., and Weng, Z. (2014). ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. Bioinformatics *30*, 1771-1773.

Pikman, Y., Lee, B. H., Mercher, T., McDowell, E., Ebert, B. L., Gozo, M., Cuker, A., Wernig, G., Moore, S., Galinsky, I.*, et al.* (2006). MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. PLoS Med *3*, e270.

Piovan, E., Yu, J., Tosello, V., Herranz, D., Ambesi-Impiombato, A., Da Silva, A. C., Sanchez-Martin, M., Perez-Garcia, A., Rigo, I., Castillo, M.*, et al.* (2013). Direct reversal of glucocorticoid resistance by AKT inhibition in acute lymphoblastic leukemia. Cancer Cell *24*, 766-776.

Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E.*, et al.* (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517-527.

Popovic, R., and Licht, J. D. (2012). Emerging epigenetic targets and therapies in cancer medicine. Cancer discovery *2*, 405-413.

Popovic, R., Martinez-Garcia, E., Giannopoulou, E. G., Zhang, Q., Zhang, Q., Ezponda, T., Shah, M. Y., Zheng, Y., Will, C. M., Small, E. C.*, et al.* (2014). Histone Methyltransferase MMSET/NSD2 Alters EZH2 Binding and Reprograms the Myeloma Epigenome through Global and Focal Changes in H3K36 and H3K27 Methylation. PLoS Genetics *10*.

Poulin, M. B., Schneck, J. L., Matico, R. E., Hou, W., McDevitt, P. J., Holbert, M., and Schramm, V. L. (2016a). Nucleosome Binding Alters the Substrate Bonding Environment of Histone H3 Lysine 36 Methyltransferase NSD2. J Am Chem Soc *138*, 6699-6702.

Poulin, M. B., Schneck, J. L., Matico, R. E., McDevitt, P. J., Huddleston, M. J., Hou, W., Johnson, N. W., Thrall, S. H., Meek, T. D., and Schramm, V. L. (2016b). Transition state for the NSD2-catalyzed methylation of histone H3 lysine 36. Proc Natl Acad Sci U S A *113*, 1197-1201.

Qiao, Q., Li, Y., Chen, Z., Wang, M., Reinberg, D., and Xu, R. M. (2011). The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. J Biol Chem *286*, 8361-8368.

Quentmeier, H., Eberth, S., Romani, J., Zaborski, M., and Drexler, H. G. (2011). BCR-ABL1-independent PI3Kinase activation causing imatinib-resistance. J Hematol Oncol *4*, 6.

Ramirez, C. L., Foley, J. E., Wright, D. A., Muller-Lerch, F., Rahman, S. H., Cornu, T. I., Winfrey, R. J., Sander, J. D., Fu, F., Townsend, J. A.*, et al.* (2008). Unexpected failure rates for modular assembly of engineered zinc fingers. Nat Methods *5*, 374-375.

Ravandi, F., Cortes, J., Estrov, Z., Thomas, D., Giles, F. J., Huh, Y. O., Pierce, S., O'Brien, S., Faderl, S., and Kantarjian, H. M. (2002). CD56 expression predicts occurrence of CNS disease in acute lymphoblastic leukemia. Leuk Res *26*, 643-649.

Ray-Gallet, D., Woolfe, A., Vassias, I., Pellentz, C., Lacoste, N., Puri, A., Schultz, D. C., Pchelintsev, N. A., Adams, P. D., Jansen, L. E., and Almouzni, G. (2011). Dynamics of histone H3 deposition in vivo reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. Mol Cell *44*, 928-941.

Real, P. J., and Ferrando, A. A. (2009). NOTCH inhibition and glucocorticoid therapy in T-cell acute lymphoblastic leukemia. Leukemia *23*, 1374-1377.

Roberts, K. G., and Mullighan, C. G. (2015). Genomics in acute lymphoblastic leukaemia: insights and treatment implications. Nat Rev Clin Oncol *12*, 344-357.

Romana, S. P., Mauchauffe, M., Le Coniat, M., Chumakov, I., Le Paslier, D., Berger, R., and Bernard, O. A. (1995). The t(12;21) of acute lymphoblastic leukemia results in a tel-AML1 gene fusion. Blood *85*, 3662-3670.

Ryckaert, J.-P. C., G.; Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys *23*, 327-341.

Salesse, S., and Verfaillie, C. M. (2002). BCR/ABL: from molecular mechanisms of leukemia induction to treatment of chronic myelogenous leukemia. Oncogene *21*, 8547-8559.

Saloura, V., Cho, H.-S., Kiyotani, K., Alachkar, H., Zuo, Z., Nakakido, M., Tsunoda, T., Seiwert, T., Lingen, M., Licht, J*., et al.* (2015). WHSC1 promotes oncogenesis through regulation of NIMA-related kinase-7 in squamous cell carcinoma of the head and neck. Molecular cancer research : MCR *13*, 293-304.

