Washington University in St. Louis
Washington University Open Scholarship

Arts & Sciences Electronic Theses and Dissertations

Arts & Sciences

5-3-2024

Integrating DNA Methylation and 3D-genome Architecture to Identify Functional Regulatory Sequences in IDH-mutant AML

Elisabeth Wilson
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Recommended Citation
Wilson, Elisabeth, "Integrating DNA Methylation and 3D-genome Architecture to Identify Functional Regulatory Sequences in IDH-mutant AML" (2024). *Arts & Sciences Electronic Theses and Dissertations*. 3072.
https://openscholarship.wustl.edu/art_sci_etds/3072

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Molecular Genetics and Genomics

Dissertation Examination Committee:
David Spencer, Chair
John Edwards
Timothy Ley
Jacqueline Payton
Matthew Walter

Integrating DNA Methylation and 3D-genome Architecture to Identify Functional Regulatory Sequences in *IDH*-mutant AML

by
Elisabeth R. Wilson

A dissertation presented to
Washington University in St. Louis
in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2024
St. Louis, Missouri

© 2024, Elisabeth R. Wilson

Table of Contents

List of Figures

Acknowledgments

It wasn't easy but it's done. I feel incredibly fortunate to have worked and trained with brilliant and talented scientists during my time at Washington University in St. Louis. It is largely due to the mentoring I've received from my DBBS program directors James Skeath, Tim Schedl, and John Edwards, the training from faculty in the Division of Oncology, the guidance of my thesis committee, and of course, the patient and unwavering support of my mentor, Dr. David Spencer, that I move forward with confidence in my ability to tackle scientific problems with courage and competence. I also acknowledge my family and close friends for their constant reassurance that I was (and continue to be) good enough to be here and do the hard work amidst giants and legends. I will be forever grateful to my partner, Matthew, who never stopped believing in me and who carried so much of the load during this process, making it possible for me to focus my time and energy on the science and ultimately achieve my goal. To Olive and Beau, thank you for providing unconditional love and comfort every day of the past six years—I would not be the same human without you. I also offer a special acknowledgment for my mother, Jolene, who not only prioritized resources so my siblings and I could pursue higher education, but who also continues to be my example of what grit, determination, and willpower can earn you in life.

In the wise words of my mentor's mentor:

"A Ph.D. is the beginning, not the end."
-Maynard Olson

Elisabeth Wilson

Washington University in St. Louis

May 2024

Dedicated to my family.

ABSTRACT OF THE DISSERTATION

Integrating DNA Methylation and 3D-genome Architecture to Identify Functional Regulatory Sequences in *IDH*-mutant AML

by

Elisabeth R. Wilson

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular Genetics and Genomics

Washington University in St. Louis, 2024

Associate Professor David Spencer, Chair

Acute myeloid leukemia (AML) is a genetically and phenotypically heterogeneous disease. Recurrent mutations in genes involved in epigenetic pathways are common in AML and are thought to contribute to this variability by disrupting epigenetic patterns relative to normal hematopoietic cells. The most well-studied epigenetic modification in AML is DNA methylation. Although methods for studying DNA methylation have improved significantly in the past decade, the landscape of methylation changes in AML and how they relate to chromatin architecture and gene regulation is still not fully understood.

This thesis is focused on the DNA methylation phenotype of *IDH1* or *IDH2* mutations in AML cells and the functional consequences of these changes. *IDH1* or *IDH2* mutations occur in approximately 20% of AML patients and are associated with altered DNA methylation patterns. The *IDH1* and *IDH2* genes encode metabolic enzymes that are not normally involved in DNA methylation. However, the mutant forms of IDH1 and IDH2 in AML produce 2-hydroxyglutarate (2HG) that inhibits the TET family of enzymes, which act to remove methylated cytosines on DNA through successive oxidation reactions. Hence, the net effect of *IDH* mutations is an

increase in DNA methylation by inhibiting its removal. Although the specific consequences of this hypermethylation phenotype have been difficult to define, *in vivo* mouse models have established that *IDH* mutations contribute to AML development.

We used a comprehensive set of whole-genome bisulfite sequencing data from primary AML to characterize AML-associated DNA methylation phenotypes and identify *IDH* mutation-specific methylation changes in primary AML samples. Such an approach had not yet been taken to understand the direct consequences of *IDH* mutations, and our study identified a unique hypermethylation signature that is enriched for active enhancer regions in *IDH*-mutant AML. The differentially methylated regions (DMRs) from this analysis were characterized by active methylation and demethylation turnover, evidenced by an increase in TET-mediated 5-hydroxymethylation, which may account for the unique susceptibility of enhancers to become hypermethylated in the presence of *IDH* mutations.

We then used high-resolution chromatin conformation data to link these IDH^{mut}-specific hypermethylated enhancer DMRs with their cognate gene(s), which we found to be highly expressed in hematopoietic cells. The interactions often coincided with CTCF binding sites that orchestrated chromatin loop formation and formed insulating TAD boundaries. We observed some evidence of disruption of loop formation and boundary insulation in *IDH* mutant samples compared with CD34+ normal hematopoietic cells, which correlated with decreased target gene expression in a subset of interactions. These results suggest that hypermethylation accumulates at regions that are important for genome organization and in certain cases may affect gene expression in favor of leukemia development.

Mechanistic studies of candidate loci were performed to better define the nature of enhancer DMR regulatory interactions. Results from targeted deletion of the DMR in a well-characterized *MYC* enhancer were consistent with our hypothesis of inherent regulatory activity associated with DMR regions, demonstrating a trend towards decreased *MYC* expression in DMR KO cell lines compared with WT cells.

Chapter 1: Introduction

1.1 Acute Myeloid Leukemia

1.1.1 Introduction

Acute myeloid leukemia (AML) is an aggressive disease that affects ~20,000 adults and children annually and is typically associated with poor prognosis[1–3]. AML is characterized by abnormal growth of myeloid cells which rapidly divide and accumulate in the bone marrow, ultimately interfering with normal blood cell development. This process can lead to symptoms such as fatigue, infections, easy bleeding/bruising, and anemia, which left unmanaged/treated, can quickly progress to death. AML is relatively rare, accounting for <2% of all cancer diagnoses each year, however, it's the second most common form of leukemia, representing ~30% of all adult cases (>19 y)[3]. Although survival is variable and often highly dependent on age, disease subtype, cytogenetic and molecular abnormalities, and treatment response, patients with AML have an average 5-year survival rate of ~28% compared with nearly 70 to 80% for patients with lymphoid malignancies[4]. Accordingly, significant development in our understanding and treatment of the disease is necessary to improve the overall health and survival of AML patients.

The greatest challenge to finding effective therapy for AML is reflected in the genetic and phenotypic heterogeneity of the disease. Since the 1960s it has been understood that AML is a clonal disease that results from both inherited and acquired genetic lesions, but the full extent of genetic diversity amongst patients has only been made clear in the last 10-15 years by extensive next-generation sequencing studies[1,5]. Early characterizations of primary AML cells detected a variety of recurrent chromosomal abnormalities by cytogenetic karyotyping including large-scale

inversions and translocations. Detection of these cytogenetic events served as the earliest disease classification and prognostic scoring system for AML and, with some refinement, continues to be the most powerful predictor of patient outcome when present. However, a significant number of AML patients lack gross chromosomal abnormalities and have historically been difficult to stratify and develop tailored treatments for. With the advent and evolution of sequencing technology, and namely, the ability to detect genetic alterations at base pair resolution, our understanding of disease heterogeneity and pathophysiology has grown tremendously. Notably, AML sub-classification, patient stratification, and disease management have been improved by testing for recurrent genetic mutations, particularly when considering those patients with otherwise normal karyotypes[1,3–9].

Despite advances in characterizing the genetic landscape of AML, treatment hasn't changed significantly since the 1960s, when the first uses of chemotherapy shifted the needle on patient survival and established the backbone of frontline therapy used by clinicians for the past 40 years. Refinements in standard therapy, including treatment intensification, better supportive care, and introduction of bone marrow transplantation have improved outcomes in younger patients, however, mortality rates remain high in the elderly who typically have lower tolerance for aggressive treatment strategies. Targeted therapies, designed to address the root genetic cause of disease, are one approach to limit toxicity in patients but have shown modest efficacy to date. Even still, precision medicine continues to be a goal in the treatment of AML, and the development of new agents targeting biochemical pathways implicated in disease pathophysiology as well as immunotherapeutic strategies poised to harness the patient's own immune system to recognize and eliminate malignant cells offer promising routes forward.

1.1.2 Diagnosis and Prognosis

The diagnosis of AML involves a combination of clinical presentations and laboratory tests. Patients often present with symptoms of bone marrow failure including fever, infection, weakness/fatigue, enlarged spleen and/or lymph nodes, etc. Commonly, patients will have an initial complete blood count (CBC) and differential workup to assess for significant changes in mature blood cell populations (e.g. anemia, thrombocytopenia, neutropenia) as well as the presence of excess immature myeloid cells, or "blasts", in the peripheral blood [2,10]. Concerning results, or counts outside of the normal range, often indicate bone marrow biopsy for subsequent morphologic, cytogenetic, immunophenotypic, and molecular testing. AML diagnosis is classically established by the presence of >20% blasts in the bone marrow or peripheral blood by morphology (recently updated to >10% in multiple genetic contexts)[10,11], while cell lineage and mutational subclassifications are assessed by immunophenotypic and molecular testing. Consideration of the patient's clinical features as well as the results of these critical diagnostic tests help clinicians predict the likely course and outcome of the disease and guide therapeutic decisions.

Age of diagnosis and overall health are important clinical factors that dictate prognosis and treatment approaches in AML patients. Age has proven to be a consistent determinant of outcome, as older patients (>70 y) historically experience the highest incidence of comorbidities, treatment-related toxicity, and poor responsiveness to intensive chemotherapy[11,12]. Perhaps not surprisingly, older patients also have a high frequency of adverse AML-associated factors including complex karyotypes, unfavorable mutations, and increased incidence of therapy-related or secondary AML. Accordingly, old age is considered an independent predictor of

adverse outcome and indicates low-intensity therapy (often in combination with targeted agents such as venetoclax) for the best chance of long-term remission[10].

As previously mentioned, cytogenetics has been a long-standing cornerstone in AML prognosis. The well-adopted European LeukemiaNet (ELN) classification system stratifies patients into three risk categories (favorable, intermediate, adverse) based on karyotypic features. Patients with chromosomal abnormalities such as t(8;21)(q22;q22), inv(16)(p13.1q22) or t(16;16)(p13.1;q22), and t(15;17)(q22;q12) generally have a better response to standard chemotherapy and a more favorable prognosis, while patients with complex karyotypes (three or more chromosomal abnormalities), -5 or del(5q), -7 or abnormal(17p) tend to have poor response to standard chemotherapy and are categorized by adverse risk, or worse overall prognosis[10,13]. The intermediate risk group represents a heterogenous population of patients including those with normal karyotypes and other non-defined abnormalities. In current practice, cytogenetics reliably predicts outcomes when patients have favorable or adverse karyotypes. For example, patients with favorable cytogenetics including inv(16) or t(8;21) translocations have an increased 5-year survival rate of 70% while individuals harboring unfavorable cytogenetic abnormalities such as monosomy 7 or complex karyotype have a dismal 5-year survival rate of 10%[3,10,13]. However, cytogenetics alone fails to capture the diversity of outcomes for intermediate-risk patients, where 5-year overall survival ranges from 25-50%[10].

The detection of gene mutations has helped further refine risk stratification groups in conjunction with cytogenetic events. Internal tandem duplications in the *FLT3* gene (*FLT3-ITD*) are associated with a higher risk of relapse and poorer overall survival, especially when the allelic ratio is high[1,7,14]. However, significant developments in targeted therapy for FLT3-ITD is improving outcome for these patients, classifying them as intermediate risk. Mutations in the

NPM1 gene, in the absence of *FLT3-ITD* or with low-allelic ratio *FLT3-ITD*, are associated with a favorable prognosis. Biallelic mutations in *CEBPA* are associated with a favorable prognosis, while mutations in *TP53*, *RUNX1*, and *ASXL1* genes (amongst others) are generally associated with an adverse prognosis[5,10,13].

Response to treatment has become an important prognostic factor dictating patient survival. Patients who achieve complete remission (CR) after induction therapy, defined as blast counts below 5% and normalization of other blood cell counts, typically have better outcomes than those who do not respond to treatment and/or relapse shortly after achieving remission[9,15]. However, the most recent updates to the ELN classification system reflect the importance of early measurable residual disease (MRD) monitoring in addition to assessing for CR[10]. Established techniques, such as multiparametric flow cytometry and RT-qPCR are currently used to detect MRD, or small numbers of leukemia cells remaining after treatment, and provide insights into the risk of relapse and the need for more intensive therapy or stem cell transplantation.

1.1.3 Treatment strategies

Standard treatment approaches for AML patients have historically revolved around the use of chemotherapy with the goal for patients to achieve complete remission after induction therapy while limiting therapy-related toxicities. Early randomized clinical trials laid a foundation for chemotherapy dosing regimens and schedules aimed at achieving this goal. Standard of care for multiple decades has followed a '3+7' regimen where patients receive 3 consecutive doses of an anthracycline, followed by 7 consecutive high-dose treatments with cytarabine[16]. Subsequent consolidation treatment, typically involving additional rounds of intensive cytarabine chemotherapy, is given to reduce the risk of relapse and prepare the patient for stem cell

transplant when appropriate. Over the past few decades, additional randomized clinical trials have helped refine and tailor chemotherapy regimens to reflect the clinical and genetic heterogeneity of AML[16,17]. Broadly speaking, current front-line induction chemotherapy is administered as a combination of either daunorubicin or idarubicin anthracyclines and high-dose cytarabine, with the addition of nucleoside analogs, CD33 targeted monoclonal antibodies, and other targeted agents when indicated[17].

The choice to administer intensive chemotherapy, often in combination with other agents, is highly nuanced and dependent on several factors including the patient's age/fitness, cytogenetic and molecular markers, and predicted CR rate/long-term outcome given the available therapies. For example, older patients (typically >70 y) have historically experienced high treatment-related mortality leading to general de-intensification of chemotherapy and increased use of alternative drugs including hypomethylating agents[10]. After considering a patient's age, the detection of certain cytogenetic abnormalities can be directive for treatment. Acute promyelocytic leukemia (APL) and core binding factor (CBF) AML, are two subsets that have near-curative regimens in place. While both subsets were initially treated with standard chemotherapy with low efficacy, studies in the early 2000s demonstrated cure rates >90% for APL patients treated with a nonchemotherapy regimen of ATRA and arsenic trioxide and 80-90% for CFB AML patients treated with fludarabine anthracycline plus cytarabine with the addition of gemtuzumab ozogamicin[18]. Most AML patients, however, require improvements to current chemotherapy combinations and dosing schedules as well as better-adapted agents that target molecular and epigenetic vulnerabilities of their individual leukemias to reach similar rates of long-term survival.

Developments in our understanding of the genetic landscape of AML have led to new targeted therapies offering more personalized treatment approaches. In certain cases, specific cytogenetic or molecular abnormalities qualify patients for targeted therapies in addition to or instead of standard treatment. For example, patients with *FLT3* mutations benefit from receiving FLT3 inhibitors such as midostaurin or gilteritinib[14]. Utilized in the setting of newly diagnosed AML with *FLT3* mutations, midostaurin in combination with induction chemotherapy has been shown to increase overall survival rates and reduce the rate of disease relapse. For patients with relapsed or refectory AML with *FLT3* mutations, the use of gilteritinib as a single agent has shown efficacy in controlling disease burden and extending survival[19]. Targeted therapies for *IDH1* and *IDH2* mutant AMLs, ivosidenib and enasidenib respectively, have also been approved for use in patients with relapsed or refractory disease, and during induction treatment in cases where aggressive chemotherapy is contraindicated, specifically when qualifying *IDH* mutations are present[20–24]. Additionally, the BCL-2 inhibitor venetoclax has shown promise, especially when combined with hypomethylating agents or low-dose cytarabine, for elderly patients or those who may not be candidates for intensive chemotherapy[18,25]e. As is true of many targeted therapies, toxicity and the risk of developing resistance often prevent long-term effectiveness. Therefore, defining combinatorial approaches to target multiple susceptibilities in a patient's leukemia is an active area of research.

Outside the use of chemotherapy and targeted therapies, stem cell transplant (SCT) offers many patients the best chance at long-term remission or disease cure. The curative potential of SCT is predicated on eradicating the patient's leukemia cells in the bone marrow and peripheral blood with chemotherapy and often irradiation therapy, and successfully engrafting healthy stem cells to repopulate a normal hematopoietic system. Stem cells can be derived from a

histocompatibility-matched related or unrelated donor in the case of an allogeneic transplant (allo-SCT) or from the patient themselves in the case of an autologous transplant (auto-SCT). While auto-SCT is indicated in certain situations, allo-SCT is more commonly used in AML patients and offers the added benefit of graft-vs-leukemia effect to target remaining cancer cells in the recipient. SCT is not without substantial risk; as such, several factors inform the decision for SCT including a patient's age, cytogenetic and molecular risk, response to initial therapy, and availability of a suitable donor. Even still, 20-40% of patients with the best predictors of transplant success will have disease recurrence[10,26]. Based on the most up-to-date guidelines, allo-SCT is generally indicated for patients whose relapse probability without transplant is >35%; adverse- and intermediate-risk AML patients most often qualify by this metric, however, clearance of measurable residual disease after induction is becoming a more widely accepted indicator for transplant success irrespective of cytogenetic risk[10,13]. Those patients with favorable risk who reach CR1 with adequate MRD clearance are not recommended for transplant, because the risk of non-relapse complications and mortality with treatment is considered to outweigh the reduction in relapse risk without[4,10,13,16]. Improvements in supportive care and prophylaxis have reduced the severity of leading comorbidities such as infection and graft-vs-host disease (GVHD) [27] and the possibility of SCT in older patients is increasing with the availability of reduced-intensity conditioning (RIC) regimens[10,28]. Even with these improvements, many patients choose to forego SCT given the risk of comorbidities and post-transplant relapse.

1.1.4 AML pathophysiology and classification

AML can develop in patients with a prior history of hematologic disorder, including myelodysplastic syndrome (MDS) or a myeloproliferative neoplasm (MPN), or as a result of previous therapy for an unrelated disease (therapy-related AML, tAML), but most commonly,

AML arises as a *de novo* entity. Irrespective of origin, AML manifests through the aberrant proliferation of immature myeloid cells, effectively crowding out normal hematopoiesis in the bone marrow. The stages of myeloid differentiation arrest are variable between patients, representing undifferentiated myeloblasts as well as more differentiated blast pollutions with characteristics of myelocytes, monocytes, erythrocytes, and/or megakaryocytes. The French-American-British (FAB) classification system accounts for this variability using a staging system to label each leukemia as one of seven subtypes (M0-M7) depending on morphologic features[10,29].

Recurrent genetic abnormalities including chromosomal translocations and gene mutations are well documented in malignant cells of patients and serve as the dominant biomarkers in a second staging system for AML as defined by the World Health Organization (WHO)[10]. Subclassifications are defined by the presence of specific chromosomal translocations or gene mutations and are used to stratify patients into favorable, intermediate, or adverse risk groups. Importantly, several AML patients lack chromosomal abnormalities and/or risk-defining mutations, despite comprehensive sequencing studies suggesting an average of three acquired mutations are detected per patient leukemia. These findings not only highlight the diversity of underlying mutations in AML but also the ongoing need to functionally characterize the leukemogenic potential of the individual as well as co-occurring mutations.

First established in the setting of familial cancer, the two-hit hypothesis was used to describe cancer predisposition as the inheritance of a mutated tumor suppressor allele, followed by a somatic hit to the opposite allele, leading to loss of both functional copies and ultimately uncontrolled proliferation[30]. Soon after, the two-hit model was extended to sporadic, or *de novo*, cancer whereby two rare genetic alterations affect both alleles of a critical gene in an individual

cell. This model was further adapted to describe the pathophysiology of AML where co-occurring mutations impair normal hematopoiesis and activate pro-proliferative pathways. In this model, genetic events don't necessarily need to occur in opposing alleles but rather can affect distinct regions that cooperate to drive leukemogenesis. Germline mutations are accounted for in the model of AML development for some patients, and in certain contexts have been shown to drive leukemogenesis in the absence of a second hit (predominantly seen cases of MLL-rearranged pediatric AML). Although late-onset leukemia with germline predisposition is represented in the population, *de novo* adult AML is typically characterized by subsequent somatic mutations that arise in myeloid progenitors and confer a clonal advantage to the affected cells[6,31,32].

While the co-occurrence of mutations is foundational to our understanding of AML development, studies of clonal origin and evolution in the past decade suggest that the traditional "two-hit" hypothesis oversimplifies AML pathophysiology. Early models of AML pathogenesis proposed that two different types of mutations were required for malignant transformation. Recurrent mutations were characterized by functional consequence into two groups: Class I mutations, leading to uncontrolled proliferation and anti-apoptotic signaling, confer constitutively activate tyrosine kinases or dysregulate down-stream signaling factors, while Class II mutations disrupt normal hematopoietic differentiation by inhibiting key transcription factors and transcriptional regulators[33,34]. Combinations of Class I and Class II mutations are detected in AML patients and may accurately describe disease etiology in those cases, but more recently, epigenetic modifiers have emerged as a prominent third class of recurrent mutations estimated to be present in < 50% of patients[1,24]. Mutations in these factors have profound impacts on DNA methylation, histone modifications, and chromatin architecture, and while they aren't thought to

cause leukemia in isolation, often cooperate with other genetic events to promote leukemogenesis.

1.2 Epigenetic dysregulation in AML

1.2.1 Recurrently mutated epigenetic factors in AML

Recurrent mutations in genes involved in epigenetic pathways are common in AML and are thought to contribute to leukemia development by disrupting epigenetic patterns relative to normal hematopoietic cells. Genomic analysis of AML performed by the Cancer Genome Atlas consortium showed that epigenetic modulators are heavily represented among recurrently mutated genes, and specifically, more than 50% of patients have at least one mutation in *DNMT3A, IDH1* and *IDH2*, or *TET2*, all of which are involved in DNA methylation either in their normal state or when mutated[1]. Importantly, mutations in these factors occur at high enough frequency in AML patients to suggest their roles as initiating events or 'driver' mutations in leukemia pathogenesis. Many studies have reported on consistent methylation phenotypes associated with each of these mutations, demonstrating their reproducible consequences at the level of the epigenome. Although extensive work has been done to characterize the role of these mutations in the developmental processes of leukemia, our understanding of the functional consequences associated with altered methylation is still incomplete.