Sankaran, S. M., Wilkinson, A. W., Elias, J. E., and Gozani, O. (2016). A PWWP Domain of Histone-Lysine N-Methyltransferase NSD2 Binds to Dimethylated Lys-36 of Histone H3 and Regulates NSD2 Function at Chromatin. J Biol Chem *291*, 8465-8474.

Santra, M., Zhan, F., Tian, E., Barlogie, B., and Jr, J. S. (2003). Brief report A subset of multiple myeloma harboring the t (4 ; 14)(p16 ; q32) translocation lacks FGFR3 expression but maintains an IGH / MMSET fusion transcript. *101*, 2374-2376.

Sarai, N., Nimura, K., Tamura, T., Kanno, T., Patel, M. C., Heightman, T. D., Ura, K., and Ozato, K. (2013). WHSC1 links transcription elongation to HIRA-mediated histone H3.3 deposition. The EMBO journal *32*, 2392-2406.

Schindler, J. W., Van Buren, D., Foudi, A., Krejci, O., Qin, J., Orkin, S. H., and Hock, H. (2009). TEL-AML1 corrupts hematopoietic stem cells to persist in the bone marrow and initiate leukemia. Cell Stem Cell *5*, 43-53.

Schwartzentruber, J., Korshunov, A., Liu, X. Y., Jones, D. T., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A. M., Quang, D. A., Tonjes, M*., et al.* (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature *482*, 226-231.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res *31*, 3381-3385.

Shah, M. a., Denton, E. L., Arrowsmith, C. H., Lupien, M., and Schapira, M. (2014). A global assessment of cancer genomic alterations in epigenetic mechanisms. Epigenetics & chromatin *7*, 29.

Shah, M. Y., Martinez-Garcia, E., Phillip, J. M., Chambliss, A. B., Popovic, R., Ezponda, T., Small, E. C., Will, C., Phillip, M. P., Neri, P.*, et al.* (2016). MMSET/WHSC1 enhances DNA damage repair leading to an increase in resistance to chemotherapeutic agents. Oncogene.

Shi, J., Wang, E., Milazzo, J. P., Wang, Z., Kinney, J. B., and Vakoc, C. R. (2015). Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. Nature Biotechnology *33*.

Shurtleff, S. A., Buijs, A., Behm, F. G., Rubnitz, J. E., Raimondi, S. C., Hancock, M. L., Chan, G. C., Pui, C. H., Grosveld, G., and Downing, J. R. (1995). TEL/AML1 fusion resulting from a cryptic t(12;21) is the most common genetic lesion in pediatric ALL and defines a subgroup of patients with an excellent prognosis. Leukemia *9*, 1985-1989.

Silva, A., Girio, A., Cebola, I., Santos, C. I., Antunes, F., and Barata, J. T. (2011). Intracellular reactive oxygen species are essential for PI3K/Akt/mTOR-dependent IL-7-mediated viability of T-cell acute lymphoblastic leukemia cells. Leukemia *25*, 960-967.

Silva, A., Yunes, J. A., Cardoso, B. A., Martins, L. R., Jotta, P. Y., Abecasis, M., Nowill, A. E., Leslie, N. R., Cardoso, A. A., and Barata, J. T. (2008). PTEN posttranslational inactivation and hyperactivation of the PI3K/Akt pathway sustain primary T cell leukemia viability. J Clin Invest *118*, 3762-3774.

Skorski, T., Bellacosa, A., Nieborowska-Skorska, M., Majewski, M., Martinez, R., Choi, J. K., Trotta, R., Wlodarski, P., Perrotti, D., Chan, T. O.*, et al.* (1997). Transformation of hematopoietic cells by BCR/ABL requires activation of a PI-3k/Akt-dependent pathway. EMBO J *16*, 6151-6161.

Society, A. C. (2014). Cancer Facts & Figures 2014. Atlanta: American Cancer Society.

Society, A. C. (2016). Cancer Facts & Figures 2016. Atlanta: American Cancer Society.

Soria, G., Polo, S. E., and Almouzni, G. (2012). Prime, repair, restore: the active role of chromatin in the DNA damage response. Mol Cell *46*, 722-734.

Stavnezer, J., Guikema, J. E., and Schrader, C. E. (2008). Mechanism and regulation of class switch recombination. Annu Rev Immunol *26*, 261-292.

Stec, I., Wright, T. J., van Ommen, G. J., de Boer, P. a., van Haeringen, a., Moorman, a. F., Altherr, M. R., and den Dunnen, J. T. (1998). WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphy gene maps in the

Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. Human molecular genetics *7*, 1071-1082.