The *DNMT3A* gene encodes DNA methyltransferase 3 alpha, one of two enzymes responsible for establishing de novo DNA methylation patterns during embryonic development and hematopoiesis[35–40]. *DNMT3A* mutations are the third most commonly occurring mutation in AML patients and have been associated with poor prognosis and reduced overall survival[1,5]. While mutations are known to occur in all three active domains of the gene, missense mutations in the R882 residue within the methyltransferase domain are by far the most common. The R882

variant has consequently become the most well-studied *DNMT3a* mutation, now known to significantly impair its enzymatic activity and result in aberrant DNA hypomethylation thought to disrupt the differentiation of hematopoietic stem cells and give rise to pre-leukemic clones. Research suggests that *DNMT3A* mutations alone might not be sufficient to induce overt leukemia but can collaborate with other mutations to drive leukemogenesis[35,38,39,41,42]. We've come to understand that *DNMT3A* mutations are genetic events associated with the aging process and are present in roughly 20% of individuals over the age of 60, a phenomenon referred to as age-related clonal hematopoiesis (ARCH)[31,43,44]. Individuals carrying these mutations in their hematopoietic stem/progenitor cells have an increased risk of developing hematological malignancies but may remain asymptomatic for years, or until a second mutation arises to drive leukemia transformation. *In vitro* and *in vivo* models of *DNMT3A* mutations have been instrumental in understanding the role of *DNMT3A* in normal and abnormal hematopoiesis. Notably, *DNMT3A* knockout mice have been developed to model and test hypotheses related to clonal hematopoiesis. Importantly, these mice display expanded hematopoietic stem cell compartments, skewed differentiation, and increased self-renewal, reinforcing the gene's critical role in normal hematopoiesis[35,38,40,42].

Ten-eleven translocation 2 (*TET2*) mutations are the fifth most commonly occurring mutation in adult AML, affecting approximately 20-25% of newly diagnosed individuals[5,37,45–47]. The prognostic significance of *TET2* mutations in AML remains somewhat controversial, however, recent studies suggest that *TET2* mutations in association with intermediate-risk cytogenetics may confer a more favorable prognosis, particularly in the absence of concurrent mutations in other epigenetic regulators or signaling pathways which may indicate a poor prognosis[10,46,48]. As an enzyme that catalyzes the oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine

(5hmC), TET2 plays a critical step in DNA demethylation[49–51]. Typically, *TET2* mutations arising in AML result in loss of function gene products and in many cases are bi-allelic, effectively depleting the cell of TET2 activity[1,5,45]. In the presence of these mutations, DNA demethylation is significantly inhibited, resulting in genome-wide patterns of hypermethylation. Similar to *DNMT3A* mutations, *TET2* mutations are commonly observed in individuals with ARCH and are thought to arise in a primitive hematopoietic stem or progenitor cell, conferring a state of abnormal HSC expansion without causing overt leukemia. Additional cooperating mutations are therefore thought to be required for leukemia transformation[31]. Deletion of *TET2* in the bone marrow compartment of mice successfully recapitulates the DNA hypermethylation phenotype observed in AML patients and is sufficient to increase the self-renewal capacity of HSCs and skew cell differentiation toward monocytic/granulocytic lineages[49,52–56]. Recently, the loss of *TET2* activity in *in vitro* and *in vivo* models has been associated with hypermethylation of key enhancers whose activity is thought to be necessary to regulate the expression of factors involved in myeloid differentiation[57]. However, the direct functional consequences of enhancer hypermethylation on regulatory activity are still unclear.

IDH1/2 mutations are the third and fourth most commonly recurring mutations in AML, affecting ~20% of patients, and are associated with altered DNA methylation patterns that are thought to occur through the disruption of active DNA demethylation[1,5,58,59]. *IDH1* and *IDH2* encode metabolic enzymes not normally involved in DNA methylation, but when mutated produce 2-hydroxyglutarate (2HG) that inhibits the TET family of enzymes, thereby reducing active demethylation. Analysis of DNA methylation in primary AML samples using array-based technologies and enhanced reduced-representation bisulfite sequencing has demonstrated that DNA methylation is increased in samples with *IDH* mutations[60–63]. While the direct effects of

these changes on gene regulation have been challenging to identify, the contribution of *IDH* mutations to leukemogenesis has been established in mouse models[60,61,61,64]. Expression of either *IDH1*^{R132H} or *IDH2*^{R140Q} blocks normal hematopoietic differentiation, promotes myeloproliferation, and can result in AML transformation in the presence of cooperating mutations. These studies establish the contribution of *IDH* mutations to AML development and suggest this may occur by disrupting the balance between DNA methylation and demethylation.

1.2.2 The epigenetic landscape of DNA methylation

Epigenetics refers to reversible, but heritable, chemical modifications added to DNA and RNA that regulate gene expression and genome organization without changing the primary DNA sequence. Epigenetic modifications have different effects on genome regulation depending the specific epigenetic 'mark' and its location in the genome, which often makes it difficult to reliably predict their effect on functional outputs, including gene expression. However, associations between epigenetic modifications and transcriptional outputs have been defined from measurements made in model systems and primary human samples, providing a basis for understanding the contribution of epigenetic regulation to normal development and how aberrant epigenetic patterns may contribute to cancer. DNA methylation (DNAme) is perhaps the most well-studied epigenetic modification in normal development and malignant states, and is thought to regulate gene expression by altering the accessibility of DNA to transcription factors and other regulatory proteins[65–68]. Although methods for studying DNA methylation have improved significantly in the past decade, the landscape of methylation changes in AML and how they relate to chromatin organization and gene regulation is still not fully understood.

DNAme refers to the addition of a methyl group to the 5' carbon of cytosine bases and is generally associated with processes of cell differentiation, X inactivation, locus imprinting, and

tumorigenesis[69,70]. DNAme is non-randomly distributed across the genomes of normal cells, almost exclusively occurring at cytosines in a C-G dinucleotide (CpG) context. At a gross level, the absence of methylation is associated with permissive chromatin and access to the transcriptional sequences on DNA, while the presence is correlated with repression. Interestingly, CpG methylation levels exhibit a bimodal distribution across the genome, representing a population of largely unmethylated cytosines and a population of cytosines that remain fully methylated, suggesting a sequence and/or context specificity for the modification. While methylation patterns are tissue and cell-type-specific due to their association with gene expression, most CpGs in the genome are hypermethylated (60%), whereas those in CpG-dense regions such as gene promoters, tend to be hypomethylated[70]. One possible explanation for this site specificity is related to the propensity for methylated cytosines to become deaminated, which can lead to the conversion of 5mC to thymine (T) and result in a C>T transition mutations in the DNA sequence (one of the most common genetic substitutions found in human cancers[24,71]. By restricting DNA methylation from promoters, the likelihood for mutations to occur in critical regulatory regions is greatly reduced.

Although we often measure DNA methylation as a steady state, the life cycle of DNA methylation is a dynamic and finely regulated process crucial for gene expression and cellular identity. It begins during embryonic development when DNA is initially unmethylated[69]. De novo DNA methyltransferases establish methylation patterns by adding methyl groups to cytosine residues at specific sites. These patterns are faithfully propagated through cell divisions by maintenance methyltransferases, ensuring epigenetic inheritance[72]. While DNA methylation is relatively stable, it is not static. The kinetics of DNA methylation involve the potential for changes in response to various stimuli including cell differentiation signals, environmental

exposures, aging, and disease. DNMT3A and TET proteins are two of the key factors involved in DNA methylation turnover, actively adding, and removing methyl groups and allowing for dynamic adjustments to the epigenetic landscape. The rates and extent of these changes can vary between different genomic regions and cell types, influencing gene regulation and cellular responses. For example, in the case of many stem and progenitor cell populations, the process of lineage commitment and cell fate are driven by the transcriptional landscape which is thought to be heavily dependent on DNA methylation. As a modification known to influence the binding of certain transcription factors to DNA, high DNA methylation turnover rates have been associated with tissue-specific enhancer regions during development, as was reported in an elegant study of methylation kinetics in human embryonic stem cells[73–75].

1.2.3 DNA methylation in AML

In AML cells, DNAme patterns become dysregulated, and often present as diffuse, global hypomethylation across the genome and/or focal hypermethylation. In cancer more broadly, hypermethylation has been linked to the silencing of tumor suppressor genes involved in important cellular processes such as DNA repair, cell cycle control, and apoptosis, however, very few examples of altered methylation directly affecting gene expression have been reported in AML. Moreover, AML exhibits significant heterogeneity in DNA methylation patterns associated with different molecular subtypes of the disease, making it difficult to define the unifying consequences of aberrant methylation in AML. As mentioned previously, AMLs with mutations in DNAme modifying enzymes have pronounced methylation phenotypes, but AMLs lacking these mutations also display subtype-specific methylation patterns. For example, AML with NPM1 mutations or MLL rearrangements often exhibits specific DNA methylation patterns

that stratify with molecular classification although these phenotypes have been less well studied[76,77].

More recently, the detection and classification of aberrant methylation as epi-alleles, also referred to as epi-mutations, has improved in line with our ability to phase epigenetic modifications with specific haplotypes. Importantly, the detection of epi-mutations helps us account for epigenetic hits in the two-hit hypothesis of cancer development. One example of this phenomenon in AML occurs at the *GATA2* locus. Allele-specific expression of *GATA2* has been well-documented in AML samples, but more recently, the role of allele-specific hypermethylation has emerged as an initiating event of AML pathogenesis. Specifically, allele-specific methylation of the *GATA2* promoter has been shown to occur nearly in 90% *CEBPA*-mutated AML, whereby active transcription from that allele gets shut down, leaving only a single copy of *GATA2*[78]. Concurrent overactivation of the opposite allele by a distal superenhancer region compensates for the cellular dosage of *GATA2* often leading to increased levels. As a protein whose balance is known to be finely tuned in normal hematopoiesis, altered *GATA2* expression is thought to contribute to a pre-leukemic phenotype that likely cooperates with *CEPBa* mutations to drive leukemia development. There are likely other examples of methylation epimutations in AML that we have yet to identify that may help link aberrant methylation phenotypes with expression consequences that aren't immediately obvious using previous approaches.

As mentioned above, aberrant methylation at non-promoter regulatory elements such as enhancers has been described in AML, particularly in the presence of *TET2* mutations. It is yet unclear if the altered methylation of enhancers directly affects activity. There is reason to believe that methylation within enhancer regions could prevent proper localization of transcription

factors, especially those known to have methylation-sensitive binding like CTCF. As our ability to link enhancers to their cognate genes continues to improve, we will be able to refine our understanding of the functional consequences associated with methylation at enhancers.

1.3 3-Dimensional Genome Organization

1.3.1 Hierarchy of genome organization

The human genome, linearly comprised of three billion base pairs of DNA nucleotides, is organized hierarchically, where different levels of structure are thought to play a role in regulating gene expression, ensuring DNA integrity, and managing the vast amount of genetic information within the cell[79–82]. Recent advancements in genomic technologies, including chromosome conformation capture techniques, have helped unravel the hierarchical organization of the 3D genome with greater resolution, facilitating the study of spatial organization as it relates to genome function. In combination with molecular studies of nucleosome remodeling, histone modifications, and DNA methylation, we are starting to better understand the mechanisms governing 3D architecture and detect alterations that may be responsible for or related to cancer pathogenesis.

At the broadest scale, the genome is partitioned into chromosomal territories within the nucleolus, where each chromosome occupies non-random, distinct regions. Within these territories, chromosomes exhibit hierarchical organization at multiple levels, including topologically associating domains (TADs), sub-TADs, loops, and diffuse chromatin interactions[90–94]. TADs are megabase-sized genomic regions characterized by high levels of self-interactions. They act as structural and functional units of the genome, segregating chromatin into distinct regulatory domains. TAD boundaries are demarcated by insulator elements, such as CTCF-binding sites, which restrict chromatin interactions between adjacent TADs, thereby

regulating gene expression by spatially orienting proper enhancer-promoter interactions and preventing aberrant regulatory crosstalk between neighboring regions[83–85].

Within TADs, finer-scale structures known as sub-TADs or contact domains have been identified. While TAD architecture tends to be stable, these substructures exhibit some level of variability and are thought to be associated with regulatory interactions that drive and support cellular differentiation and identity. Interaction within these sub-structures can be somewhat diffuse, helping to nucleate the domains, or highly focal in the case of chromatin loops formed by robust interactions between distal regulatory regions, such as enhancers and gene promoters. These loops are often mediated by DNA binding proteins and complexes, including the CTCF, cohesin complex, and the mediator complex, which facilitate long-range chromatin interactions[79,83,86,87]. Loop extrusion models propose that the cohesin complex drives the formation of chromatin loops by progressively extruding DNA until it encounters boundary elements or convergent CTCF-binding sites, highlighting the importance of proper localization of CTCF in the genome[88].

The Mediator complex is also known to play a crucial role in regulating distal interactions. As a key transcriptional coactivator in eukaryotic gene expression, the mediator complex facilitates communication between enhancer regions and gene promoters by physically interacting with both transcription factors bound at enhancers and the preinitiation complex (PIC) assembled at promoters[89,90]. This interaction helps to bring distal enhancers into proximity with their target promoters, promoting the initiation of transcription. Some subunits of the Mediator complex have been shown to possess histone acetyltransferase (HAT) activity, which can lead to the acetylation of histones near gene promoters[91]. This histone modification is associated with transcriptional activation by promoting an open chromatin conformation conducive to

transcription. Additionally, the Mediator complex interacts with chromatin remodeling complexes, further facilitating changes in chromatin structure that are necessary for DNA accessibility and transcriptional activation[90].

Dynamic changes in the 3D genome organization occur during cellular differentiation, development, and response to environmental cues, highlighting the plasticity and regulatory complexity of the genome architecture. Dysregulation at any level of organization can potentially wreak havoc on genome regulation and expression programs necessary for healthy cell development. For example, the disruption of TAD boundaries and chromatin loops in the absence or inhibition of CTCF binding has been correlated with dysregulation of gene expression linked to developmental disorders, cancer, and other diseases[84,85,92]. Ongoing characterization of AML-associated changes in the epigenetic landscape as they relate to functional consequences on 3D genome architecture will provide novel insights into leukemia pathogenesis and offer potential strategies for targeted treatment.

1.3.2 AML-associated alterations in 3D genome architecture

In vitro studies provide circumstantial evidence suggesting that AML cells can remodel genome interactions in the presence of hypermethylated elements to enable alternative enhancer usage. This may be necessary to maintain gene expression when a cognate enhancer has decreased chromatin accessibility and prohibits activators from binding DNA. In certain cases of insulator hypermethylation, TAD boundaries become compromised and allow cross-talk of regulatory elements between adjacent regions. Termed enhancer-hijacking, elevated expression of oncogenes in experimental models have been described concomitantly with the formation of novel interactions with highly active enhancers in neighboring TADs where insulators break

down. Loss of insulation in many of these examples has been linked to hypermethylation of CTCF binding sites and loss of CTCF occupancy[93].

A recently published study reported the first comprehensive analysis of 3D genome architecture in primary AML samples[94]. Their investigation focused on recurrent structural variation (chromosomal fusions) as a potential cause for altered genome organization, describing subtle subtype-specific features of A/B compartmentalization, topologically associating domains, and chromatin loops. Perhaps the most striking outcome of their analysis was the detection of AML-specific promoter-enhancer and promoter-silencer loops surrounding fusion events. They further validated the role of repressive loops on target gene expression in experimental models providing some of the first evidence for silencer-hijacking events in AML samples. Studies that expand the search for altered interactions beyond regions implicated in structural variation will be important in describing the full spectrum of AML-associated changes in genome organization.

1.3 Study Rationale, Hypothesis, and Specific Aims

The ability to characterize functional consequences of aberrant DNA hypermethylation in *IDH* mutant AML has been impeded by the lack of genome-wide studies in primary AML samples. Although previous studies using targeted DNA methylation approaches have reported the general effects of *IDH1* and *IDH2* mutations on DNA methylation[62,76,95,96], a genome-wide methylation analysis in primary AML samples had not yet been described at the outset of our project. It has therefore been unclear whether *IDH1* and *IDH2* mutations display a truly unique methylation phenotype compared to other AML subtypes. In addition, most regions with altered methylation occur beyond gene promoters, which makes it challenging to directly link these changes to expression of the appropriate gene. However, the development of chromatin capture assays that measure direct interactions between regulatory sequences and genes now make it possible to

assign regulatory elements with altered methylation to their cognate gene promoters. We, therefore, hypothesized that whole-genome DNA methylation datasets combined with chromatin interaction data will identify functionally important regulatory sequences with altered DNA methylation that may contribute to leukemogenesis by directly influencing gene expression. To test this hypothesis, we have conducted experiments to address the following aims.

Aim 1: Define *IDH* mutation-associated changes in genome-wide DNA methylation in AML. We identified regions of the genome with unique methylation changes in primary AML samples with *IDH* mutations (*IDH*^{mut}) compared with normal bone marrow and other AML subtypes. Using an extensive set of whole genome bisulfite sequencing (WGBS) data from 55 primary AML samples and normal hematopoietic stem and progenitor cells (HSPCs), differentially methylated regions were identified and assessed for unique methylation levels specific to *IDH*^{mut} AML samples.

Aim 2: Identify functional regulatory elements affected by *IDH*^{mut}-associated methylation changes. We integrated WGBS with histone modification data, measurements of 3D genome interactions, and gene expression profiles in primary AML samples to characterize the activity of regions with *IDH*^{mut} -specific DNA methylation. Through this line of investigation, we aimed to understand whether focal changes in methylation in these regions affect functional regulatory elements and whether these changes are associated with alterations in cis-regulatory activity, local chromatin interactions, and/or changes in the expression of key target genes.

Aim 3: Characterize the activity of candidate enhancer DMRs in AML cell lines. We used targeted genome editing in AML cell lines to define the functional role of DNA methylation at DNA elements with *IDH*^{mut} -associated hypermethylation. To determine whether sequences with

altered methylation have regulatory activity, we deleted selected DNA elements identified in Aim 2 to determine whether they are involved in regulating 3D architecture and/or expression of their target genes.

1.4 References

1. Mardis, E. R. *et al.* Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
2. Khwaja, A. *et al.* Acute myeloid leukaemia. *Nat. Rev. Dis. Primer* **2**, 1–22 (2016).
3. Kantarjian, H. Acute myeloid leukemia—Major progress over four decades and glimpses into the future. *Am. J. Hematol.* **91**, 131–145 (2016).
4. Kantarjian, H. M. *et al.* Acute Myeloid Leukemia: Historical Perspective and Progress in Research and Therapy Over 5 Decades. *Clin. Lymphoma Myeloma Leuk.* **21**, 580–597 (2021).
5. Papaemmanuil, E. *et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N. Engl. J. Med.* **374**, 2209–2221 (2016).
6. Welch, J. S. *et al.* The Origin and Evolution of Mutations in Acute Myeloid Leukemia. *Cell* **150**, 264–278 (2012).
7. Schlenk, R. F. *et al.* Mutations and Treatment Outcome in Cytogenetically Normal Acute Myeloid Leukemia. *N. Engl. J. Med.* **358**, 1909–1918 (2008).
8. Burd, A. *et al.* Precision medicine treatment in acute myeloid leukemia using prospective genomic profiling: feasibility and preliminary efficacy of the Beat AML Master Trial. *Nat. Med.* **26**, 1852–1858 (2020).

9. Patel, J. P. *et al.* Prognostic Relevance of Integrated Genetic Profiling in Acute Myeloid Leukemia. *N. Engl. J. Med.* **366**, 1079–1089 (2012).

10. Döhner, H. *et al.* Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood* **140**, 1345–1377 (2022).

11. De Kouchkovsky, I. & Abdul-Hay, M. 'Acute myeloid leukemia: a comprehensive review and 2016 update'. *Blood Cancer J.* **6**, e441–e441 (2016).

12. de Botton, S. *et al.* Enasidenib vs conventional care in older patients with late-stage mutant-IDH2 relapsed/refractory AML: a randomized phase 3 trial. *Blood* **141**, 156–167 (2023).

13. Lachowiez, C. A. *et al.* Comparison and validation of the 2022 European LeukemiaNet guidelines in acute myeloid leukemia. *Blood Adv.* **7**, 1899–1909 (2023).

14. Gilliland, D. G. & Griffin, J. D. The roles of FLT3 in hematopoiesis and leukemia. *Blood* **100**, 1532–1542 (2002).

15. Mims, A. S. *et al.* Comparison of clinical and molecular characteristics of patients with acute myeloid leukemia and either TP73 or TP53 mutations. *Leukemia* **35**, 1188–1192 (2021).

16. Morton, L. M. *et al.* Evolving risk of therapy-related acute myeloid leukemia following cancer chemotherapy among adults in the United States, 1975-2008. *Blood* **121**, 2996–3004 (2013).

17. Daver, N., Cortes, J., Kantarjian, H. & Ravandi, F. Acute myeloid leukemia: advancing clinical trials and promising therapeutics. *Expert Rev. Hematol.* **9**, 433–445 (2016).

18. Iyer, S. G., Elias, L., Stanchina, M. & Watts, J. The treatment of acute promyelocytic leukemia in 2023: Paradigm, advances, and future directions. *Front. Oncol.* **12**, 1062524 (2023).

19. Perl, A. E. *et al.* Gilteritinib or Chemotherapy for Relapsed or Refractory *FLT3* -Mutated AML. *N. Engl. J. Med.* **381**, 1728–1740 (2019).

20. Linch, D. C., Hills, R. K., Burnett, A. K., Russell, N. & Gale, R. E. Therapy for isocitrate dehydrogenase 2 (IDH2)R172-mutant acute myeloid leukaemia. *Br. J. Haematol.* **196**, 1348–1352 (2022).

21. Pirozzi, C. J. & Yan, H. The implications of IDH mutations for cancer development and therapy. *Nat. Rev. Clin. Oncol.* **18**, 645–661 (2021).

22. Wang, F. *et al.* Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science* **340**, 622–626 (2013).

23. Abdel-Wahab, O. & Levine, R. L. Mutations in epigenetic modifiers in the pathogenesis and therapy of acute myeloid leukemia. *Blood* **121**, 3563–3572 (2013).

24. Shih, A. H. *et al.* Mutational Cooperativity Linked to Combinatorial Epigenetic Gain of Function in Acute Myeloid Leukemia. *Cancer Cell* **27**, 502–515 (2015).

25. Ehsan, H., Iqbal, Q., Masood, A. & Grunwald, M. R. Durable remission of acute myeloid leukemia in an elderly patient following a limited course of azacitidine and venetoclax. *Leuk. Res. Rep.* **18**, 100345 (2022).

26. Li, Y. & Zhou, F. Efficacy of bone marrow transplantation in treating acute myeloid leukemia: a systematic review and meta-analysis. *Am. J. Transl. Res.* **15**, 1–12 (2023).

27. Sweeney, C. & Vyas, P. The Graft-Versus-Leukemia Effect in AML. *Front. Oncol.* **9**, 1217 (2019).

28. Niederwieser, D., Lange, T., Cross, M., Basara, N. & Al-Ali, H. Reduced intensity conditioning (RIC) haematopoietic cell transplants in elderly patients with AML. *Best Pract. Res. Clin. Haematol.* **19**, 825–838 (2006).

29. Vardiman, J. W., Harris, N. L. & Brunning, R. D. The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood* **100**, 2292–2302 (2002).

30. Knudson, A. G. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc. Natl. Acad. Sci.* **68**, 820–823 (1971).

31. Jaiswal, S. *et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).

32. Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478 (2014).

33. Takahashi, S. Current findings for recurring mutations in acute myeloid leukemia. *J. Hematol. Oncol.J Hematol Oncol* **4**, 36 (2011).

34. Liang, D.-C. *et al.* Cooperation of Gene Mutations Including Class I, Class II and Tumor Suppressor Genes In Childhood Acute Myeloid Leukemia and Their Impacts on Survivals. *Blood* **116**, 2713 (2010).

35. Challen, G. A. *et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* **44**, 23–31 (2012).

36. Mayle, A. *et al.* Dnmt3a loss predisposes murine hematopoietic stem cells to malignant transformation. *Blood* **125**, 629–638 (2015).

37. Ley, T. J. *et al.* DNMT3A Mutations in Acute Myeloid Leukemia. *N. Engl. J. Med.* **363**, 2424–2433 (2010).

38. Russler-Germain, D. A. *et al.* The R882H DNMT3A Mutation Associated with AML Dominantly Inhibits Wild-Type DNMT3A by Blocking Its Ability to Form Active Tetramers. *Cancer Cell* **25**, 442–454 (2014).

39. Spencer, D. H. *et al.* CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression. *Cell* **168**, 801-816.e13 (2017).

40. Smith, A. M. *et al.* Functional and epigenetic phenotypes of humans and mice with DNMT3A Overgrowth Syndrome. *Nat. Commun.* **12**, 4549 (2021).

41. Jeong, M. *et al.* Loss of Dnmt3a Immortalizes Hematopoietic Stem Cells *In Vivo*. *Cell Rep.* **23**, 1–10 (2018).

42. Challen, G. A. *et al.* Dnmt3a and Dnmt3b Have Overlapping and Distinct Functions in Hematopoietic Stem Cells. *Cell Stem Cell* **15**, 350–364 (2014).

43. Steensma, D. P. *et al.* Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126**, 9–16 (2015).

44. Midic, D. *et al.* Prevalence and dynamics of clonal hematopoiesis caused by leukemia-associated mutations in elderly individuals without hematologic disorders. *Leukemia* **34**, 2198–2205 (2020).

45. Weissmann, S. *et al.* Landscape of TET2 mutations in acute myeloid leukemia. *Leukemia* **26**, 934–942 (2012).

46. Abdel-Wahab, O. & Levine, R. L. Mutations in epigenetic modifiers in the pathogenesis and therapy of acute myeloid leukemia. *Blood* **121**, 3563–3572 (2013).

47. Delhommeau, F. *et al.* Mutation in TET2 in Myeloid Cancers. *N. Engl. J. Med.* **360**, 2289–2301 (2009).

48. Bowman, R. L. & Levine, R. L. TET2 in Normal and Malignant Hematopoiesis. *Cold Spring Harb. Perspect. Med.* **7**, a026518 (2017).

49. Ko, M. *et al.* Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc. Natl. Acad. Sci.* **108**, 14566–14571 (2011).

50. Nakajima, H. & Kunimoto, H. TET2 as an epigenetic master regulator for normal and malignant hematopoiesis. *Cancer Sci.* **105**, 1093–1099 (2014).

51. Huang, Y. & Rao, A. Connections between TET proteins and aberrant DNA modification in cancer. *Trends Genet.* **30**, 464–474 (2014).

52. Shide, K. *et al.* TET2 is essential for survival and hematopoietic stem cell homeostasis. *Leukemia* **26**, 2216–2223 (2012).

53. Moran-Crusio, K. *et al.* Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. *Cancer Cell* **20**, 11–24 (2011).

54. Quivoron, C. *et al.* TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. *Cancer Cell* **20**, 25–38 (2011).

55. Cimmino, L. *et al.* Restoration of TET2 Function Blocks Aberrant Self-Renewal and Leukemia Progression. *Cell* **170**, 1079-1095.e20 (2017).

56. Li, Z. *et al.* Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* **118**, 4509–4518 (2011).

57. Fortin, J. *et al.* Distinct and opposite effects of leukemogenic Idh and Tet2 mutations in hematopoietic stem and progenitor cells. *Proc. Natl. Acad. Sci.* **120**, e2208176120 (2023).

58. Zarnegar-Lumley, S. *et al.* Characteristics and prognostic impact of IDH mutations in AML: a COG, SWOG, and ECOG analysis. *Blood Adv.* **7**, 5941–5953 (2023).

59. Davis, A. R., Canady, B. C., Aggarwal, N. & Bailey, N. G. Clinicopathologic Features of IDH2 R172–Mutated Myeloid Neoplasms. *Am. J. Clin. Pathol.* **160**, 89–97 (2023).

60. Sasaki, M. *et al.* IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. *Nature* **488**, 656–659 (2012).

61. Turcan, S. *et al.* Mutant-IDH1-dependent chromatin state reprogramming, reversibility, and persistence. *Nat. Genet.* **50**, 62–72 (2018).

62. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell* **18**, 553–567 (2010).

63. Gruber, E. *et al.* Inhibition of mutant IDH1 promotes cycling of acute myeloid leukemia stem cells. *Cell Rep.* **40**, (2022).

64. Lu, C. *et al.* IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474–478 (2012).

65. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).

66. Guillamot, M., Cimmino, L. & Aifantis, I. The Impact of DNA Methylation in Hematopoietic Malignancies. *Trends Cancer* **2**, 70–83 (2016).

67. Blecua, P., Martinez-Verbo, L. & Esteller, M. The DNA methylation landscape of hematological malignancies: an update. *Mol. Oncol.* **14**, 1616–1639 (2020).

68. Izzo, F. *et al.* DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nat. Genet.* **52**, 378–387 (2020).

69. Smallwood, S. A. *et al.* Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat. Genet.* **43**, 811–814 (2011).

70. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339–344 (2012).

71. Shi, K. *et al.* Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat. Struct. Mol. Biol.* **24**, 131–139 (2017).

72. Sen, G. L., Reuter, J. A., Webster, D. E., Zhu, L. & Khavari, P. A. DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature* **463**, 563–567 (2010).

73. Farlik, M. *et al.* DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell Stem Cell* **19**, 808–822 (2016).

74. Alajem, A. *et al.* DNA methylation patterns expose variations in enhancer-chromatin modifications during embryonic stem cell differentiation. *PLoS Genet.* **17**, e1009498 (2021).

75. Prasad, R., Yen, T. J. & Bellacosa, A. Active DNA demethylation—The epigenetic gatekeeper of development, immunity, and cancer. *Adv. Genet.* **2**, e10033 (2021).

76. Figueroa, M. E. *et al.* DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17**, 13–27 (2010).

77. Spencer, D. H. *et al.* Epigenomic analysis of the HOX gene loci reveals mechanisms that may control canonical expression patterns in AML and normal hematopoietic cells. *Leukemia* **29**, 1279–1289 (2015).

78. Mulet-Lazaro, R. *et al.* Allele-specific expression of GATA2 due to epigenetic dysregulation in CEBPA double-mutant AML. *Blood* **138**, 160–177 (2021).

79. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**, 1110–1121 (2016).

80. Dekker, J. & van Steensel, B. The Spatial Architecture of Chromosomes. in *Handbook of Systems Biology* 137–151 (2013). doi:10.1016/B978-0-12-385944-0.00007-1.

81. Kloetgen, A., Thandapani, P., Tsirigos, A. & Aifantis, I. 3D Chromosomal Landscapes in Hematopoiesis and Immunity. *Trends Immunol.* **40**, 809–824 (2019).

82. Leidescher, S. *et al.* Spatial organization of transcribed eukaryotic genes. *Nat. Cell Biol.* **24**, 327–339 (2022).

83. Liu, Y. & Dekker, J. CTCF–CTCF loops and intra-TAD interactions show differential dependence on cohesin ring integrity. *Nat. Cell Biol.* **24**, 1516–1527 (2022).

84. Narendra, V., Bulajié, M., Dekker, J., Mazzoni, E. O. & Reinberg, D. CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev.* **30**, 2657–2662 (2016).

85. Valton, A.-L. & Dekker, J. TAD disruption as oncogenic driver. *Curr. Opin. Genet. Dev.* **36**, 34–40 (2016).

86. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).

87. Seitan, V. C. *et al.* Cohesin-Based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res.* **23**, 2066–2077 (2013).

88. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338–1345 (2019).

89. El Khattabi, L. *et al.* A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers. *Cell* **178**, 1145-1158.e20 (2019).

90. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).

91. Richter, W. F., Nayak, S., Iwasa, J. & Taatjes, D. J. The Mediator complex as a master regulator of transcription by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **23**, 732–749 (2022).

92. Wang, M., Sunkel, B. D., Ray, W. C. & Stanton, B. Z. Chromatin structure in cancer. *BMC Mol. Cell Biol.* **23**, 35 (2022).

93. Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).

94. Xu, J. *et al.* Subtype-specific 3D genome alteration in acute myeloid leukaemia. *Nature* **611**, 387–398 (2022).

95. Chan, S. M. & Majeti, R. Role of DNMT3A, TET2, and IDH1/2 mutations in pre-leukemic stem cells in acute myeloid leukemia. *Int. J. Hematol.* **98**, 648–657 (2013).

96. Chen, C. *et al.* Cancer-associated IDH2 mutants drive an acute myeloid leukemia that is susceptible to Brd4 inhibition. *Genes Dev.* **27**, 1974–1985 (2013).

Chapter 2: Focal disruption of DNA methylation at enhancers in *IDH* mutant AML cells

Elisabeth R. Wilson, Nichole M. Helton, Sharon E. Heath, Robert S. Fulton, Jacqueline E. Payton, John S. Welch, Matthew J. Walter, Peter Westervelt, John F. DiPersio, Daniel C. Link, Christopher A. Miller, Timothy J. Ley, and David H. Spencer

From

Focal disruption of DNA methylation dynamics at enhancers in *IDH* mutant AML cells.

Wilson, E.R., Helton, N.M., Heath, S.E. *et al.* Focal disruption of DNA methylation dynamics at enhancers in IDH-mutant AML cells. *Leukemia* **36**, 935–945 (2022).

Declarations

Conception and design: ERW, DHS. Acquisition of data: SEH, RSF, JSW, JEP, PW, MJW, JFD, DCL, CAM, TJL. Data processing, statistical and computational analysis: ERW, DHS. Manuscript preparation: ERW, DHS. Scientific input and manuscript comments: JSW, TJL. Administrative, technical, or material support: SEH, RSF, JEP, PW, MJW, JFD, DCL, CAM, TJL, DHS. Study supervision: DHS.

2.1 Abstract

Recurrent mutations in *IDH1* or *IDH2* in acute myeloid leukemia (AML) are associated with increased DNA methylation, but the genome-wide patterns of this hypermethylation phenotype have not been comprehensively studied in AML samples. We analyzed whole-genome bisulfite sequencing data from 15 primary AML samples with *IDH1* or *IDH2* mutations, which identified ~4,000 focal regions that were uniquely hypermethylated in IDH^{mut} samples vs. normal CD34+ cells and other AMLs. These regions had modest hypermethylation in AMLs with biallelic *TET2*

mutations, and levels of 5-hydroxymethylation that were diminished in *IDH* and *TET*-mutant samples, indicating that this hypermethylation results from inhibition of TET-mediated demethylation. Focal hypermethylation in IDH^{mut} AMLs occurred at regions with low methylation in CD34+ cells, implying that DNA methylation and demethylation are active at these loci. AML samples containing *IDH* and $DNMT3A^{R882}$ mutations were significantly less hypermethylated, suggesting that IDH^{mut}-associated hypermethylation is mediated by DNMT3A. IDH^{mut}-specific hypermethylation was highly enriched for enhancers that form direct interactions with genes involved in normal hematopoiesis and AML, including *MYC* and *ETV6*. These results suggest that focal hypermethylation in *IDH*-mutant AML occurs by altering the balance between DNA methylation and demethylation, and that disruption of these pathways at enhancers may contribute to AML pathogenesis.

2.2 Introduction

DNA methylation changes in acute myeloid leukemia (AML) are caused by disruptions in the processes that add or remove 5-methyl groups to cytosines (1,2). In normal and malignant hematopoietic cells, *de novo* DNA methylation is catalyzed primarily by the DNA methyltransferase DNMT3A (3,4), which methylates unmethylated DNA substrates. Demethylation occurs passively after DNA synthesis in the absence of DNMT1-mediated propagation of hemi-methylated DNA, and actively via hydroxylation of 5mC by the TET family of hydroxylases. Alterations in these opposing forces result in either increased or decreased DNA methylation in AML cells. These changes include diffuse hypomethylation across large genomic regions and focal hypermethylation in CpG islands (CGIs). We recently showed that CGI hypermethylation in AML is mediated by DNMT3A and is present in nearly all AML subtypes (5). In addition to these changes, specific DNA methylation patterns correlate with AML

mutations that influence DNA methylation. This includes the *DNMT3A*^{R882} mutation, which impairs DNA methylation activity and results in a focal, canonical hypomethylation phenotype (5).

Mutations in *IDH1* and *IDH2* are also associated with altered DNA methylation patterns (6,7) that are thought to occur by disrupting active DNA demethylation. *IDH1* and *IDH2* encode metabolic enzymes not normally involved in DNA methylation, but when mutated produce 2-hydroxyglutarate (2HG) (8) that inhibits the TET family of enzymes (9), thereby reducing active demethylation. Analysis of DNA methylation in primary AML samples using array-based technologies and enhanced reduced-representation bisulfite sequencing has demonstrated that DNA methylation is increased in samples with *IDH* mutations (6,10). While the direct effects of these changes on gene regulation have been challenging to identify, the contribution of *IDH* mutations to leukemogenesis has been established in mouse models. Expression of either *IDH1*^{R132H} or *IDH2*^{R140Q} blocks normal hematopoietic differentiation, promotes myeloproliferation (11–13), and can result in AML transformation in the presence of cooperating mutations (13,14). These studies establish the contribution of *IDH* mutations to AML development and suggest this may occur by disrupting the balance between DNA methylation and demethylation.

Although previous studies using targeted DNA methylation approaches have reported the general effects of *IDH1* and *IDH2* mutations on DNA methylation (6,7,10,15), a genome-wide methylation analysis in primary AML samples has not yet been described. It is therefore unclear whether *IDH1* vs. *IDH2* mutations cause hypermethylation at the same or different genomic loci, and whether these methylation changes are distinct from DNMT3A-mediated CGI hypermethylation. Additionally, although *IDH* mutations are thought to cause hypermethylation

via inhibition of TET enzymes, the overlap in methylation phenotypes between AML samples with these mutations is unclear. Here, we performed a genome-wide analysis of DNA methylation in primary AML samples with recurrent mutations in *IDH1*, *IDH2*, or *TET2* using whole-genome bisulfite sequencing (WGBS). WGBS data from normal hematopoietic cells and AML samples with other mutational profiles were included to define the methylation phenotypes specific to *IDH* mutations, and to determine their relationship to 'generic' AML-associated methylation changes. We integrated these data with epigenetic modifications and three-dimensional genome architecture from primary AML samples to characterize the functional genomic elements that may be affected by disruption of the balance between DNA methylation and demethylation in AML.

2.3 Materials and Methods

Patient samples

Primary AML samples and normal hematopoietic cells for epigenetic studies were obtained from presentation AML and normal bone marrow aspirates, following informed consent using protocol (201011766) approved by the Human Research Protection Office at Washington University as described previously (5) . All experiments with AML samples used total bone marrow cells for DNA preparation.

Whole genome bisulfite and oxidative bisulfite sequencing and data analysis

Whole-genome bisulfite sequencing data for 38 samples were described previously (5). Data for 13 additional samples were generated using 50ng of DNA with the Swift Accel-NGS Methyl-Seq library preparation kit. Oxidative bisulfite sequencing libraries were prepared following treatment of 200ng of DNA with the TrueMethyl oxBS module (Cambridge Epigenetix) prior to bisulfite conversion and Swift library construction and sequencing on NovaSeq 6000 instruments

(Table S1). Data were aligned to the GRCh38 reference and processed into methylated read counts using biscuit (16) with default parameters. Differentially methylated CpGs (DMCs) were identified between AML groups and CD34+ cells using read count data via DSS (17) and required a minimum methylation difference of 0.2. DMCs were then used to identify differentially methylated regions (DMRs) with >10 CpGs and a difference in mean methylation of 0.2. *IDH*^{mut}- and *TET2*^{mut}-specific DMCs and DMRs were subsequently identified by comparing these samples to all other AML samples via the DSS beta-binomial test in the methylkit Bioconductor package (18). 5hmC values were obtained by subtracting the methylation ratios from OxWGBS data from WGBS data at all CpGs with coverage > 10x.

ChIP-seq for histone modifications

ChIP-seq was performed using ChIPmentation (19) with the following antibodies: H3K27me3 (9733S), and H3K27ac (8173S) from Cell Signaling Technology, and H3K4me1 (ab1012) from Abcam. Sequencing was performed on a NovaSeq 6000 (Illumina, San Diego, CA) to obtain ~50 million 2x150 bp reads. Data were analyzed via adapter trimming with trimgalore and alignment to GRCh38 using bwamem (20). Normalized coverage for visualization and analysis used the deeptools "bamCoverage" tool (21), and peaks were called with MACS2 (22). Statistical comparisons with DESeq2 (23) used raw fragment counts at peak summits, and visualizations were prepared with Gviz (24). Superenhancer analysis was conducted using ROSE software (25,26) with default parameters.

RNA-seq analysis

RNA-seq data from AML samples were obtained from the AML TCGA study (15). TPM values were obtained using kallisto (27) and gene counts were generated using the tximport Bioconductor package (28) in R with the tx2gene option set to accomplish gene-level

summarization. Previously published RNA-seq data for normal CD34+ cells generated using the same procedures that were used for the AML samples (29,30) were obtained as raw sequencing reads from the short-read archive (GSE48846) and processed as described above.

Hi-C data analysis

Hi-C data were obtained from previous studies of 3D genome interactions in primary AML samples (31) and normal hematopoietic stem/progenitors (32). All libraries were generated using MboI digestion prior to proximity ligation and data were analyzed using the juicer pipeline (33). Loops were identified with HICCUPs and were analyzed using bedtools (34) to identify overlap with genes and putative enhancers. Visualizations used the GenomicInteractions and Gviz R packages (24).

Data Availability

All raw data from primary AML samples presented in this study is available in dbGaP (accession number phs000159). Processed data from this study are available for public download at the following site: [1,68].

2.4 Results

2.4.1 Primary AML samples with *IDH1* or *IDH2* mutations are focally hypermethylated at regions with low methylation in normal hematopoietic cells.

We performed WBGS using 15 primary bone marrow aspirate samples from AML patients with canonical *IDH* mutations, including seven with *IDH1*$^{R132C/G}$, seven with *IDH2*^{R140Q}, and one with an *IDH2*^{R172K} allele (referred to throughout as *IDH*^{mut}). These data were analyzed with WGBS data from 36 other primary AML samples representing nine mutational categories, including six with biallelic loss-of-function mutations in *TET2*, and primary CD34+ cells from six healthy

adult bone marrow donors (5). All AML samples were previously sequenced using whole genome and/or whole exome sequencing (15,35) that confirmed the mutations affecting DNA methylation were present in the dominant leukemic clone (Figure 1A). Importantly, the 15 AML samples with *IDH* mutations were wild type for *DNMT3A* and *TET2* to minimize effects of other mutations on DNA methylation patterns. We first performed an unsupervised analysis of genome-wide methylation in 1 kb bins using principal component analysis (PCA). This demonstrated that most AML samples formed a diffuse cluster separate from CD34+ cells (Figure 1B). AML samples with either *DNMT3A*^{R882} or *IDH* mutations (and some with *TET* mutations) formed sub-clusters on opposite sides of the main AML group, which is consistent with the hypomethylation phenotype of AML cells with the *DNMT3A*^{R882} mutation (5) and suggests *IDH*^{mut} samples may also have unique methylation features compared to other AMLs.

We next determined whether *IDH* mutations have global or context-dependent effects on DNA methylation by analyzing methylation levels in regions defined by chromatin states in hematopoietic stem/progenitors (36). This demonstrated that quiescent and repressed chromatin states had lower methylation in most AMLs compared to CD34 cells, whereas bivalent regions (which are enriched for CpG islands) were hypermethylated in nearly all samples (Figure 1C). Enhancer and regions flanking transcriptional start sites (TSS) supervised a cluster of hypermethylated AMLs containing 14 of the 15 *IDH*^{mut} samples. Mean methylation in *IDH*^{mut} AMLs at enhancer regions was significantly higher vs. both CD34 cells and AMLs without *IDH* mutations (Figure 1D, P=0.009 and P=0.0002, respectively). *IDH*^{mut} AMLs also tended to have higher mean methylation vs. other AML groups both genome-wide (Figure 1E, adjusted P=0.02) and in regions with other chromatin states (Figures S1A-B), but not in CpG islands (Figure 1F;

P=0.14), indicating that *IDH* mutations do not result in an exaggerated CGI hypermethylation phenotype.

We next determined the extent to which *IDH* mutations result in focal methylation changes by performing differentially methylated CpG (DMC) and region (DMRs) analysis (37) between AMLs with *IDH1* or *IDH2* mutations and CD34 cells. There were 6,309 DMRs in *IDH1*^{mut} AMLs, of which 99% were hypermethylated relative to CD34 cells (methylation difference >0.2, FDR<0.05 with >10 CpGs; Figure 1G); this was more than any mutation-defined AML group. *IDH2*^{mut} AMLs had fewer DMRs (N=4,915), although most were also hypermethylated (85%). AMLs with *IDH1* or *IDH2* mutations also had the highest fraction of hypermethylated CpGs (DMCs) (85% and 87%, respectively; see Figure S1C), most of which were contained in DMRs (Figure S1D-E). Interestingly, although *IDH* mutations are thought to inhibit active demethylation, most *IDH*^{mut} DMRs had low methylation in normal hematopoietic cells. For example, 60% of the *IDH*^{mut} DMRs had a mean methylation <0.3 in both CD34+ cells (Figure 1H) and more mature myeloid cell populations (Figure S1F), suggesting that DNA methylation pathways must be active in these regions despite the low methylation levels at these loci in normal cells.

2.4.2 *IDH*^{mut}-specific methylation changes are distinct from AML-associated CGI hypermethylation and are influenced by *IDH* mutation type.