Subramanian, A., Kuehn, H., Gould, J., Tamayo, P., and Mesirov, J. P. (2007). GSEA-P: a desktop application for Gene Set Enrichment Analysis. Bioinformatics *23*, 3251-3253.

Suvà, M. L., Riggi, N., and Bernstein, B. E. (2013). Epigenetic reprogramming in cancer. Science (New York, NY) *339*, 1567-1570.

Szenker, E., Ray-Gallet, D., and Almouzni, G. (2011). The double face of the histone variant H3.3. Cell Research *21*, 421-434.

Tasian, S. K., Teachey, D. T., Li, Y., Shen, F., Harvey, R. C., Chen, I. M., Ryan, T., Vincent, T. L., Willman, C. L., Perl, A. E.*, et al.* (2017). Potent efficacy of combined PI3K/mTOR and JAK or ABL inhibition in murine xenograft models of Ph-like acute lymphoblastic leukemia. Blood *129*, 177-187.

Tisi, D., Chiarparin, E., Tamanini, E., Pathuri, P., Coyle, J. E., Hold, A., Holding, F. P., Amin, N., Martin, A. C., Rich, S. J.*, et al.* (2016). Structure of the Epigenetic Oncogene MMSET and Inhibition by N-Alkyl Sinefungin Derivatives. ACS Chem Biol *11*, 3093-3105.

Tvardovskiy, A., Schwammle, V., Kempf, S. J., Rogowska-Wrzesinska, A., and Jensen, O. N. (2017). Accumulation of histone variant H3.3 with age is associated with profound changes in the histone methylation landscape. Nucleic Acids Res *45*, 9272-9289.

Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999). Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol *285*, 1735-1747.

Wu, G., Broniscer, A., McEachron, T. A., Lu, C., Paugh, B. S., Becksfort, J., Qu, C., Ding, L., Huether, R., Parker, M.*, et al.* (2012). Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat Genet *44*, 251-253.

Wynn, R. F. (2010). Acute Lymphoblastic Leukemia. In Pediatric Hematology and Oncology, (Wiley-Blackwell), pp. 75-94.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J Biol Chem *286*, 7983-7989.

Yun, M., Wu, J., Workman, J. L., and Li, B. (2011). Readers of histone modifications. Cell Res *21*, 564-578.

Zelent, A., Greaves, M., and Enver, T. (2004). Role of the TEL-AML1 fusion gene in the molecular pathogenesis of childhood acute lymphoblastic leukaemia. Oncogene *23*, 4275-4283.

Zhang, Y., Ge, X., Yang, F., Zhang, L., Zheng, J., Tan, X., Jin, Z. B., Qu, J., and Gu, F. (2014). Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. Sci Rep *4*, 5405.

Zheng, Y., Sweet, S. M. M., Popovic, R., Martinez-Garcia, E., Tipton, J. D., Thomas, P. M., Licht, J. D., and Kelleher, N. L. (2012). Total kinetic analysis reveals how combinatorial methylation patterns are established on lysines 27 and 36 of histone H3. Proceedings of the National Academy of Sciences of the United States of America *109*, 13549-13554.

Zheng, Y., Tipton, J. D., Thomas, P. M., Kelleher, N. L., and Sweet, S. M. (2014). Site-specific human histone H3 methylation stability: fast K4me3 turnover. Proteomics *14*, 2190-2199.

Zuckerman, T., and Rowe, J. M. (2014). Pathogenesis and prognostication in acute lymphoblastic leukemia. F1000Prime Rep *6*, 59.

Vita

PERSONAL INFORMATION

Name: Alok Swaroop
Date of Birth: August 19, 1989
Place of Birth: New Haven, CT

EDUCATION

2011-present	**Northwestern University, Feinberg School of Medicine** Chicago, IL M.D., anticipated May 2019
2011-present	**Northwestern University, Feinberg School of Medicine** Driskill Graduate Program in Life Sciences Chicago, IL Ph.D., anticipated May 2019
2007-2011	**University of Michigan** College of Literature, Science, and the Arts Ann Arbor, MI B.S. in Cellular and Molecular Biology, minor in Classical Civilization

HONORS

2016-present	National Institutes of Health - Ruth L. Kirschstein National Research Service Award - F30CA203292
2014-2015	T32 Carcinogenesis Training Program Trainee
2010	American Society of Hematology Trainee Research Award
2008	American Society of Hematology Trainee Research Award

RESEARCH EXPERIENCE

2013-2017	**Ph.D. student researcher** Northwestern University, Feinberg School of Medicine Department of Medicine University of Florida Health Cancer Center Department of Medicine Mentor: Jonathan Licht, M.D. Thesis title: An Activating Mutation of the NSD2 Histone Methyltransferase Drives Oncogenic Reprogramming in Acute Lymphocytic Leukemia

Role: Primary researcher using cell lines and mouse models to study an oncogenic mutation commonly found in leukemia and lymphoma.