We next performed a second statistical comparison of the DMRs (and DMCs) identified in AMLs with *IDH* mutations vs. CD34+ cells to identify loci with methylation levels in the *IDH*^{mut} samples that was significantly different from all other AML samples. AMLs with mutations in *DNMT3A* or *TET2* were excluded from this analysis given their established hypomethylation phenotype (*DNMT3A*) and potential to phenocopy *IDH* mutations (*TET2*). This resulted in 4,388

and 2,552 *IDH1*^{mut} and *IDH2*^{mut}-specific DMRs, respectively, nearly all of which were hypermethylated relative to the other AML samples (Figures 2A-C, Tables S2-3). Similar results were observed at the DMC level (Figure S2A-B). Most of these DMRs displayed low methylation in normal cells, with 60% of *IDH1*^{mut}-specific and 58% of *IDH2*^{mut}-specific loci having a methylation level <0.3 in CD34+ and mature myeloid cells (Figure S2C-D). There was extensive overlap between the *IDH* mutation-specific DMRs (94% [2,399/2,552] of *IDH2*^{mut}-specific DMRs overlapped an *IDH1*^{mut}-specific DMR), and AML samples with either mutation were hypermethylated at both DMR sets (Figure 2D). However, hierarchical clustering demonstrated considerable variability in methylation between the *IDH1*^{mut} and *IDH2*^{mut} samples (Figure 2E). Notably, three *IDH2*^{mut} AMLs had lower methylation across the union of *IDH*^{mut}-specific DMRs. *IDH2*^{mut} AML samples were also less methylated than *IDH1*^{mut} samples at the combined set of *IDH*^{mut} DMRs (0.54 vs. 0.70, respectively; P=0.04), but were hypermethylated relative to CD34+ cells (Figure 2E). This was not related to mutant *IDH* allele abundance (all samples had VAFs >30%, Table S1), and did not correlate with other recurrent mutations, including *NPM1*c (4 in *IDH1*^{mut} and 3 in *IDH2*^{mut} samples, Figure 2D; all samples were wild type for *DNMT3A* and *TET2*). Comparable differences in methylation were observed at the DMC level (Figure S2E-F), suggesting this phenomenon was not an artifact of DMR identification.

Interestingly, the *IDH*^{mut}-specific DMRs demonstrated markedly different CpG density and overlap with genomic annotations compared to hypermethylated regions in other AML samples. For example, both *IDH1*^{mut}-specific and *IDH2*^{mut}-specific DMRs displayed significantly less overlap with annotated CGIs compared to 4,573 hypermethylated regions identified in at least 2 other AML mutation categories (18% and 20% of *IDH1*^{mut} and *IDH2*^{mut} DMRs overlapped a CGI, respectively, compared to 54% of commonly hypermethylated regions; see Figure 2F), and

had lower CpG density (mean CpG density of 0.81 and 0.79 vs. 1.26, respectively; P-values<0.0001; Figure 2G). Promoters were also underrepresented in *IDH1*^{mut}-specific and *IDH2*^{mut}-specific DMRs (21% and 20% of *IDH1*^{mut} and *IDH2*^{mut} DMRs overlapped a promoter, vs 31% of commonly hypermethylated regions; Figure 2G). *IDH*^{mut}-specific DMCs showed similar levels of overlap with annotated regions as DMRs (Figure S2G), further suggesting that *IDH*-associated hypermethylation is distinct from AML-associated CGI hypermethylation.

2.4.3 Hypermethylation in *TET2*^{mut} AMLs overlaps with *IDH*^{mut}-specific hypermethylation but does not phenocopy the extent of methylation changes.

We next determined whether AML samples with biallelic loss-of-function mutations in *TET2* shared similar genome-wide patterns of hypermethylation with *IDH*^{mut} AMLs. Initial comparison of the *TET2*^{mut} AMLs vs. normal CD34+ cells yielded fewer DMRs (and DMCs) and a lower proportion of hypermethylated regions compared to the combined set of DMRs in *IDH*^{mut} samples (3,083 vs. 7,569 DMRs, and 75% vs 99% hypermethylated regions, respectively; see Figures 1G and S1A), consistent with previous reports (6,10). Hierarchical clustering of *TET2*^{mut} samples with the set of *IDH*^{wt}/*TET2*^{wt}/*DNMT3A*^{wt} AMLs at these regions did not reveal striking methylation differences between the two groups (Figure S3A). Consistent with this result, only 188 *TET2*^{mut}-specific DMRs were identified using the approach described above (with *IDH*^{mut} and *DNMT3A*^{R882} AMLs excluded from the analysis) (Figure 3A). Although most *TET2*^{mut}-specific DMRs were hypermethylated relative to CD34+ cells and other AMLs (171 of 188), the fraction was less than in either *IDH1*^{mut} or *IDH2*^{mut} AMLs (89% vs 99% and 99%, respectively). Similarly, *TET2*^{mut}-specific DMCs showed subtle hypermethylation (Figure S3B). *TET2*^{mut}-specific DMRs were also not enriched for CGIs and promoters compared to a set of regions commonly hypermethylated in AML (12% of *TET2* DMRs overlapped a CGI vs. 54% of common hypermethylated DMRs; 17% of *TET2* DMRs overlapped a promoter vs. 31% of

hypermethylated DMRs; see Figure S3C-D), suggesting these regions do not reflect CGI hypermethylation.

To investigate the interaction between *IDH*^{mut} and *TET2*-mediated demethylation, we compared *TET2*^{mut}-specific and *IDH*^{mut}-specific DMRs and performed oxidative bisulfite sequencing (38) to measure 5-hydroxymethylation (5hmC) in *TET2*^{mut}, *IDH*^{mut}, and *TET2*^{wt}/*IDH*^{wt} samples. This analysis showed that 68% (127 of 188) of the *TET2*^{mut}-specific DMRs overlapped an *IDH*^{mut}-specific hypermethylated region (Figure 3B; $P<0.0001$ using a permutation test for overlaps using all DMRs identified in any AML group). *TET2*^{mut} AMLs also displayed higher methylation levels at the combined set of 4,541 *IDH*^{mut}-specific DMRs compared to CD34+ cells (mean methylation of 0.35 vs. 0.26; 40% of DMRs with increased methylation via beta binomial hypothesis testing with $q<0.05$; Figure 3C-D). Analysis of 5hmC using paired oxidative and standard whole-genome bisulfite sequencing (oxWGBS and WGBS with conversion rates ranging from 73%-83%; see Figure S3E) demonstrated low calculated levels of 5hmC across the genomes of all samples (0.44-0.66% in *TET2*^{mut}, 0.52-1.22% in *TET2*^{wt}, 0.17-0.25% in *IDH*^{mut}; Figure S3F), with higher levels in enhancer regions (Figure S3G) and identifiable peaks at selected loci (Figure S3H). Calculated 5hmC was statistically higher in *IDH*^{mut} DMRs compared to regions that were hypermethylated in other AML samples or in constitutively methylated heterochromatic regions (adjusted $P=0.0009$ and $P=4x10^{-7}$ for a difference in mean 5hmC in all samples at *IDH*^{mut} DMRs vs. 4,586 commonly hypermethylated DMRs and 105,519 heterochromatin regions, respectively; see Figures 3E, S2I-J). AML samples with *TET2*, *IDH1*, or *IDH2* mutations had lower calculated 5hmC levels at *IDH*^{mut} DMRs compared to AMLs that were wild type for these genes (Figures 3E, S2I-J), providing evidence that these mutations influence methylation turnover at these loci.

2.4.4 DNA hypermethylation in *IDH*^{mut} AML cells requires *DNMT3A*

To assess whether *de novo* DNA methylation by DNMT3A contributes to *IDH*^{mut}-associated hypermethylation, we analyzed methylation levels at *IDH*^{mut}-specific DMRs in seven AML samples with co-occurring *IDH1* (N=5) or *IDH2* (N=2) and *DNMT3A*^{R882} mutations (R882 mutations have a more severe hypomethylation phenotype than other *DNMT3A* mutations (4,39)). Interestingly, although *DNMT3A*^{R882}/*IDH*^{mut} AMLs were still hypermethylated at *IDH*^{mut}-specific DMRs, the degree of hypermethylation was diminished, with 67% of these regions having significantly lower DNA methylation levels than samples with *IDH* mutations alone (3024 of 4541 regions having a beta-binomial adjusted P<0.05; see Figures 4A-C, S3A). Similar findings were observed in 7 additional *DNMT3A*^{R882}/*IDH*^{mut} AML samples using methylation array data from the TCGA AML study (15) (Figure S4B). To further characterize the extent of this interaction, we analyzed DNA methylation levels in *DNMT3A*^{R882}/*IDH*^{mut} AML samples at hypomethylated DMRs in AMLs with the *DNMT3A*^{R882} allele (5). Surprisingly, these regions remained nearly fully methylated in the *DNMT3A*^{R882}/*IDH*^{mut} double mutant samples, with 93% of the regions having significantly higher methylation than AMLs with *DNMT3A*^{R882} alone (4209 of 4541 regions having a beta-binomial adjust P<0.05; see Figures 4D-F, Figure S4C). Similar findings were observed in the AML TCGA data (15) (Figure S4D), strongly suggesting that DNMT3A-mediated methylation and TET-mediated demethylation occur at the same places in the genome.

2.4.5 *IDH*^{mut}-specific hypermethylated DMRs are enriched for enhancers

We next asked whether *IDH*^{mut}-specific DMRs were associated with certain chromatin states. Annotation of these DMRs with chromatin states in CD34+ cells (36) demonstrated that 44% occurred in enhancers, which was a 2-fold enrichment over regions commonly hypermethylated (Figure 5A). This enrichment was not observed in analyses on DMRs identified in other AML

subtypes (Figure S5A-C). We further defined this association using ChIP-seq peaks for active, weak, and poised enhancers using ChIP-seq data for H3K27ac, H3K4me1, and H3K27me3 modifications from 16 primary AML samples, including 2 with *IDH* mutations. This demonstrated that 47% of the *IDH*^{mut} DMRs overlapped an active enhancer, compared to 3% and 1% that overlapped poised and weak regions, respectively (Figures 5B-D). In comparison, commonly hypermethylated regions showed less overlap with active enhancers (13% of DMRs) and greater intersection with repressive H3K27me3 marks (Figure 5D). Analysis of *IDH*^{mut}-specific DMRs for transcription factor (TF) binding motifs identified binding sites for hematopoietic-associated TFs, including *SPI1*, *RUNX1*, and *MYC* (Figure 5E), further supporting the occurrence of *IDH*^{mut}-specific hypermethylation in regions with potential regulatory activity. However, quantitative analysis of H3K27ac signal over these regions in samples with and without *IDH* mutations did not identify appreciable differences (P=0.24, Figure 5F), suggesting that hypermethylation does not modify H3K27ac levels within these regions.

2.5 Discussion

Recurrent gain-of-function *IDH* mutations increase DNA methylation, but the genomic locations and functional consequences of these changes have not previously been clearly defined. Our analysis of whole-genome bisulfite sequencing data from primary AML samples shows that methylation changes caused by these mutations is not widespread, but instead manifests as thousands of focal regions that are uniquely hypermethylated compared to normal CD34+ cells and AML cells without *IDH* mutations. These regions had lower CpG density and fewer CGIs than loci that are commonly hypermethylated in AML, suggesting that *IDH*^{mut}-associated hypermethylation is caused by a distinct mechanism. The *IDH2*^{mut} AMLs in our dataset had less pronounced hypermethylation than those with *IDH1* mutations, but both were hypermethylated

at a highly overlapping set of loci. AMLs with biallelic inactivating *TET2* mutations had a far less dramatic methylation phenotype, although many of the hypermethylated DMRs identified in these samples overlapped an *IDH*^{mut}-specific DMR. Further, oxidative bisulfite sequencing demonstrated increased levels of 5hmC in these regions in AML samples that were wild type for *TET2*, *IDH1* and *IDH2*; 5hmC levels were significantly lower in *IDH*^{mut} or *TET2*^{mut} samples in these regions, providing evidence that these mutations cause increased DNA methylation by impairing TET-mediated DNA demethylation. Regions with *IDH*^{mut}-specific hypermethylation were enriched for active enhancers, many of which formed direct interactions with highly expressed AML genes, including *MYC* and *ETV6*. Although increased methylation at these loci was not associated with repressed chromatin or lower gene expression in *IDH*^{mut} AML samples, this finding demonstrates that *IDH*^{mut}-associated hypermethylation affects the regulatory sequences of genes that may contribute to AML pathogenesis.

This study adds new context to the dynamics of *de novo* DNA methylation and active demethylation pathways in normal hematopoietic cells and in AML. The fact that *IDH*^{mut}-associated hypermethylation occurs at regions with low levels of DNA methylation in normal CD34+ cells suggests that *de novo* DNA methylation and TET-mediated demethylation are both active in these regions, despite their low steady-state methylation levels. This is supported by the observation that AML samples with co-occurring *IDH* and *DNMT3A*^{R882} mutations show significantly attenuated hypermethylation, and that *IDH*^{mut}-specific DMRs have high levels of 5hmC, which is produced from 5mC as a substrate. Remodeling of DNA methylation by these processes in specific regions has been reported previously in studies of embryonic stem cells, which have shown that methylation and active demethylation are in equilibrium at many loci (1,2), and may be maintained by the occupancy of methylation and demethylation complexes

(43). Our analysis suggests this equilibrium also exists in normal hematopoietic stem/progenitor cells and is disrupted in the presence of mutant *IDH* alleles, leaving *de novo* DNA methylation unopposed. The focal nature of *IDH*^{mut}-associated hypermethylation implies that activity (or occupancy) of DNMT3A and TET enzymes is not diffuse and may instead be targeted to specific genomic regions. The genomic or epigenetic features directing this activity are unclear (44), but the enrichment of *IDH*^{mut} DMRs in active enhancers suggests that components of active chromatin may recruit methylation and demethylation machinery. The convergence of these processes at enhancers could provide clues as to why mutations with opposite effects on DNA methylation both contribute to AML development via dysregulation of common target genes.

Our analysis of 3D genome interactions involving *IDH*^{mut}-specific DMRs found that these sequences directly interact with genes that are highly expressed in hematopoiesis and AML (e.g., *MYC* and *ETV6*). Contrary to the canonical relationship between DNA methylation and activity, hypermethylation in the *IDH*^{mut} AML samples does not appear to repress either the enhancer elements or expression of their target genes. Other regulatory factors may therefore be dominant to DNA methylation at these loci, and result in persistently high gene expression. It is also possible that regions of active chromatin, such as enhancers (and superenhancers), have high rates of methylation turnover, and are therefore more susceptible to perturbations in methylation and demethylation (1,2). Focal hypermethylation may occur in DNA elements bound by factors that contribute to 'fine-tuning' these enhancers in specific cellular or developmental contexts, but that do not drive their activity in AML cells. Additional studies will be necessary to understand whether enhancer hypermethylation is a consequence of decreased occupancy of these modulating factors (45), or whether it directly prevents proper regulation in ways that contribute to AML development.

2.6 Acknowledgments

This work was supported by the National Cancer Institute (K08CA190815) and The Cancer Research Foundation Young Investigator Award to D.H.S. Support for procurement, annotation, and sequencing of human samples was provided by the Genomics of AML Program Project (P01CA101937, to Dr. Ley) and the Specialized Program of Research Excellence in AML (P50CA171963, to Dr. Link) from the NCI.

2.7 Figures

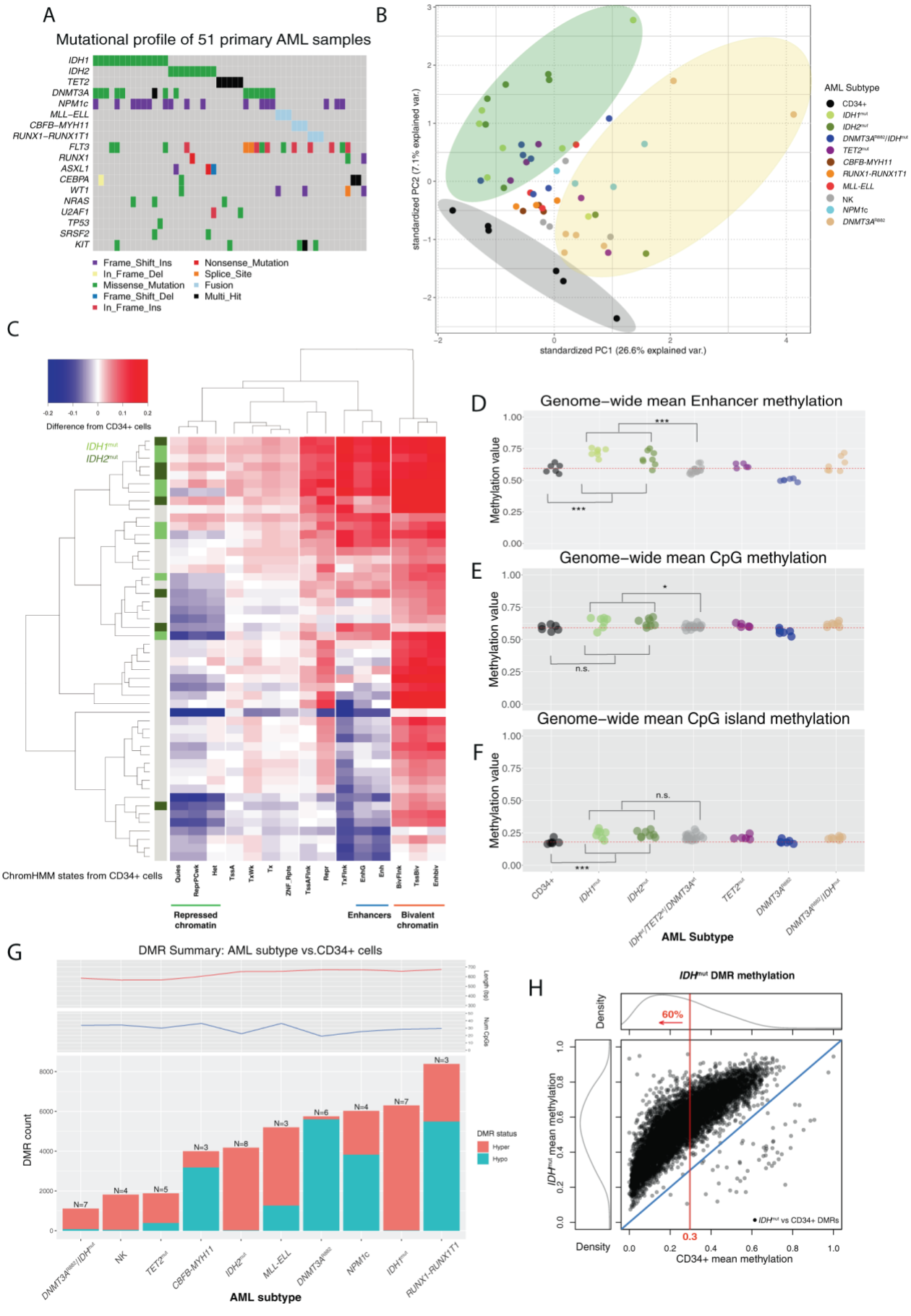

A
Mutational profile of 51 primary AML samples
IDH1
IDH2
TET2
DNMT3A
NPM1c
MLL−ELL
CBFB−MYH11
RUNX1−RUNX1T1
FLT3
RUNX1
ASXL1
CEBPA
WT1
NRAS
U2AF1
TP53
SRSF2
KIT
Frame_Shift_Ins
In_Frame_Del
Missense_Mutation
Frame_Shift_Del
In_Frame_Ins
Nonsense_Mutation
Splice_Site
Fusion
Multi_Hit
B
standardized PC2 (7.1% explained var.)
standardized PC1 (26.6% explained var.)
AML Subtype
CD34+
IDH1mut
IDH2mut
DNMT3AR882/IDHmut
TET2mut
CBFB-MYH11
RUNX1-RUNX1T1
MLL-ELL
NK
NPM1c
DNMT3AR882
C
Difference from CD34+ cells
IDH1mut
IDH2mut
ChromHMM states from CD34+ cells
Quies
ReprPCwk
Het
TssA
TxWk
Tx
ZNF_Rpts
TssAFlnk
Repr
TxFlnk
EnhG
Enh
BivFlnk
TssBiv
EnhBiv
Repressed chromatin
Enhancers
Bivalent chromatin
D
Genome−wide mean Enhancer methylation
Methylation value
E
Genome−wide mean CpG methylation
n.s.
F
Genome−wide mean CpG island methylation
AML Subtype
G
DMR Summary: AML subtype vs.CD34+ cells
Length (bp)
Num Cpgs
DMR count
DMR status
Hyper
Hypo
AML subtype
H
IDHmut DMR methylation
Density
60%
IDHmut mean methylation
IDHmut vs CD34+ DMRs
0.3
CD34+ mean methylation

Figure 2.1. Genome-wide DNA methylation patterns in 51 primary AML samples and normal CD34+ cells

A) Summary of the mutations in 51 primary AML patients analyzed using whole-genome bisulfite sequencing. B) Principal component analysis of genome-wide methylation in AML samples and CD34+ cells. Points show the values of the first and second principal components by variance explained from an analysis of genome-wide methylation summarized in mean methylation in 1 kb bins, with colors representing the defining mutation for each sample. Stratification of *IDH*^{mut} samples from CD34+ cells and AML samples with *DNMT3A*^{R882} mutations are highlighted by colored ellipses. C) Two-way hierarchical clustering of relative (difference from CD34+ cells) mean methylation levels in genomic regions define by 15 chromatin states (36) in CD34+ cells, where rows are AML samples and columns are chromatin states. Blue is less methylated than CD34+ cells and red is more methylated. *IDH1* and *IDH2* mutation status are indicated in the colored bar on the left, and selected chromatin states are shown underneath the panel. D) Mean methylation levels in the enhancer chromHMM state (derived from publicly available CD34+ epigenetic data) from WGBS for CD34+ cells (N=6) and AML subtypes (*IDH1*^{mut} or *IDH2*^{mut}, n=15; *TET2*^{mut}, n=5; *DNMT3A*^{R882}, n=6; *DNMT3A*^{R882}/*IDH*^{mut}, n=7; normal karyotype with *NPM1c* and wild-type *IDH1*, *IDH2*, *TET2*, and *DNMT3A,* n=4; Normal karyotype with wild-type *NPM1*, *IDH1*, *IDH2*, *TET2*, and *DNMT3A,* n=4; *CBFB-MYH11*, n=3; *KMT2A-ELL*, n=3; *RUNX1-RUNX1T1*, n=3). E. Mean methylation levels for ~28 million genome-wide CpGs in CD34+ cells and AML subtypes. F) Mean methylation at CpG islands in CD34+ cells and AML subtypes. G) Number of differentially methylated regions (DMRs) identified for each AML subtype compared with normal CD34+ cells. Teal and orange bars represent hypomethylated and hypermethylated DMRs with respect to normal CD34+ cells, respectively. Mean number of CpGs per DMR (top panel) and DMR length (bottom panel) are shown for each AML subtype. H. Mean methylation in *IDH*^{mut} DMRs in *IDH*^{mut} samples versus CD34+ cells. The red line indicates the percent of all DMRs with a mean methylation in CD34+ cells less than 0.3.

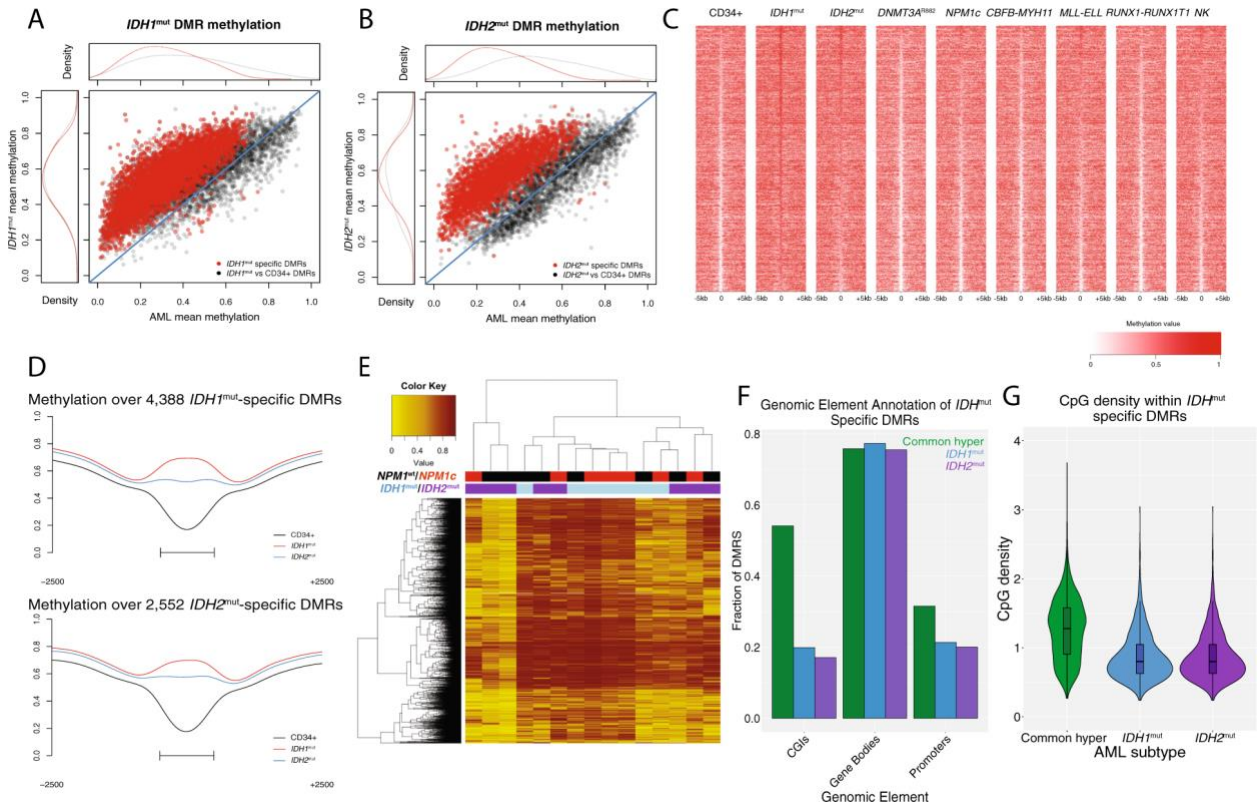


Figure 2.2. Characterization of *IDH*^{mut}-specific DMRs.

A) Mean methylation in all *IDH1*^{mut} associated DMRs vs. CD34+ cells (n=6,309; black dots) and in *IDH1*^{mut}-specific DMRs that are unique compared to all other AMLs (n=4,388; red dots). B) Mean methylation in all *IDH2*^{mut} associated DMRs vs. CD34+ cells (n=4,195; black dots) and in *IDH2*^{mut}-specific DMRs that are unique compared to all other AMLs (n=2,552;red dots). C) Locus heatmap showing mean methylation values by group for the union of *IDH1*^{mut} and *DH2*^{mut}-specific DMRs, where each column is centered over the DMR with the window extending 5kb upstream and downstream of the DMR center. D) Mean methylation across all *IDH1/2*^{mut}-specific DMRs (rows) in 15 individual *IDH*^{mut} cases (columns). The mutation status of *IDH1*, *IDH2*, and *NPM1* are indicated by the colored bars above the heatmap. Note that all AML samples in this panel are wild type for *DNMT3A* and *TET2*. E) Aggregate DMR methylation across 4,388 *IDH1*^{mut} specific DMRs and 2,552 *IDH2*^{mut} specific DMRs respectively. F) Fraction of generically hypermethylated DMRs, *IDH1*^{mut} specific DMRs, and *IDH2*^{mut} specific DMRs overlapping functional genomic elements. G) Distribution of CpG densities across generically hypermethylated regions in primary AML and *IDH1*^{mut}- and *IDH2*^{mut}-specific DMRs.

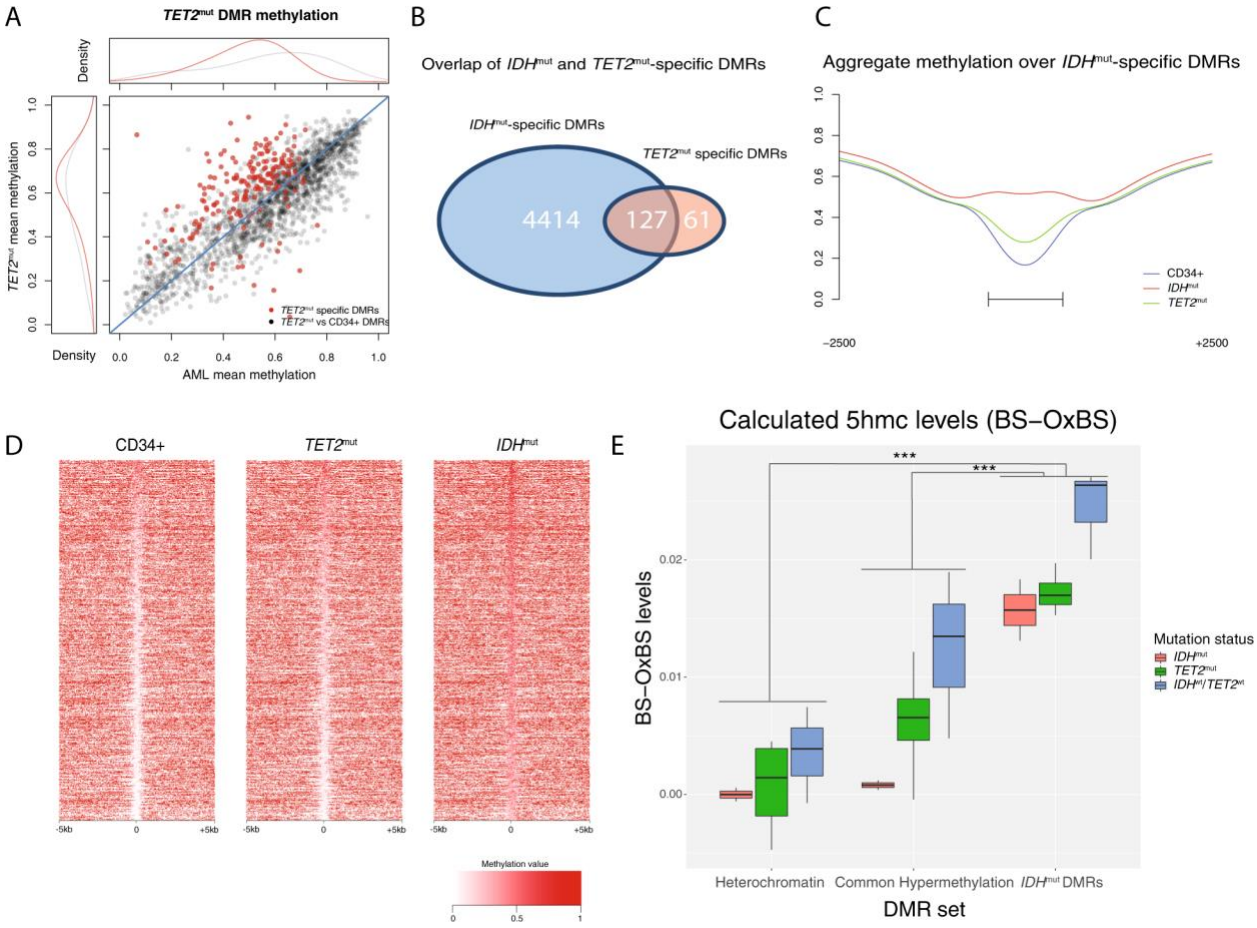


Figure 2.3. *TET2*^{mut} AMLs have modest hypermethylation that overlaps *IDH*^{mut}-specific DMRs.

A) Mean DMR methylation across 2512 *TET2*^{mut} DMRs called vs. CD34+ cells (black points) and 121 in *TET2*^{mut}-specific DMRs (red points) in *TET2* mutant samples versus all other AMLs that are wild type for *IDH, TET2, and DNMT3A*. B) Intersection of in *TET2*^{mut}-specific and *IDH*^{mut}-specific DMRs. C) Aggregate methylation over *IDH*^{mut}-specific DMRs in *IDH*^{mut} and in *TET2*^{mut} AML and CD34+ cells. D) Locus heatmap of mean methylation values for all *IDH*^{mut}-specific DMRs (rows), where each column is centered over the DMR with the window extending 5kb up- and down-stream the DMR center point. E) Mean 5hmC (WGBS minus oxWGBS) levels in *IDH*^{mut}, *TET2*^{mut}, and *IDH*^{wt}/*TET2*^{wt} AML samples at 105,519 ChromHMM heterochromatic regions, 4,586 generically hypermethylated regions (i.e., regions that were hypermethylated vs. CD34+ cells in at least 2 AMLs without *IDH1*, *IDH2*, or *TET2* mutations), and 4,541 *IDH*^{mut}-specific hypermethylated DMRs.

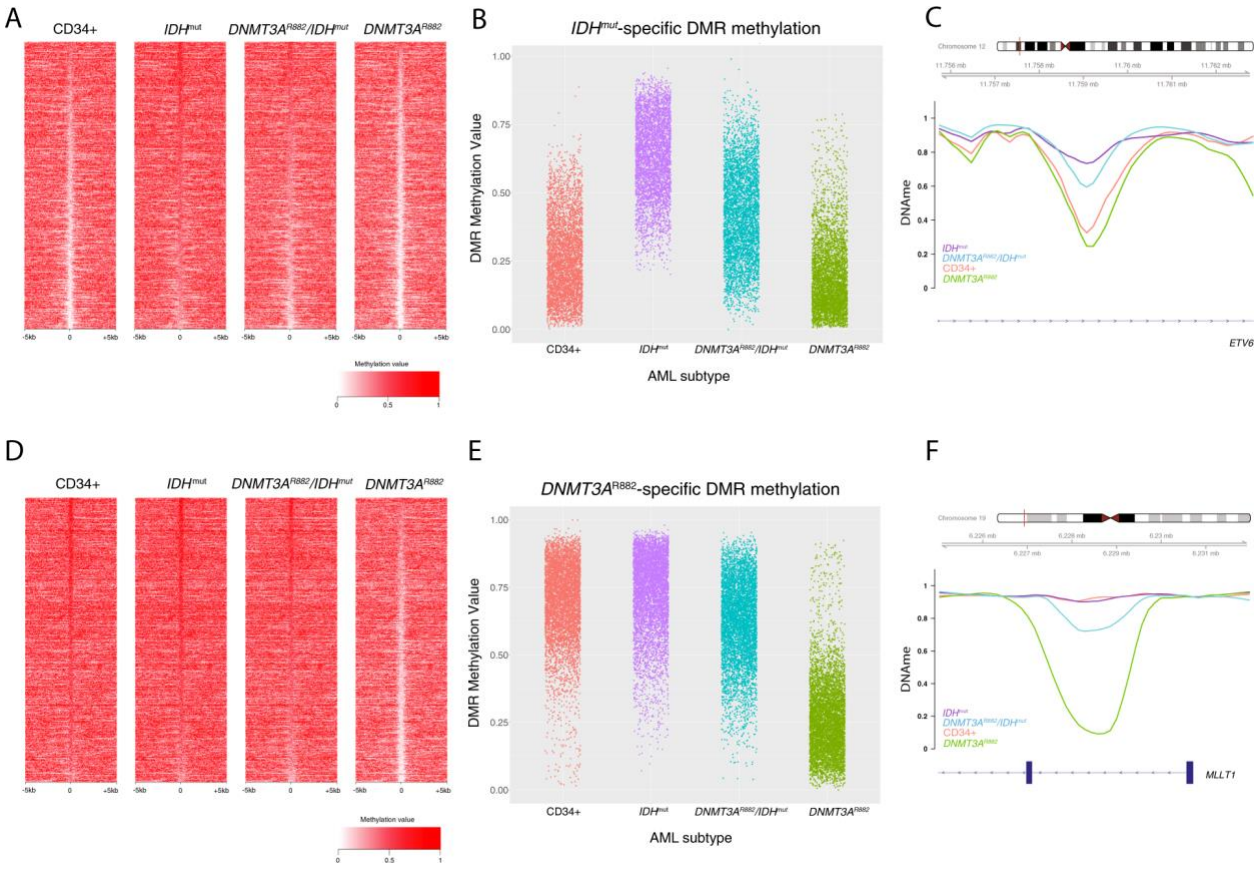


Figure 2.4. *DNMT3A*^{R882}/*IDH*^{mut} double mutant AMLs display an attenuated focal hypermethylation phenotype.

A) Locus heatmap of mean methylation at *IDH*^{mut} DMRs (rows) in *IDH1* or *IDH2* mutant, *DNMT3A*^{R882}/*IDH*^{mut} double mutant, and *DNMT3A*^{R882} AMLs, and CD34+ cells. B) Distribution of *IDH*^{mut}-specific DMR methylation levels by AML subtype. C) Example *IDH*^{mut}-specific DMR locus within the *ETV6* gene demonstrating an intermediate methylation phenotype of double mutant samples with respect to *IDH*^{mut} and *DNMT3A*^{R882} mutant AMLs. D) Methylation locus heatmap of average subtype methylation across *DNMT3A*^{R882} DMRs called vs. CD34+ cells in *IDH*^{mut}, *DNMT3A*^{R882}/*IDH*^{mut} double mutant, and *DNMT3A*^{R882} AMLs, and CD34+ cells. E) Distribution of *DNMT3A*^{R882} DMR methylation levels by AML subtype. F) Example *DNMT3A*^{R882} DMR locus within the *MLLT1* gene, demonstrating the hypomethylation phenotype of *DNMT3A*^{R882} mutant samples with respect to *IDH*^{mut} and *DNMT3A*^{R882}/*IDH*^{mut} double mutant AML samples.

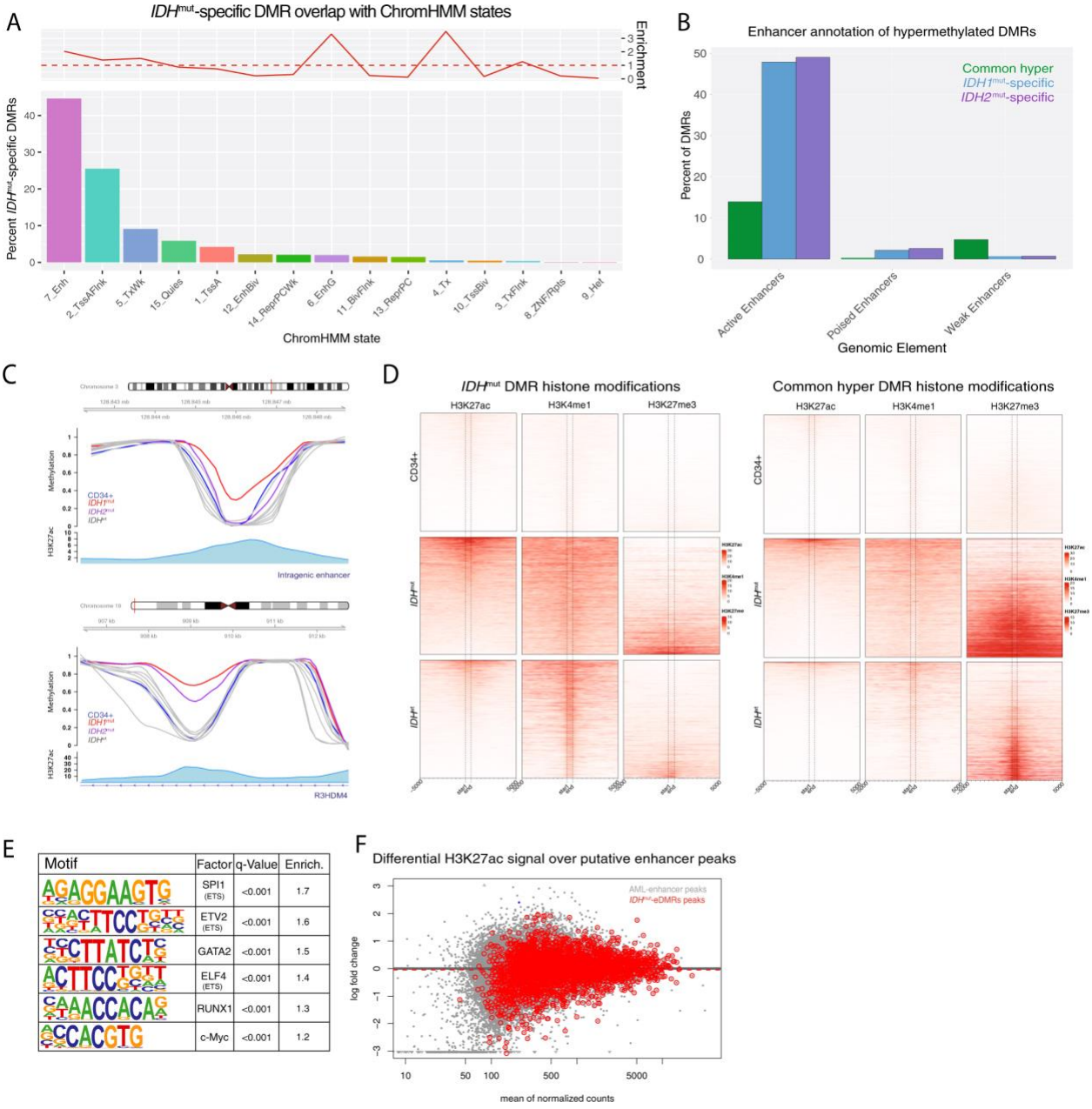


Figure 2.5. *IDH*^{mut}-specific DMRs are enriched for putative enhancers.

A) Distribution of ChromHMM chromatin states from CD34+ cells represented in *IDH*^{mut}-specific DMRs. Enrichment of chromatin states within *IDH*^{mut}-specific DMRs is shown with respect to the frequency of states overlapping regions of common CpG island hypermethylation. B) Enhancer-based annotation of common hypermethylated regions, *IDH1*^{mut}, and *IDH2*^{mut} DMRs, where DMRs intersecting an H3K27ac peak alone or in combination with H3K4me1 constitute active enhancers, an H3K27ac peak in combination with H3K27me3 constitutes poised enhancers, and H3K4me1 alone constitutes a weak enhancer profile. C) Examples of

intragenic and genic enhancer regions exhibiting *IDH1*^{mut}, *IDH2*^{mut}, or *IDH1/2*^{mut} hypermethylation compared with CD34+ cells and other AML subtypes. D) Heatmap of enhancer histone modifications and heterochromatin modifications over *IDH*^{mut}-specific DMRs (left) and generic hypermethylation (right) in CD34+ cells (N=4 H3K27ac, N=7 H3K3me1, and N=7 H3K27me3), *IDH*^{mut} AML (n=3), and *IDH*^{wt} AML samples (N=9 H3K27ac, N=10 H3K3me1, and N=24 H3K27me3). E) Differential active enhancer signal (H3K27ac) for all AML-associated putative enhancers (black points) compared to putative enhancers intersecting an *IDH1/2*^{mut}-specific DMR (red points). F) HOMER motif enrichment analysis of *IDH1/2*^{mut}-specific DMRs with respect to a background set of generically hypermethylated regions. G) Enrichment analysis of TF binding events for 445 TFs within *IDH1/2*^{mut}-specific DMRs.

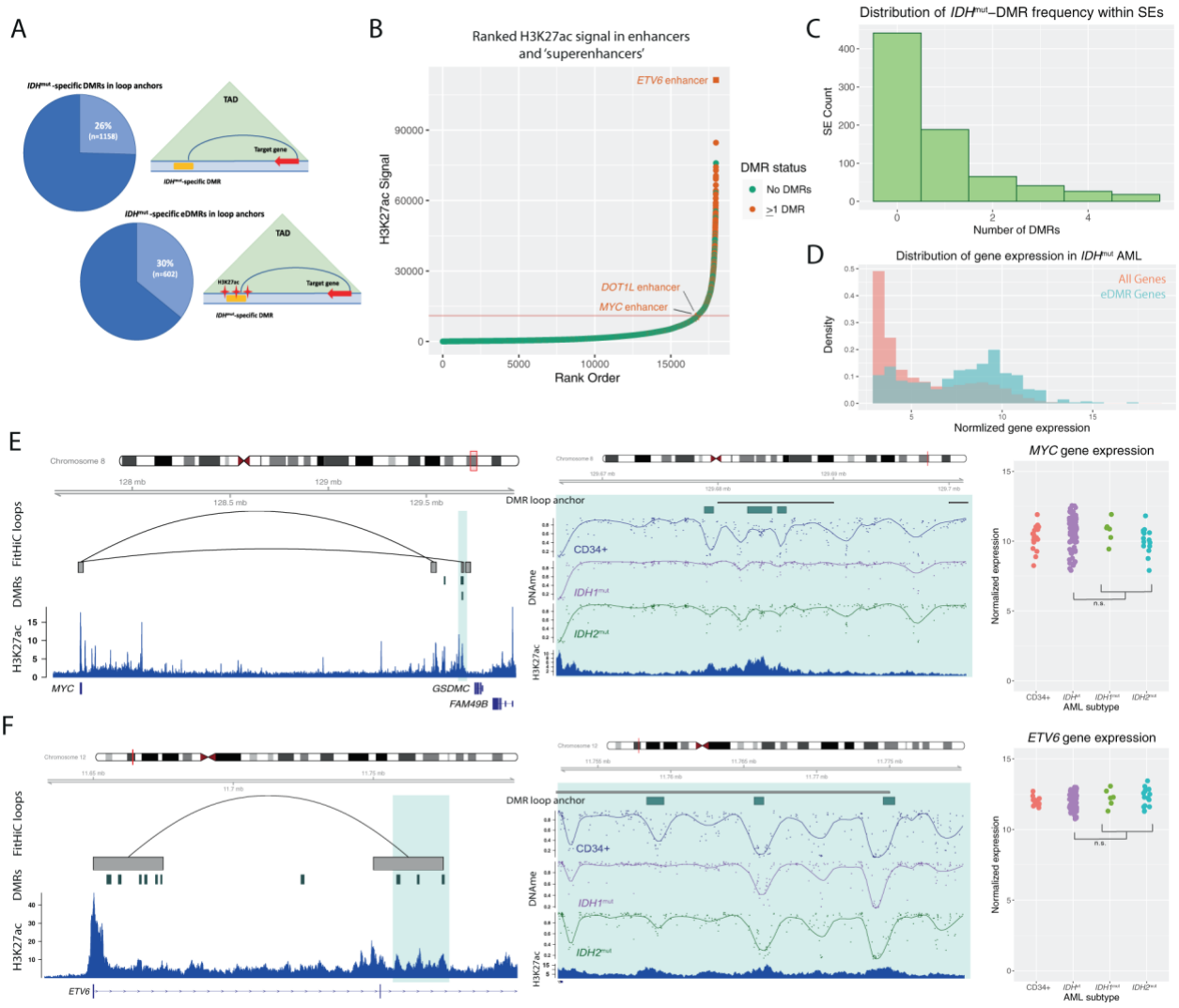


Figure 2.6. *IDH*^{mut}-specific DMRs are enriched in superenhancers and interact with highly expressed genes in AML.