2009-2011 **Student researcher**
University of Michigan Medical School
Mentor: Jordan Shavit, M.D./Ph.D.
Modeled blood clotting disorders in zebrafish using morpholinos and developed a lethal model of thrombosis using zinc finger nucleases.

2009 **Summer intern in biomedical research**
National Institutes of Health – NIDDK
Mentor: Douglas Forrest, Ph.D.
Combined analysis directly tracking photoreceptor generation using immunofluorescence to monitor Nrl and TRβ2 expression.

2007-2009 **Student researcher**
University of Michigan Medical School
Mentors: David Ginsburg, M.D., Jordan Shavit, M.D./Ph.D.
Modeled gamma carboxylase deficiency and other blood clotting disorders in zebrafish using morpholinos.

2006-2007 **Student researcher**
University of Michigan – Department of Molecular, Cell, and Developmental Biology
Mentor: Pamela Raymond, Ph.D.
Explored genes involved in retinal development and cone photoreceptor patterning in zebrafish using immunohistochemistry techniques.

PUBLICATIONS

Vo, A. H.*, **Swaroop, A.***, Liu, Y., Norris, Z.G., & Shavit, J.A. (2013). Loss of Fibrinogen in Zebrafish Results in Symptoms Consistent with Human Hypofibrinogenemia. *PloS One*, 8(9), e74682. doi:10.1371/journal.pone.0074682. **[*-authorship equally shared]**

Ng, L., Lu, A., **Swaroop, A.**, Sharlin, D.S., Swaroop, A., & Forrest, D. (2011). Two transcription factors can direct three photoreceptor outcomes from rod precursor cells in mouse retinal development. *Journal of Neuroscience*, 31(31), 11118–11125.

Bennett, R.L., **Swaroop, A.,** Troche, C., Licht, J.D. (2018). The Role of Nuclear Receptor-Binding SET Domain Family Histone Lysine Methyltransferases in Cancer. *Cold Spring Harb Perspect Med*, ;7(6), a026708. doi: 10.1101/cshperspect.a026708

Swaroop, A.*, Oyer, J.A.*, Will C.M., Huang X., Yu, W., Troche, C., Bulic, M., Durham, B.H., Wen, Q.J., Crispino, J.D., MacKerell, A.D., Bennett, R.L., Kelleher, N.L., Licht, J.D. (2018). An activating mutation of the NSD2 histone methyltransferase drives oncogenic reprogramming in acute lymphocytic leukemia. *Oncogene* **(in press)**. **[*-authorship equally shared]**

ABSTRACTS/POSTERS

Swaroop, A., Oyer, J.A., Will C.M., Popovic, R, Licht, J.D. 2014. Generation of cellular model to study to *MMSET E1099K* mutant histone methyltransferase mediated chromatin dysregulation using CRISPR genomic editing. Wellcome Trust Advanced Course: Chromatin Structure and Function.

Swaroop, A., Oyer, J.A., Will C.M., Bulic, M., Wen, Q.J., Crispino, J.D., Licht, J.D. 2016. Point Mutation E1099K in *MMSET* Disrupts Balance of Histone Methylation and Drives Widespread Carcinogenic Reprogramming in Acute Lymphocytic Leukemia. University of Florida Department of Medicine Celebration of Research.

Swaroop, A., Oyer, J.A., Will C.M., Bulic, M., Wen, Q.J., Crispino, J.D., Bennett, R.L., Licht, J.D. 2017. An activating mutation of the *NSD2* histone methyltransferase drives oncogenic reprogramming in acute lymphocytic leukemia. AAP-ASCI-APSA Joint Meeting.

Swaroop, A., Oyer, J.A., Will C.M., Bulic, M., Wen, Q.J., Crispino, J.D., Bennett, R.L., Licht, J.D. 2017. An activating mutation of the *NSD2* histone methyltransferase drives oncogenic reprogramming in acute lymphocytic leukemia. University of Florida Department of Medicine Celebration of Research.

EXTRACURRICULAR ACTIVITIES

2011-present	**Medical Scientist Training Program Student Council** Northwestern University, Feinberg School of Medicine Vice President, 2014
2016	**Data Sharing Working Group** Physical Sciences-Oncology Network – National Cancer Institute Chicago Region PSOC representative Worked with scientists across the country to come up with ideas to facilitate exchange of data and resources between different programs

MEMBERSHIPS

2008 & 2010	American Society of Hematology
2011-present	American Physician Scientist Association
2013-present	American Medical Student Association
2014-2015	American Association for Cancer Research
2018-present	American Medical Association