A) Schematic of DMR and enhancer-associated DMRs (eDMR) and their interaction with target genes based on intersection HiC-defined genome loops. B) Rank ordered enhancer regions based on H3K27ac signal in a representative *IDH*^{mut} AML sample, annotated by presence of overlapping *IDH*^{mut}-specific DMRs (absence of DMRs indicated by green points, greater than one DMR indicated by orange points) and computationally-defined 'superenhancer' (above red line). Enhancers of specific hematopoietic genes are designated. C) Distribution of number of *IDH*^{mut}-specific DMRs overlapping a set of AML consensus superenhancers from H3K27ac data from 4 primary samples (N=779). D) Distribution of normalized gene expression values for all expressed genes (orange histogram) and a set of 750 eDMR target genes (blue histogram) in *IDH*^{mut} AML samples. E) Example *IDH*^{mut} -eDMR locus displaying interactions with the *MYC* promoter. A zoomed-in view of the locus demonstrates focal enhancer hypermethylation in *IDH1*^{mut} (purple) and *IDH2*^{mut} (green) samples compared with CD34+ cells (blue). Normalized *MYC* expression is shown for 17 CD34+ cord blood cell samples, 6 and 14 *IDH1*^{mut} and *IDH2*^{mut} samples, and 91 *IDH*^{wt} samples. F) Example *IDH*^{mut}-DMR locus in a candidate enhancer that displays robust interactions with the *ETV6* promoter. A zoomed-in locus view demonstrates focal enhancer hypermethylation in *IDH1*^{mut} (purple) and *IDH2*^{mut} (green) samples compared with CD34+ cells (blue). Normalized *ETV6* expression is shown for CD34+ cells, *IDH1*^{mut} and *IDH2*^{mut} samples, and *IDH*^{wt} samples (see E for sample sizes).

2.9 References

1. Charlton J, Jung EJ, Mattei AL, Bailly N, Liao J, Martin EJ, et al. TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. Nat Genet. 2020;52(8):819–27.
2. Ginno PA, Gaidatzis D, Feldmann A, Hoerner L, Imanci D, Burger L, et al. A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. Nat Commun. 2020;11(1):2680.
3. Challen GA, Sun D, Jeong M, Luo M, Jelinek J, Berg JS, et al. Dnmt3a is essential for hematopoietic stem cell differentiation. Nat Genet. 2012;44(1):23.
4. Russler-Germain DA, Spencer DH, Young MA, Lamprecht TL, Miller CA, Fulton R, et al. The R882H DNMT3A Mutation Associated with AML Dominantly Inhibits Wild-Type

DNMT3A by Blocking Its Ability to Form Active Tetramers. Cancer Cell. 2014;25(4):442–54.

5. Spencer DH, Russler-Germain DA, Ketkar S, Helton NM, Lamprecht TL, Fulton RS, et al. CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression. Cell. 2017;168(5):801-816.e13.
6. Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, et al. Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. Cancer Cell. 2010;18(6):553–67.
7. Akalin A, Garrett-Bakelman FE, Kormaksson M, Busuttil J, Zhang L, Khrebtukova I, et al. Base-Pair Resolution DNA Methylation Sequencing Reveals Profoundly Divergent Epigenetic Landscapes in Acute Myeloid Leukemia. Plos Genet. 2012;8(6):e1002781.
8. Losman J-A, Looper RE, Koivunen P, Lee S, Schneider RK, McMahon C, et al. (*R*)-2-Hydroxyglutarate Is Sufficient to Promote Leukemogenesis and Its Effects Are Reversible. Science. 2013;339(6127):1621–5.
9. Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim S-H, et al. Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of α-Ketoglutarate-Dependent Dioxygenases. Cancer Cell. 2011;19(1):17–30.
10. Glass JL, Hassane D, Wouters BJ, Kunimoto H, Avellino R, Garrett-Bakelman FE, et al. Epigenetic Identity in AML Depends on Disruption of Nonpromoter Regulatory Elements and Is Affected by Antagonistic Effects of Mutations in Epigenetic Modifiers. Cancer Discov. 2017;7(8):868–83.

11. Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, et al. Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. Cancer Cell. 2011;20(1):11–24.
12. Sasaki M, Knobbe CB, Munger JC, Lind EF, Brenner D, Brüstle A, et al. IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. Nature. 2012;488(7413):656.
13. Kats LM, Reschke M, Taulli R, Pozdnyakova O, Burgess K, Bhargava P, et al. Proto-Oncogenic Role of Mutant IDH2 in Leukemia Initiation and Maintenance. Cell Stem Cell. 2014;14(3):329–41.
14. Yoshimi A, Lin K-T, Wiseman DH, Rahman MA, Pastore A, Wang B, et al. Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis. Nature. 2019;574(7777):273–7.
15. TJ L, C M, Ding L. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. New Engl J Medicine. 2013;368(22):2059–74.
16. GitHub - huishenlab/biscuit: BISulfite-seq CUI Toolkit [Internet]. [cited 2021 Apr 7]. Available from: https://github.com/huishenlab/biscuit
17. Feng H, Conneely KN, Wu H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. Nucleic Acids Res. 2014;42(8):e69–e69.
18. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. 2012;13(10):R87.

19. Schmidl C, Rendeiro AF, Sheffield NC, Bock C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat Methods. 2015;12(10):963–5.
20. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013;
21. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(W1):W160–5.
22. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.
23. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
24. Hahne F, Ivanek R. Methods in Molecular Biology. Methods Mol Biology Clifton N J. 2016;1418:335–51.
25. Lovén J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, et al. Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. Cell. 2013;153(2):320–34.
26. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell. 2013;153(2):307–19.
27. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34(5):525–7.
28. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000research. 2016;4:1521.

29. Pabst C, Bergeron A, Lavallée V-P, Yeh J, Gendron P, Norddahl GL, et al. GPR56 identifies primary human acute myeloid leukemia cells with high repopulating potential in vivo. Blood. 2016;127(16):2018–27.

30. MacRae T, Sargeant T, Lemieux S, Hébert J, Deneault É, Sauvageau G. RNA-Seq Reveals Spliceosome and Proteasome Genes as Most Consistent Transcripts in Human Cancer Cells. Plos One. 2013;8(9):e72884.

31. Ghasemi R, Struthers H, Wilson ER, Spencer DH. Contribution of CTCF binding to transcriptional activity at the HOXA locus in NPM1-mutant AML cells. Leukemia. 2020;1–13.

32. Zhang X, Jeong M, Huang X, Wang XQ, Wang X, Zhou W, et al. Large DNA Methylation Nadirs Anchor Chromatin Loops Maintaining Hematopoietic Stem Cell Identity. Mol Cell. 2020;78(3):506-521.e6.

33. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. 2016;3(1):95–8.

34. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.

35. Klco JM, Miller CA, Griffith M, Petti A, Spencer DH, Ketkar-Kulkarni S, et al. Association Between Mutation Clearance After Induction Therapy and Outcomes in Acute Myeloid Leukemia. Jama. 2015;314(8):811–22.

36. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. Nat Protoc. 2017;12(12):2478–92.

37. Wu H, Xu T, Feng H, Chen L, Li B, Yao B, et al. Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. Nucleic Acids Res. 2015;43(21):gkv715.

38. Kernaleguen M, Daviaud C, Shen Y, Bonnet E, Renault V, Deleuze J-F, et al. Epigenome Editing, Methods and Protocols. Methods Mol Biology. 2018;1767:311–49.

39. Smith AM, LaValle TA, Shinawi M, Ramakrishnan SM, Abel HJ, Hill CA, et al. Functional and epigenetic phenotypes of humans and mice with DNMT3A Overgrowth Syndrome. Nat Commun. 2021;12(1):4549.

40. Shi J, Whyte WA, Zepeda-Mendoza CJ, Milazzo JP, Shen C, Roe J-S, et al. Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. Gene Dev. 2013;27(24):2648–62.

41. Pulikkan JA, Hegde M, Ahmad HM, Belaghzal H, Illendula A, Yu J, et al. CBFβ-SMMHC Inhibition Triggers Apoptosis by Disrupting MYC Chromatin Dynamics in Acute Myeloid Leukemia. Cell. 2018;174(1):172-186.e21.

42. Bahr C, Paleske L von, Uslu VV, Remeseiro S, Takayama N, Ng SW, et al. A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. Nature. 2018;553(7689):515–20.

43. Dixon G, Pan H, Yang D, Rosen BP, Jashari T, Verma N, et al. QSER1 protects DNA methylation valleys from de novo methylation. Science. 2021;372(6538):eabd0875.

44. Weinberg DN, Papillon-Cavanagh S, Chen H, Yue Y, Chen X, Rajagopalan KN, et al. The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. Nature. 2019;573(7773):281–6.

45. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. Nature. 2012;489(7414):75.

Chapter 3: Identification of functional regulatory elements affected by *IDH*^{mut}-associated methylation changes.

Elisabeth R. Wilson, Mohamed Mahgoub, Heidi Struthers, Reza Ghasemi, and David H. Spencer

Declarations

Conception and design: ERW, DHS. Acquisition of data: ERW, MM, RG, HS. Data processing, statistical and computational analysis: ERW, MM, DHS. Administrative, technical, or material support: HS, DHS. Study supervision: DHS.

3.1 Introduction

Considering our findings that *IDH*-mutation specific hypermethylation is enriched at genomic regions with features of active enhancers[1], we decided to conduct a rigorous study of 3D chromatin interactions in samples with and without *IDH* mutations to better define this relationship and the genes that may be affected by this enhancer hypermethylation phenotype. We performed a multi-tiered analysis of 3D genome architecture in 15 primary AML samples with and without *IDH1* or *IDH2* mutations. Analysis of these data was focused on the ~4,400 *IDH*^{mut}-specific DMRs identified in Aim 1 of this thesis. We evaluated potential associations between these DMRs and all levels of genome organization, including genome compartments, topologically associated domains (TADs), and chromatin loops and used 'generically' hypermethylated regions defined in Chapter 3 as 'control' regions to identify patterns that were uniquely associated with *IDH*-specific hypermethylation.

Genome interactions in this study were measured using Micro-C, an updated version of Hi-C that performs 'all-vs-all' chromatin conformation analysis at nucleosome-level resolution. While these data provide genome-wide information about genome architecture, this project focused specifically on characterizing features of the 3D landscape that overlap with enhancer-associated methylation uniquely observed in *IDH* mutant AML. Additionally, we performed capture on the same Micro-C libraries for >84,000 annotated gene promoters that enriched for promoter-anchored interactions to more accurately evaluate interactions that may directly affect gene regulation. Promoter capture data was also used to search for evolved interactions that may not be enriched/detected in genome-wide Micro-C sequencing. Analysis of all interaction data allowed us to address specific hypotheses about the role of IDH^{mut}-specific hypermethylated DMRs in AML. These include the association of IDH^{mut} -specific DMRs with active vs. repressed compartments, the propensity of regions to undergo compartment switching when an IDH^{mut} -specific DMR is present, and whether DMRs influence TAD boundary strength or the interaction frequency or expression of genes that directly interact with distal enhancers. At the level of loop interactions, we were able to perform differential analysis between CD34 normal cells, IDH^{wt} AML, and IDH^{mut} cases to assess differential interaction frequencies involving the set of IDH^{mut}-specific DMRs that may indicate altered regulation of target genes. We also performed CUT&Tag for selected histone modifications (H3K4me3, H3K9me3, H3K27me3, H3K27ac) that enabled annotation of regions that form interactions with genes with putative regulatory functions.

3.2 Methods

Patient samples

Banked samples were selected based on mutation status and previous generation of whole-genome bisulfite sequencing data. MicroC libraries were generated from 15 primary AML patients with varying mutation statuses including *IDH1/2* mutations (n=4), *DNMT3a* R882 (n=6), *RUNX1-RUNX1T1* (n=2), and *MLL* rearrangements (n=3). Additionally, we generated libraries from normal donor-mobilized CD34+ cells to use as a comparison for assessing AML-associated changes in 3D architecture.

Micro-C and Capture-C

Library preparation was performed with the Dovetail Micro-C Kit following the manufacturer's protocol[2]. In brief, crosslinking of chromatin is performed in intact nuclei to stabilize the interactions between DNA and associated proteins. Crosslinked chromatin is then digested using Micrococcal nuclease (MNase) to fragment the genome at the sites of naked chromatin. DNA fragments are ligation under dilute conditions to form chimeric DNA molecules. Crosslinking is reversed through heat treatment to remove proteins, and the ligated DNA fragments are subsequently purified and prepped for high-throughput sequencing. The resulting library of DNA fragments represents the spatial proximity of genomic regions within the chromatin.

ChIP-seq for histone modifications

ChIP-seq was performed using ChIPmentation[3] with the following antibodies: H3K27me3 (9733S), and H3K27ac (8173S) from Cell Signaling Technology, and H3K4me1 (ab1012) from Abcam. Sequencing was performed on a NovaSeq 6000 (Illumina, San Diego, CA) to obtain ~50 million 2x150 bp reads. Data were analyzed via adapter trimming with trimgalore and alignment to GRCh38 using bwamem[4]. Normalized coverage for visualization and analysis used the

deeptools "bamCoverage" tool[5], and peaks were called with MACS2[6]. Statistical comparisons with DESeq2[7] used raw fragment counts at peak summits, and visualizations were prepared with Gviz.

RNA-seq analysis

RNA-seq data from AML samples were obtained from the AML TCGA study[8]. TPM values were obtained using kallisto and gene counts were generated using the tximport Bioconductor package in R with the tx2gene option set to accomplish gene-level summarization. Previously published RNA-seq data for normal CD34+ cells generated using the same procedures that were used for the AML samples were obtained as raw sequencing reads from the short-read archive (GSE48846) and processed as described above.

CUT & TAG

CUT&TAG experiments were performed according to published recommendations from the Heinekoff lab[9]. Briefly, cells were fixed with formaldehyde to crosslink proteins to DNA. Membranes were subsequently permeabilized and cells were treated with primary antibodies specific to the nuclear protein of interest. Protein A-Micrococcal Nuclease (pA-MN) is added to DNA and localizes to target proteins via the Fc chain of the primary antibody. The addition of calcium triggers the MN activity, causing it to cleave the DNA and create small fragments of DNA. A transposase enzyme is used to insert sequencing adapters into the fragments, which are then PCR amplified and sequenced.

Data processing and analysis

BWA-MEM was used to align raw Micro-C data, while pairtools was used to record and sort valid ligation events, remove PCR duplicates, and generate pairs .bam files. Cooler tools subsequently convert .bam files to contact matrixes that were used for downstream analysis.

Conformation analysis of Micro-C data utilized multiple published tools including the fanc eigenvector module for defining genome compartmentalization, cooltools insulation and SpectralTAD callers for topologically associated domain definition, and Chicago tools for Capture-C loop identification and mustache for loop calling in whole Micro-C data. Visualization tools include the python coolpup and cooltools modules, the Genova R packages, and the HiGlass server[10–14].

3.3 Results

3.3.1 Micro-C analysis of primary AML samples and mobilized CD34+ cells from normal donors.

We used Micro-C to measure genome interaction landscapes in 15 primary AML samples, including samples with *IDH1/2* mutations (n=4), *DNMT3A* R882 (n=6), *RUNX1-RUNX1T1* (n=2), and *MLL* rearrangements (n=3). Additionally, we generated libraries from purified CD34+ hematopoietic stem/progenitor cells obtained from peripheral blood mobilized healthy normal donors. These were used as normal controls to define AML-associated changes in 3D architecture, particularly as they relate to AML-associated methylation phenotype. Libraries from all samples were generated in triplicate and pooled to reach an average number of read pairs per library of 589 million. The average fraction of non-duplicate read pairs per library was 75%, of which an average of 40% supported *cis* long-range interactions greater than 10kb, indicating optimal library construction. The yield and quality statistics of our libraries suggested that we could run subsequent genome-wide analyses at resolutions as low as 5 to 10kb with significant results. In addition to sequencing full Micro-C libraries, probe capture was performed for all samples and sequenced to enrich promoter-associated interactions, adding further support for point interactions observed in the genome-wide data.

Genome compartments in AML cells remain intact compared to normal CD34+ cells.
We initially analyzed genome compartmentalization in CD34+ normal cells and AML samples. Genome compartments are characterized by two states—the active compartment (compartment A) and the inactive compartment (compartment B). These compartments are defined by large genomic regions that tend to interact more frequently with each other than with regions in the opposing compartment and can be visualized as checkered patterns in the genome-wide interaction matrix from Micro-C or similar experiments. As is implied, the active A compartments tend to include expressed genes while the inactive B compartments are enriched for repressed genes and heterochromatic regions. Using a standard eigenvector (EV) decomposition approach from the Fanc-C suite of analysis tools[15], compartments were called in all AML samples and CD34 normal samples. The returned EV contains a positive or negative value for each region of the genome (here we used 50kb bins) where the sign indicates the two separate compartments. The sign of the EV value was correlated with GC content of the corresponding genomic regions to orient the compartments so that the aggregate compartment regions with higher GC content were called compartment A. For subsequent analysis and figures, EVs were oriented so positive values represent compartment A and negative values represent compartment B.

In agreement with previous findings demonstrating consistent compartment structure amongst cells in similar tissue types[16], we found that genome-wide compartmentalization was largely conserved between the CD34+ stem/progenitor cells and AML. In all samples, the B compartment occupied slightly more of the genome than the A compartment (average 55% B, average 45% A; Figure 1). In addition to partitioning each sample into active vs. inactive compartments, we also compared AML samples to normal CD34+ cells and determined the

frequency of compartment switching in each sample (i.e. regions classified as A compartment in CD34+ cells that are found in the B compartment in AML, A→B; regions found in B compartment of CD34+ cells vs. A compartment in AML, B→A). A→B switches were more common across all AML samples regardless of mutation subclass and involved a mean of 6.1% of the genome (range: 3.2% to 11%) compared to a mean of 3.8% of the genome switching from B→A (range: 1.2% to 5.9%) (Figure 1). When assessing the consistency of switches between samples within the same mutation classification, we found that the locations of compartment switches were highly variable. Overall, the fraction of switches shared across all AMLs was 6% of the union of switched bins, and the average fraction of overlap observed within the *IDH*^{mut}, *DNMT3A*^{R882}, and *MLL* fusion mutational subsets was 11.4% of the subset union of switched bins (10.2% to 12.1%) (Figure 2). We only had two *RUNX1-RUNX1T1* samples in our cohort, which resulted in a 54% fraction of overlap for switched bins between these two samples, however, we predict the fraction of shared bin switches in this subset would be comparable to the others if more samples were present.

Hierarchical clustering of inter-sample Pearson correlation values based on compartment bins representing the top 10% of variance across AML samples revealed a distinct group containing three out of four *IDH*^{mut} samples (including two samples with chromosomal fusions: *RUNX1-RUNX1T1* and a *MLL* fusion), separate from other *IDH*^{wt} AML and CD34+ cells (Figure 3). Although the fourth *IDH*^{mut} did not cluster with the others, it displayed a higher Pearson correlation with other *IDH*^{mut} samples compared to any other mutational subclass of AML as a group. Clustering also revealed a distinct group containing four out of five *DNMT3A*^{R882} samples; the fifth *DNMT3A*^{R882} sample had lower levels of correlation with the other *DNMT3A*^{R882} samples and increased correlation with *MLL* fusion cases despite any clear overlaps

in their mutation profiles (Figure 3). The samples with transcription factor fusions (2 *KMT2A* rearranged cases and 2 with *RUNX1-RUNXT1* gene fusions) formed a separate branch of the dendrogram from the *IDH*^{mut} and *DNMT3A*^{R882}groups (Figure 3). As expected, CD34+ normal cells displayed a high level of correlation forming a unique branch of the dendrogram. Our clustering analysis of variable regions highlights subclass-specific features of compartmentalization that are otherwise difficult to detect on a genome-wide scale and suggests that the strong methylation phenotypes associated with *IDH*^{mut} and *DNMT3A*^{R882}mutations may contribute to subtype stratification at the compartment level.

The active genome compartment is enriched for *IDH*^{mut}-specific hypermethylation.

To better characterize associations between *IDH*^{mut}-specific methylation changes and compartmentalization, we began by intersecting the set of ~4,500 *IDH*^{mut}-specific DMRs and ~4,600 commonly hypermethylated regions with CD34+ defined compartments. 92% of *IDH*^{mut}-specific DMRs and 70% of commonly hypermethylated regions overlapped A compartment bins, respectively, with the remainder of each set falling into B compartment bins. We calculated the enrichment of each DMR set for A and B compartments using methods that controlled for the total compartment size and DMR length. This analysis showed significant enrichment of *IDH*^{mut} DMRs in the A compartment compared with commonly hypermethylated regions (1.89 vs. 1.23; Figure 4), while the opposite was true for the B compartment with common DMRs being overrepresented (0.48 vs. 0.07; Figure 4). These differences were not concentrated within in subset of compartment bins but rather were dispersed across the genome (Figure 5). This finding is consistent with our previous observation that *IDH*^{mut}-specific DMRs are enriched within active enhancer regions, which, based on histone modification data from hematopoietic cells published by our group and others, tend to occur within the active compartment of the genome.

We next surveyed the distribution of compartment scores for all DMR-containing compartments in CD34+ normal cells and *IDH*^{mut} AML to determine whether *IDH*-associated hypermethylation had more subtle effects on compartment-level interactions. Compartment scores summarized over the set of CD34+ defined compartments containing a DMR showed an overall trend towards lower compartment scores for A compartment DMR bins in three of four *IDH* mutant AML samples compared to normal and slightly decreased B compartment DMR bin strengths in two samples compared to normal (Figure 6). For example, a slight downward shift was observed in the distribution of positive compartment scores for three samples along with a slight upward shift in negative values for two samples (less negative values are consistent with weaker B compartment scores) when subsetted for compartments containing hypermethylated DMRs, suggesting these regions may form less robust interactions with genomic regions in their respective nuclear compartments.

***IDH*^{mut}-associated hypermethylation influences compartment assignment.**

We next directly tested the relationship between regions with altered compartment assignments and regions of *IDH*-associated hypermethylation. For each sample, we intersected the bins that had undergone a compartment switch with the set of *IDH*^{mut}-specific DMRs and commonly hypermethylated regions. This analysis identified an average of 7.7% of A->B or B->A switched bins overlapped an *IDH*^{mut} DMR in each sample. When evaluating the rate of DMR overlap based on switch direction, we found that DMRs more often coincided with a switch from the active compartment to the repressed compartment (fraction of A->B switches: mean 6.4%, range: 3.9%-9.6%; fraction of B->A switches: mean 2.8%, range: 2%-3.6%). The same analysis repeated for the commonly hypermethylated regions identified an average of 8% of all switched bins overlapping a common DMR per sample. A to B switches were again over-represented in

this set of common DMR-associated switches, but to a lesser extent than the *IDH*^{mut}-specific DMRs (A->B mean 9%, range: 8%-11.4%; B->A mean 5.6%, range: 4.7%-8%). Given that a similar fraction of switched bins overlap *IDH*^{mut}-specific DMRs as they do commonly methylated regions, these data suggest that hypermethylation associated with *IDH* mutations more often coincides with a switch from the active compartment to the repressed compartment compared with common hypermethylation, which is consistent with the enrichment of *IDH*^{mut}-specific DMRs for the A compartment over the set of commonly hypermethylated regions.

To assess whether the A>B or B>A switching is a potential consequence of DNA hypermethylation, we tested the enrichment of *IDH*^{mut}-specific DMRs for compartment switches in *IDH*^{wt} samples. We noted that the number of total switches and switch direction were variable amongst patients and may account for differences in the frequency of DMR overlaps. We therefore calculated an enrichment of *IDH*^{mut}-specific DMRs for switches with respect to common hypermethylation switches in each patient. Overall, switched bins in *IDH*^{mut} samples had a modest increase in enrichment for *IDH*^{mut}-specific DMRs compared with *IDH*^{wt} samples (average enrichment 0.64 vs. 0.56, respectively; Figure 7, upper left). When looking specifically at the set of A>B switches across samples, the enrichment of *IDH*^{mut} -specific DMR overlap was similarly pronounced in *IDH*^{mut} cases compared with *IDH*^{wt} cases (average enrichment 0.68 vs. 0.6, respectively; Figure 7, upper right). We did not note any clear differences in the enrichment of *IDH*^{mut} DMRs for B>A switches between *IDH*^{mut} and *IDH*^{wt} samples (average enrichment 0.51 vs. 0.53 respectively; Figure 7, lower left). These data suggest that *IDH* mutant-associated hypermethylation may influence compartmentalization not only by weakening interactions within the active compartment but also by inducing compartment switching in certain regions.

IDH mutant samples were next compared to CD34+ normal cells both in aggregate and individually to assess for changes in compartment strengths before and after switching (i.e. in normal vs malignant states). On average, bins with a DMR that underwent a compartment switch had lower absolute compartment scores in CD34+ cells, indicating these regions had weaker interactions with their respective compartments across the genome. Likewise, the 'switched' compartment score in the IDH^{mut} AMLs was also low, with a mean change in compartment score of < 1.5 (inside red lines; Figure 8). We characterized these events as weak A > weak B or weak B> weak A switches. However, we also observed events characterized by weak > strong compartmentalization switches and strong > weak switches (average change in compartment score of > 1.5 (outside red lines); Figure 8). Although we did not observe an overall correlation of DMR-containing bins with strong vs. weak switches, there were examples of strong switches associated with DMRs that may individually represent the effects of IDH^{mut}-associated hypermethylation on compartmentalization, warranting further analysis in follow up studies.

These patterns were observed in individual samples to varying degrees, representing sample-to-sample heterogeneity in compartment switching while maintaining overall trends (Figure 9). In only one sample did we observe strong B > strong A switches (average change in compartment score of > 1.5), all of which occurred on chromosome 19 suggesting that unique constitution/regulation/3D organization associated with chr19 may be present in that patient (Figure 9). This analysis of *IDH* mutant AML samples suggests that DMR-associated switches may occur at equivocal regions, for instance, compartment bins that are weakly in the A or B compartment in normal cells and may naturally fluctuate between the two, rather than at regions strongly associating with one compartment vs the other. Though not systematically measured, a manual review of DMR-associated switches supported this hypothesis, showing a prevalence of

these events at the borders of A and B compartments in normal CD34+ cells, where values inherently tend to be closer to zero.

Gene expression changes in DMR-resident switch regions correlate with compartment activity.

To better understand the functional consequences of compartment switching associated with *IDH*^{mut} hypermethylation, we summarized gene expression levels and AML-associated expression changes based on localization within compartment bins that had undergone a switch. Bulk RNAseq data generated for three of four *IDH*^{mut} patients and six CD34+ normal donor samples were used to obtain expression values for the sets of genes contained within DMR-associated switch regions; switched regions that harbored commonly hypermethylated DMRs were used as controls. Individual sample analysis showed genes within both *IDH*^{mut}-specific and common hypermethylated DMR-associated regions had a bimodal pattern of normalized expression (Figures 10 and 11). Consistent with previous observations, the regions associated with *IDH*^{mut}-specific DMRs were enriched for highly expressed genes compared with the common hypermethylated set[1]. In all cases of A to B switches, an overall shift toward higher expression was observed in the AML sample compared with CD34+ normal cells for these sets of genes (Figure 10). Though statistically difficult to quantify, this shift appeared greater for the DMR-associated A>B switch genes than the background set. A similar analysis was conducted for DMR-associated B>A switches, revealing a modest shift in gene expression to the right for AML samples compared with CD34+ normal cells that was mirrored in the background set (Figure 11). Expression changes were next evaluated on a gene-by-gene basis between CD34+ cells and *IDH*^{mut} AML samples. Interestingly, although expression differences were variable, some genes displayed consistent expression changes across samples that suggest the tendency for hypermethylation associated with *IDH* mutations to influence their expression (Figure 12).

Genes consistently upregulated in association with B>A DMR switches included previously reported cancer-related factors *COP1*[17], *FTO*[18], and *HGF*[19], while genes consistently downregulated in association with A>B DMR switches across *IDH* mutant AMLs included known hematopoietic factors *AKT3*[20], *MN1h*[21], and *HDAC9*[22]. We did not observe any consistent examples of upregulation for A>B switches or downregulation associated with B>A switches, that is, for all consistent changes (observed in at least two of three cases), upregulation correlated with a switch into the more active compartment of the nucleus while downregulation correlated with a switch to the repressed compartment. Taken together, these results suggest that IDH^{mut}-specific hypermethylation may impact factors governing nuclear compartmentalization, and in certain cases affect regional gene expression without directly targeting the regulatory sequence of a gene.

3.3.2 Relationship of IDH^{mut} associated hypermethylation with topologically associated domain architecture.

We were next interested in assessing the relationship between topologically associated domains (TADs) and DNA methylation in normal hematopoietic progenitors and primary AML samples. TADs are distinct, self-contained genomic regions that can promote interactions necessary for gene expression and prevent spurious interactions between neighboring regulatory units[23]. By definition, regulatory elements within a TAD have more frequent interactions with genes contained in that domain vs. adjacent regions. Broad topological domains tend to be conserved across species, suggesting their functional importance in genome organization while sub-TADs show a greater degree of heterogeneity across species and are thought to be consistent with tissue-specific expression patterns[24]. CTCF is a critical factor in TAD formation and maintenance, often demarcating the edges of strong domains where it is bound[25]. In certain disease contexts, loss of TAD insulation and concomitant gene expression changes have been

correlated with the inhibition of CTCF binding at normal domain boundaries[26]. Although the enrichment of CTCF binding at TAD boundaries has been well established, the epigenetic patterns governing the dynamics of CTCF binding are less well understood. Here, we investigate the potential influence of DNA methylation on CTCF-mediated TAD boundary maintenance in primary AML, along with the downstream consequences on gene regulation.

Large domains are conserved in AML cells, while local topology exhibits AML-specific features.

We started by calling TAD boundaries in CD34+ normal cells and individual AML samples using different methods including Arrowhead, the gold standard approach for Hi-C analysis developed by the Aiden lab, as well as the HiCExplorer, SpectralTAD, and cooltools domain calling modules[10,11,27]. Given the significant overlap in boundary calls across methods, we proceeded with the set of TADs identified by the cooltools suite, which has been suggested for analyzing Micro-C data. Cooltools uses a sliding diamond window method to generate an insulation profile for each contact matrix. Moving across the length of the genome, contacts within the window for each position are summed and inspected for local minima consistent with decreased contact frequencies between upstream and downstream loci. These highly insulated positions are then assigned a boundary strength using a peak prominence model and thresholded as weak vs. strong boundaries as shown below for one example region encompassing the MYC locus in CD34+ normal cells (Figure 13).

Using rigorous thresholding parameters for the cooltools insulation calculations, we found the total number of strong boundaries (aka TAD boundaries), as well as the distance between boundaries, was largely consistent amongst CD34+ normal cells and primary AML samples with varying mutational profiles at a lower resolution (100kb sliding windows), likely representing

domain architecture consistent with blood cell identity (Figure 14, right panel). As expected, the number of TADs detected increased as the resolution increased (30kb and 50 kb sliding windows), detecting local structures demonstrating sub-TAD architecture (Figure 14, left and middle panels). What was surprising, however, was that the increased resolution also revealed unique features of TAD compartmentalization in AML samples compared to normal. Notably, AML samples on average had an increased number of TAD boundaries compared with normal cells, and *IDH* mutant samples consistently exhibited a higher number of TAD boundaries than any other AML subtype when using a 30kb sliding window. The increase in local TAD formation we observed in AMLs vs. normal cells may represent a tendency for transformed cells to adopt new chromatin interaction patterns that support a proliferative phenotype, or perhaps have no functional consequence but are the result of underlying mutations and epigenetic forces. It will be important to investigate this possibility in future studies linking the mechanism of 3D genome organization with AML phenotypes beyond the differential methylation patterns studied in this thesis.

We were struck by the significant increase in TAD boundaries observed in *IDH* mutant samples compared to other AMLs and wondered whether this may be related to *IDH* mutant hypermethylation. However, no clear association was identified between the TAD boundaries and IDH^{mut}-specific DMRs via an overlap analysis. This suggests that most DMRs do not directly affect TAD boundary formation or stability but does not formally rule out the possibility of longer-range effects or other downstream consequences of hypermethylation that could increase TAD structures. It will be interesting to follow up on this finding and explore possible explanations related to cellular metabolism, inhibition of histone-modifying complexes, oxidative stress, etc., either individually or in combination with DNA hypermethylation.

CTCF binding events are enriched at strong boundaries in normal CD34+ cells.

Since TAD formation is known to be facilitated by the structural DNA binding protein CTCF, we wanted to evaluate how well CTCF binding correlated with the TAD boundaries identified in our samples. For this analysis, we used publicly available ChIPseq data from CD34+ normal cells and a consensus set of TAD boundary regions identified in the Micro-C data from CD34+ normal samples. We first compared CTCF binding signal with the boundary scores for each overlapping interval. Although CTCF ChIPseq signal was observed to varying degrees across the range of boundary scores, it was significantly enriched in regions with the highest scores (i.e. greatest insulation strengths) (Figure 15). Interestingly, when we looked at the set of TAD boundaries containing an *IDH*^{mut}-specific DMR, we found that strong boundaries (those clearing a cooltools statistical threshold of 0.67 boundary score) were enriched for *IDH*^{mut}-specific hypermethylation compared to the set of commonly hypermethylated DMRs (Figure 16, right of the threshold line). In contrast, weaker boundaries showed little difference in the frequency of overlap between the two sets (Figure 16, left of the threshold line). This observation suggested to us that differential methylation in *IDH* mutant samples may uniquely affect a set of boundaries regulating and or maintaining a set of interactions important for normal cell development. We therefore wanted to define a set of regions with CTCF occupancy that could be tested for changes in boundary strength associated with differential methylation.

The two independent thresholding models used to identify the set of strong boundaries (Li, Otsu) nicely capture the boundary strength cutoffs for clear CTCF enrichment (Figure 15), the Li threshold being more permissive while the Otsu threshold is more stringent. Using each threshold value, we identified a set of strong boundaries containing statistically significant CTCF binding peaks in CD34+ normal samples. The two sets contained a similar number of

boundaries, that is, the permissive threshold added only 6% to the total number of CTCF resident boundaries, therefore we used the more permissive Li threshold to define an inclusive set of 1,389 CTCF boundaries for subsequent TAD analysis (see below).

TAD boundaries containing hypermethylated CTCF binding sites are weakened in *IDH* mutant AML compared to normal cells.

To measure potential changes in TAD boundaries influenced by IDH^{mut}-specific hypermethylation, we intersected the boundary CTCF peaks with IDH^{mut}-specific DMRs (henceforth referred to as CTCF-DMR boundaries) and compared the strength of resultant boundaries between CD34+ normal cells and IDH^{mut} AML samples. CTCF-associated boundaries can represent strong point interactions between distal regulatory elements, consistent with corner loops or focal maximal falling off the diagonal in contact matrices, as well as borders between neighboring regions with diffuse by highly interactive chromatin. We discuss findings from our detailed analysis of the point-to-point interactions facilitated by CTCF in a later section. Here we report on the analysis of all TAD boundaries in aggregate.

Although results were variable across individual *IDH* mutant cases, average boundary strengths at the set of CTCF-DMR bounds were significantly decreased in the set of *IDH* mutant AML samples compared to CD34+ normal cells (Figure 17, left panel). We also assessed boundary strengths for the same regions in IDH^{wt} samples to help determine if the observed differences were specific to *IDH* mutant AML or generalizable across all AML samples in our data set. We found a similar level of sample-to-sample heterogeneity for IDH^{wt} samples, however, the trend towards decreased scores for CTCF-DMR boundary was not observed to the same extent for the other mutation subtypes (Figure 17, right panel). The same analysis was repeated with the control set of commonly hypermethylated regions, revealing a more even distribution of both

strengthened and weakened boundaries across AML subtypes when compared to normal cells. Taken together, these results suggest that *IDH*^{mut}-specific hypermethylation may uniquely affect genome regulation at the level of topological domain architecture, perhaps disrupting strong insulator regions through methylation-dependent inhibition of CTCF binding. Additional experiments measuring CTCF occupancy at these regions in *IDH* mutant AML samples will be important to follow up on this hypothesis.

To better quantify the differences between mutation subtypes, we performed a clustering analysis of CTCF-DMR boundary scores across all AMLs. Although the differences were relatively subtle, we observed an overall decrease in strength across the set of CTCF-DMR boundaries for AMLs compared to CD34+ normal cells. In this cohort, *IDH* mutant cases did not form a distinct clustering, suggesting a degree of heterogeneity between these samples consistent with overall heterogeneity in AML samples (Figure 18). We did, however, notice that *DNMT3A*^{R882} samples formed a distinct cluster from other mutation subclasses, which may capture a unique TAD phenotype associated with *DNMT3A*^{R882} associated hypomethylation. We previously described a relationship between regions of DNA hypermethylation in *IDH* mutant samples and high methylation turnover rates, where we believe active methylation and de-methylation occur in a balance to maintain a baseline state. The fact that other AMLs with altered methylation activity may display a structural phenotype in the same set of regions is an interesting observation that could be the subject of future investigations. The same analysis was repeated for commonly hypermethylated CTCF boundaries demonstrating greater dispersion of *DNMT3a*^{R822} subtypes compared to the analysis of CTCF-DMRs, suggesting that regions associated with features of *IDH* hypermethylation may stratify a consistent phenotype in certain patients with altered methylation.

TAD with hypermethylated CTCF boundary elements exhibit reduced insulation in *IDH* mutant AML compared to normal cells.

We next looked to see if there were any high-level patterns unifying the set of weakened boundaries or the intervening region of the associated TADs that could suggest potential consequences of insulator hypermethylation. A reduction in insulation suggests that the local interaction frequencies may be altered in these samples, therefore we assessed intra- and inter-TAD interactions across the set of CTCF-DMR boundaries in normal samples and *IDH* mutant samples. As expected, interaction frequencies between neighboring TADs decreased as a function of distance, consistent with our understanding of genome organization, however, we noticed some sample-specific differences in the immediate TAD and neighboring TADs based on methylation status. *IDH* mutant samples had consistently reduced interaction frequencies in the resident CTCF-DMR TAD (intra-TAD) compared to normal samples, although interaction frequencies were generally decreased for all CD34+ defined TADs in these samples (Figure 19). Changes in inter-TAD interactions were less clear for these samples, overall exhibiting consistent interaction frequencies for n+1, n+2, and n+3 nearest neighbor TADs with normal cells (Figure 19). We see this phenomenon hold up when visualizing the contact matrix for implicated regions including an example on chromosome 8 that demonstrates how a decrease in insulation at a CTCF-DMR TAD boundary is accompanied by loss of interactions, including focal/point interactions, for the adjacent region (Figure 20).

Lastly, we wanted to summarize gene expression in the TADs containing a CTCF-DMR in one or both boundaries to assess for any patterns in the set of affected regions and determine if decreased insulation is correlated with changes in gene expression. Based on the intersection of annotated genes with the set of affected TADs, we observed a significant increase in average gene expression for the set of all genes contained within affected TADs compared to unaffected

TADs (Figure 21). Similarly, when we summarized genes in TADs with commonly hypermethylated CTCF sites in either or both boundaries, average expression was significantly higher than unaffected TADs and comparable to CTCF-DMR-affected TADs. Though very coarse, this analysis suggests that while hypermethylation tends to be focal, its localization may be reflective of the broader activity in a region—in this case, TADs with more active transcription. At this level of summarization, we did not observe any clear differences in gene expression associated with AML mutation status, suggesting that CTCF motif hypermethylation in insulator regions may not consistently alter the regulation of genes genome-wide. However, hypermethylation of regulatory elements in specific insulators may correlate with gene expression changes that are relevant for leukemia cell proliferation and will be important to follow up on in additional analyses.

3.3.3 Relationships between *IDH*^{mut} -specific hypermethylation and chromatin loops

In addition to relationships between *IDH*^{mut}-specific hypermethylation and higher-order genome organization, we were also interested in the local effects of methylation on interactions between gene promoters and regulatory elements. One of the most striking observations from our genome-wide methylation analysis in a mutationally diverse AML cohort was the enrichment of active enhancer regions in *IDH* mutant DMR[1]. Our. published study began to explore the impact of these changes on genome interactions using Hi-C data from two AML samples without *IDH* mutations generated in our lab and one published Hi-C data set from CD34+ normal cells. We used these data to link enhancer DMRs (eDMRs) to their cognate genes but recognized the limitations of this analysis given that we did not have data from *IDH* mutant AML samples. To expand our characterization of the functional consequences of enhancer hypermethylation, we used MicroC and CaptureC data from CD34+ normal cells and *IDH* mutant and wild-type

primary AML samples to identify enhancer-promoter interactions that may be uniquely influenced by *IDH*^{mut}-specific DMRs.

IDH^{mut}-specific DMRs overlap chromatin loops to a greater extent than commonly hypermethylated regions in AML

We first began by calling loop interactions from CaptureC data using the CHiCAGO (Capture Hi-C Analysis of Genomic Organization) pipeline[28]. This approach uses statistical models to call loop interactions in a sparse data set where information is concentrated around promoter-defined bait regions. We observed variability in the total number of CHiCAGO loops identified per sample at a resolution of 10kb with ~22,000 to ~78,000 CHiCAGO interactions identified in an individual AML sample and ~120,000 identified in CD34+ normal cells (Figure 22). A simple intersection of sample-specific loops with the set of *IDH*^{mut}-specific DMRs and the control set of commonly hypermethylated regions demonstrated a proportional level of variability in the number of resulting DMR loops with the overall number of loop calls. The sample-to-sample variability in loop calls could be due to the transient nature of loop formation, or inherent differences in primary samples (patient-specific features, tissue banking, cell handling, etc.), sequencing depth, and/or the under-sampling nature of most capture datasets and analyses at the high resolution of loop interactions are likely more sensitive to variability due to such factors. For example, we noticed the large jump in loop counts for CD34+ cells was correlated with an increase in sequencing coverage in those samples.

We next investigated the relationship between *IDH*^{mut}-specific DMRs and chromatin loops by measuring overlaps between these regions and comparing this to commonly hypermethylated regions. Using this approach, we found that *IDH*^{mut}-specific DMRs were uniformly enriched for point interactions compared to background hypermethylation across primary samples and CD34+

normal samples, although to varying degrees (Figure 23). The sample-to-sample variability in enrichment is, again, likely due to technical aspects of the assay and would require greater sequencing depth per library and additional samples to adequately acknowledge any mutation-specific patterns in the enrichment level if present. The overall finding, however, is consistent with our previous annotation of *IDH*^{mut}-specific DMRs being uniquely enriched for active enhancers; regulatory elements that tend to be situated opposite gene promoters in loop interactions.

IDH^{mut}-specific DMRs occur in a set of focal interactions.

Having identified enrichment for *IDH*^{mut}-specific DMRs in point-to-point chromatin interactions, we were next interested in measuring potential differences in the interaction frequencies of eDMR loops between AML subtypes. Taking the full union of capture loops identified across AML and CD34+ normal cells, we clustered samples based on the normalized interaction frequencies for all loops intersecting an eDMR. Our analysis showed considerable heterogeneity across AML subtypes, however, the range of normalized interaction frequencies for most loops in this set was very narrow, potentially impeding the detection of any subtype-specific supervision (Figure 24). Similar results were observed when we clustered samples based on interaction frequencies of loops overlapping the set of commonly hypermethylated regions.

To further assess the possibility of altered interaction dynamics between eDMRs and distal regions in hypermethylated AML, we performed aggregate peak analysis (APA) on the set of CD34+ eDMR loops in individual samples with and without *IDH* mutations. APA quantifies the enrichment of a set of interactions in aggregate by summing up contact frequencies in submatrices of the genome-wide contact matrix, each centered on a peak pixel representing the interaction of two distinct loop anchors. The focal enrichment for the set of interactions in

aggregate manifests as larger values at the center of the APA plot. For this analysis, we used the genomic coordinates of loops defined using CaptureC data to highlight a set of enriched interactions in the context of the complete micro-C data. At a resolution of 10kb, APA plots show a clear enrichment for eDMR loops in all samples including normal CD34+ cells, *IDH* wild-type, and *IDH* mutant cases (Figure 25, column 1). When compared with the set of CD34+ loops containing a commonly hypermethylated region in a loop anchor, the *IDH*^{mut}-specific DMR loops showed increased signal intensity, suggesting they may coincide with a robust set of focal interactions in AML compared to commonly methylated regions that are more diffuse (Figure 25, column 2).

IDH^{mut}-specific hypermethylation influences CTCT-bound chromatin loops in AML compared to normal cells.

Given the subtle differences in interaction frequencies across the full set of eDMR loops, we decided to focus our analysis on the loops where an eDMR intersected an established CTCF site. DNA methylation is known to influence CTCF binding[29] and we therefore hypothesized that these loops may be preferentially affected by *IDH*^{mut} associated hypermethylation. For this analysis, we began by identifying loop anchors containing a CTCF-binding site overlapped by an eDMR (CTCF-DMR). Although the breadth of impact a DMR has on adjacent regions isn't fully clear and may be an interesting study in and of itself, we specifically wanted to identify regulatory elements whose hypermethylation could directly impact known binding factors. Measurements of direct CTCF binding in normal CD34+ cells are limited, so we used ChIPseq data from primary AML samples to infer a set of consensus interactions likely mediated by CTCF binding in AML. Using a union set of binding peaks from a diverse set of primary AML samples, we intersected CTCF binding events with *IDH*^{mut}-specific DMRs to define a set of 866 CTCF-DMRs for downstream analysis. These CTCF-DMRs were subsequently overlapped with

CD34+ normal cell CaptureC loop anchors to define interactions with a CTCF-DMR specifically in the anchor opposite the bait probe, identifying 1,080 interactions. We again performed enrichment analysis for interactions anchored on a CTCF-DMR region across AML samples. Although a statistical difference in the interaction frequency of CTCF-DMR loops was difficult to define between *IDH*^{mut} samples and CD34+ normal cells, the overall interaction frequency of CTCF-DMR resident loops was decreased in *IDH*^{mut} AML compared to normal (Figure 25, column 3). There was a similar decrease in the enrichment of CTCF-DMR resident loops in *IDH*^{wt} AMLs compared with normal cells, however, we did not observe any clear differences between the AML subtypes (Figure 26, column 3). We reason that group-level differences in interaction frequencies may be difficult to detect at the resolution of our data, if present, or there may be a feature-specific reason we don't see a difference between *IDH*^{mut} samples and other AMLs—for example, if the enhancers marked by hypermethylation are depleted of CTCF or other methylation sensitive TF binding sites, we may not expect the difference in loop formation to be altered dramatically. It is also possible that if differences do indeed exist, the directional effects of hypermethylation on interaction frequency for this set of enhancers are heterogeneous, and when assessed in aggregate appear to be neutral.

We also recognize that the set of CTCF peaks used in this analysis doesn't fully account for CTCF binding events in CD34+ normal cells and may miss CTCF-mediated interactions present only in non-malignant cells. In the absence of CTCF ChIPseq data from normal cells, we used data from a previous study that performed ChIAPET to enrich for three-dimensional genomic interactions bound by CTCF in CD34+ normal progenitors. The assay is designed to enrich a set of proximity ligation events bound by CTCF but does not necessarily indicate which loop anchor harbors the CTCF binding site in the same way that ChIPseq detects discrete CTCF binding

events. We used this set of loop coordinates to broadly compare the signal of CTCF loops in normal CD34+ cells and AMLs, given our inability to subset these loops based on a direct overlap of CTCF binding events with a DMR. In aggregate, the set of CD34+ CTCF-associated interactions were the most enriched in the MicroC data from CD34+ normal samples, while all *IDH*^{mut} AML samples consistently showed lower enrichment scores for these foci (Figure 25, column 4). Again, we did not observe any subtype-specific patterns in CTCF loop enrichment, suggesting that normal CTCF-bound interactions may be marginally reduced in AML cells compared to CD34+ cells, irrespective of mutational background (Figure 26, column 4). Although this analysis does not assess for novel CTCF-associated interactions present in AML, the fact that we observe an overall decrease in interaction frequencies mediated by CTCF binding does suggest that underlying differences in CTCF-associated interactions may be present in AML cells.

IDH^{mut}-specific DMRs are associated with diminished promoter interactions in AML compared to normal donor cells.

In addition to surveying for AML-associated changes in normal hematopoietic stem/progenitor interactions, we also wanted to detect any novel interactions in *IDH*^{mut} AMLs that may be the result of altered methylation (i.e. an *IDH*^{mut} -specific interaction landscape). To address this, we used CHiCdiff, a differential looping module offered as part of the CHiCAGO pipeline, to run direct comparisons between *IDH*^{mut} AMLs and CD34+ normal cells or other *IDH*^{wt} AMLs. CHiCdiff reported 21,126 statistically significant differential loops between *IDH*^{mut} AML and CD34+ cells, representing both up and down-regulated interactions in AML with respect to normal (43% up, 57% down; weighted-padj < 0.05; Figure 27, left panel). Of the total set, 6,490 differential interactions intersected an eDMR in the loop anchor opposite the promoter capture probe. Interactions upregulated in *IDH*^{mut} AML were greatly under-represented in this set of

differential interactions (8% up in *IDH*^{mut} AML) with most differential loops exhibiting decreased frequencies compared to CD34+ normal cells (92% down in *IDH*^{mut} AML; Figure 27, right panel). When we called differential loops between *IDH*^{wt} samples and CD34+ normal cells, only 4,895 loops were identified (56% up, 44% down; weighted-padj < 0.05; Figure 28, left panel). Of the differential loops, 1,139 intersected an eDMR opposite the bait anchor. Similar to the set of differential *IDH*^{mut} eDMR loops, these loops were largely downregulated in *IDH*^{wt} samples vs. CD34+ cells but to a slightly lesser degree (14% up, 86% down).

We also intersected the sets of differential loops with commonly hypermethylated regions and found the number of statistically significant loops was slightly reduced for this set of loops in both *IDH*^{mut} and *IDH*^{wt} AML (4,667 vs. 894, respectively; weighted-padj < 0.05). Of these loops, the distribution between up and down-regulated interactions was also skewed towards decreased interactions compared to normal, but to a lesser degree than *IDH*^{mut} eDMR loops (17% and 38% increased interactions in *IDH*^{mut} and *IDH*^{wt} AML, respectively; 83% and 62% decreased interactions in *IDH*^{mut} and *IDH*^{wt} AML, respectively). This finding suggests that while common AML-associated hypermethylation occurs within downregulated interactions, *IDH*^{mut}-specific hypermethylation marks a set of interactions that are more consistently downregulated in AML.

Differential interactions coinciding with *IDH*^{mut}-specific CTCF-DMRs identify divergent phenotypes in *IDH*^{mut} and *IDH*^{wt} AML

Finally, narrowing our focus only to differential interactions with a differentially methylated CTCF binding site in the regulatory anchor, we found 2,664 CTCF-DMR loops and 1,307 commonly hypermethylated CTCF loops reaching statistical significance (weighted-padj < 0.05). In both cases, the fraction of differentially upregulated loops in *IDH*^{mut} AML was very low compared to the number of interactions representing differentially decreased interactions

compared to normal cells (UP in *IDH*^{mut}: 4% CTCF-DMR diff loops, 3% common hyper diff loops; DOWN in *IDH*^{mut}: 96% CTCF-DMR diff loops, 97% common hyper diff loops) (CTCF-DMRs shown in Figure 27, right panel). Upon inspection of the list of gene targets, we found a few candidate genes whose significance in AML development and prognosis has been previously described in the literature. For example, target genes involved specifically in down-regulated CTCF-DMR interactions included *MED24,* and *RTEL1,* while genes involved in down-regulated interactions associated with either CTCF-DMRs or commonly hypermethylated CTCF sites included *KDM6B* and *TP53.* While the interaction with any of these genes may be representative of regulatory dynamics in AML, we were interested in the finding that *MED24*, a component of the mediator complex was found to interact uniquely with down regulated interactions in *IDH*^{mut} AML. As a critical component tethering enhancer-promoter interactions and facilitating transcriptional activation, altered regulation of *MED24* could conceivably have impacts on the genome-wide regulation of the 3D genome and gene expression.

Interestingly, when we performed this same analysis on the subset of CTCF-DMR differential loops in *IDH*^{wt} samples compared to CD34+ cells, the distribution between up and down-regulated loops was flipped (Figure 28, right panel). That is, differential interactions intersecting a CTCF-DMR were more often upregulated than downregulated in *IDH*^{wt} samples. This finding supports our hypothesis that interactions coinciding with *IDH*^{mut} associated hypermethylation tend to be robust, possibly a feature of active or tightly regulated enhancer-promoter interactions and suggests that the dynamics of these interactions may be influenced by the methylation phenotype of the cells.

3.3.4 *IDH*^{mut}-specific enhancer DMRs form loops with highly expressed genes in hematopoietic cells

To expand our previous observation that *IDH*^{mut}-specific eDMR regions interact with highly expressed genes in AML, we analyzed the expression of genes in chromatin loops identified from the Capture-C data from CD34+ normal donor cells and AML samples. We started by interrogating CHiCAGO loops to define two sets of target genes: those opposite an *IDH*^{mut} -specific eDMR and those opposite a commonly hypermethylated DMR; regions with both an *IDH*^{mut}-specific DMR and a common hypermethylated DMR were removed the from analysis, while genes interacting with both eDMR anchors and commonly hypermethylated anchors were retained. Considering all consensus loops detected in CD34+ normal donor samples that intersected an annotated gene, we identified 2,757 targets opposite an eDMR and 2,147 gene targets opposite a commonly hypermethylated anchor. When filtering only expressed gene targets based on RNAseq data from a large cohort of AML samples (n=152), we identified a union set of 1,955 eDMR target genes vs 1,439 genes opposite a commonly hypermethylated loop anchor. Consistent with our previously published results, the average normalized expression level of eDMR targets was increased compared to genes opposite a commonly hypermethylated anchor (7.85 vs 7.13 normalized counts respectively), a feature reflected in the overall right shift in the distribution of normalized expression for eDMR targets (Figure 29, left panel). We noted an overlap of 333 genes between the two sets, which had an average normalized expression level and distribution more consistent with the full set of eDMR targets (CD34+ avExp: 7.7 normalized counts) rather than the full set of commonly hypermethylated targets. Increased gene expression for these targets suggests that the influence of eDMR interaction is dominant to the effect of interactions with commonly hypermethylated regions. Analysis of loop interactions using MicroC data corroborates our previous findings of increased eDMR target expression

compared to commonly hypermethylated enhancer targets but improves the overall resolution, lending greater specificity in linking enhancers to their cognate genes.

Having identified target genes that interact with putative enhancer regions prone to hypermethylation in *IDH*^{mut} AML, we next wanted to characterize the consequences of *IDH*^{mut}-specific enhancer hypermethylation on gene expression. Accordingly, we assessed the normalized expression of the same set of eDMR targets and commonly hypermethylated enhancer targets in *IDH*^{mut} and *IDH*^{wt} AMLs. The average normalized expression for eDMR targets in *IDH*^{mut} samples was decreased compared to CD34+ normal cells to a greater degree than the set of commonly hypermethylated enhancer targets, while the average expression of all measurable genes slightly increased in *IDH*^{mut} AML compared to normal cells. The distribution of average normalized expression for both sets was slightly left-shifted in *IDH*^{mut} samples compared to normal cells, and we noticed rather than displaying the bi-modal curve observed in CD34+ samples, expression in the *IDH*^{mut} samples exhibited an intermediate peak for both gene sets. Though not dramatic, the observed trend toward decreased expression for both sets of genes, suggests that the regulation of certain targets may be impacted by hypermethylation of interacting enhancers (Figure 10b). This provides evidence that highly expressed genes may display some diminished expression in the setting of enhancer hypermethylation while still being expressed.

Highly expressed genes involved in differential CTCF-eDMR loops have decreased expression in AML compared to normal cells.

We were also curious whether genes involved in differential loop interactions, or *IDH*^{mut}-specific loops, had any unifying features. As reported above, interactions with decreased frequency in *IDH*^{mut} samples compared with CD34+ cells were present in the full set of differential loops but

significantly over-represented when looking at the subset containing an eDMR in the regulatory anchor (92% DOWN in *IDH*^{mut} vs. 8% UP in *IDH*^{mut}). Overall, genes participating in a differential loop interaction with a coinciding eDMR in the regulatory anchor, exhibited a range of normalized expression values and in aggregate didn't display any significant trends in expression level differences between *IDH*^{mut} samples and CD34+ normal samples. When looking specifically at the top 50% of genes based on expression, however, we observed a notable decrease in the average value and distribution of expression for target genes, while the bottom 50% of expressed genes displayed comparable values between *IDH*^{mut} and CD34+ normal cells (Figure 11). Further filtering of the set of interactions to include only target genes involved in a differential interaction with a CTCF-eDMR in the regulatory anchor strengthened the observed trend of decreased expression in *IDH*^{mut} samples compared to CD34+ normal samples when summarizing the top 50% of expressed targets while remaining consistent for the bottom 50% of targets (Figure 12). This finding is consistent with our previous observations that highly expressed genes display a unique relationship to *IDH*^{mut}-specific hypermethylation compared with instances of common hypermethylation and may mark a set of regulatory interactions that may become hypermethylated as a result of active epigenetic and transcriptional regulation. In addition, the expression data suggests that hypermethylation at enhancers of highly expressed genes may contribute to decreases in expression for certain targets.

In addition to looking at the overall trends in eDMR target gene expression, we also wanted to look for candidate genes with consistently increased or decreased expression across *IDH* mutant samples compared to normal cells. To think that all eDMRs impact target gene expression in a meaningful way is unlikely, and therefore we wanted to identify concrete, believable examples of enhancer hypermethylation coinciding with the decreased (or increased) expression of its gene

target/s. Narrowing our focus to differential chromatin loops with a hypermethylated CTCF binding site in the regulatory anchor, we found several candidate hematopoietic genes with observable changes in expression. Many of these genes had reduced expression compared to normal cells, including *MED24*, *RETL1*, *RRAS*, and *PTBP1* (Figure 13). The role of these genes in hematopoiesis and leukemogenesis has been discussed in previous studies[30–33], but whether the expression changes observed in relation to enhancer hypermethylation contribute to an oncogenic phenotype is unclear and could be an interesting focus for future experiments. Again, we highlight *MED24* as an interesting candidate for follow-up studies, given its role in facilitating physical interaction between distal regulator elements. It would be interesting to test if any of the phenotypes observed in our study are dependent on *MED24* based on genetic knockdown or KO studies.

3.4 Acknowledgments

A special thank you to Reza Ghasemi and Mohamed Mahgoub for the generation and processing of AML and normal donor Micro-C and Capture-C data, and to Heidi Struthers for generating all CUT&TAG libraries for primary AML samples in this study.

3.5 Figures

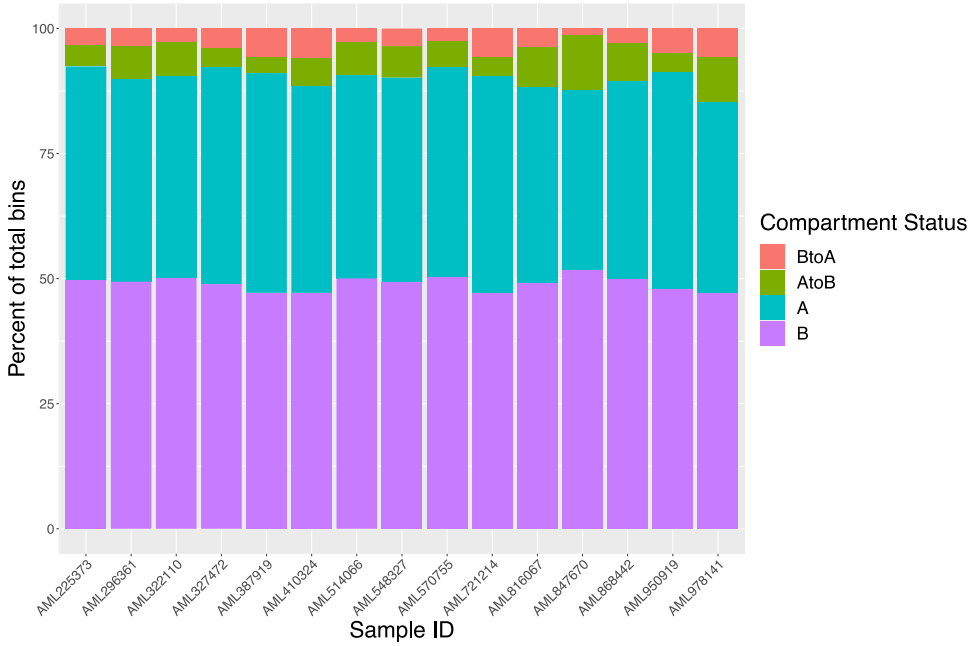


Figure 3.1. Summary of genome compartmentalization across AML samples using a 50 kb sliding window.

Each bar contains the proportion of a patient's genome in the A and B compartments (blue bars and purple bars, respectively), as well as the proportion of genomic windows that have undergone a switch from A to B or B to A with respect to CD34+ normal donor cells (green bars and pink bars, respectively).

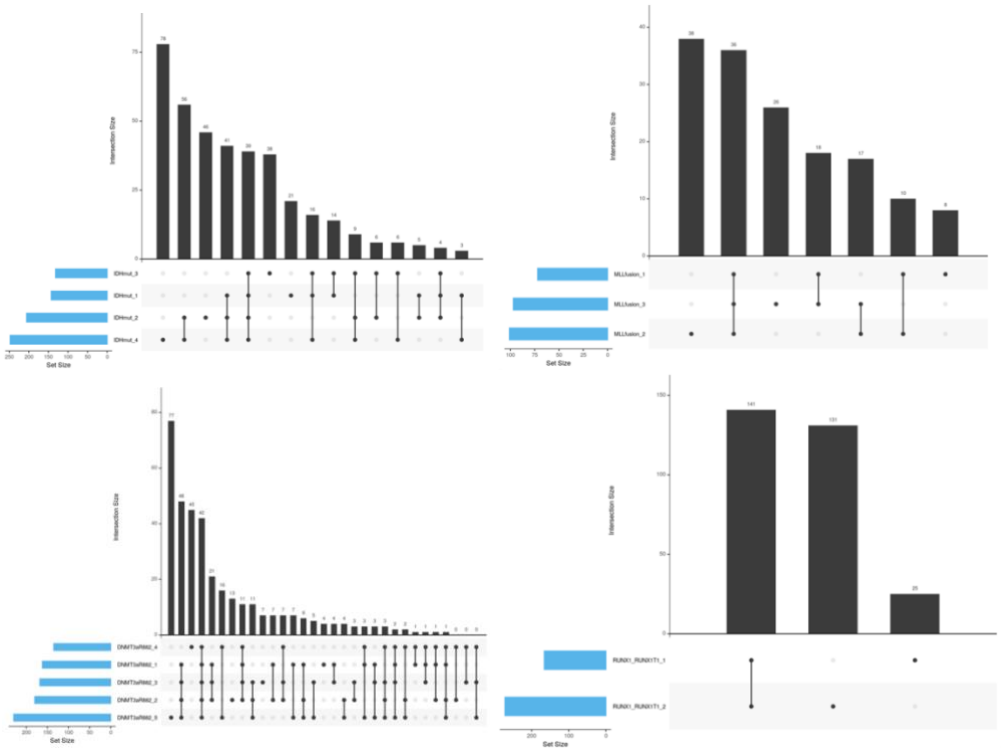


Figure 3.2. Overlap of switched bins within subtypes.

The height the of black bars represents the total number of shared switch regions in the indicated combination of samples as shown in the multi-sect line plot below. The total number of switched regions identified in a given sample is shown in the blue bars to the left.

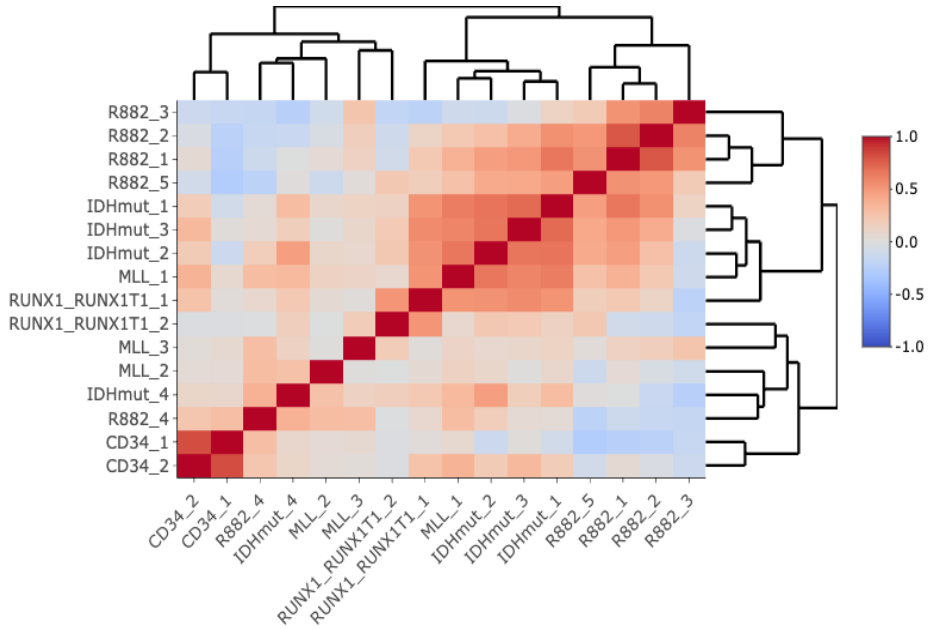


Figure 3.3. Hierarchical clustering based on between-sample correlation at the top 10% of most variable compartment regions in AML samples.

Pearson correlation values for each pairwise comparison between samples (n=4 *IDH* mutations; n=5 *DNMT3A-R882* mutations; n=3 *KMT2A* fusions; n=2 *RUNX1-RUNX1T1* fusions, n= 2 CD34+ normal cells) was calculated for a set of ~6,000 compartment bins representing the top 10% of variance. A higher correlation between samples is represented by darker red bins in the heatmap while darker blue bins represent less/anti-correlated samples.

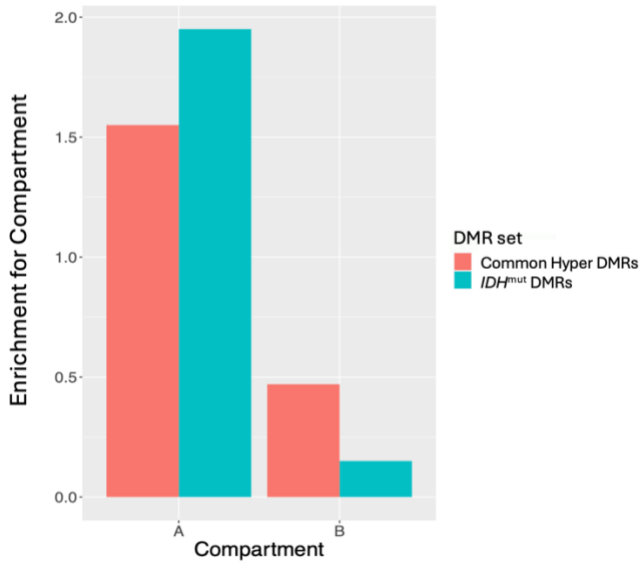


Figure 3.4. Enrichment for DMRs in A vs. B nuclear compartments.

Commonly hypermethylated DMRs (pink bars) and *IDH*^{mut}-specific DMRs (blue bars) were intersected with 50 kb compartment intervals and annotated as active (falling within an A compartment) or suppressed (falling within a B compartment) based on the activity defined in CD34+ normal cells.

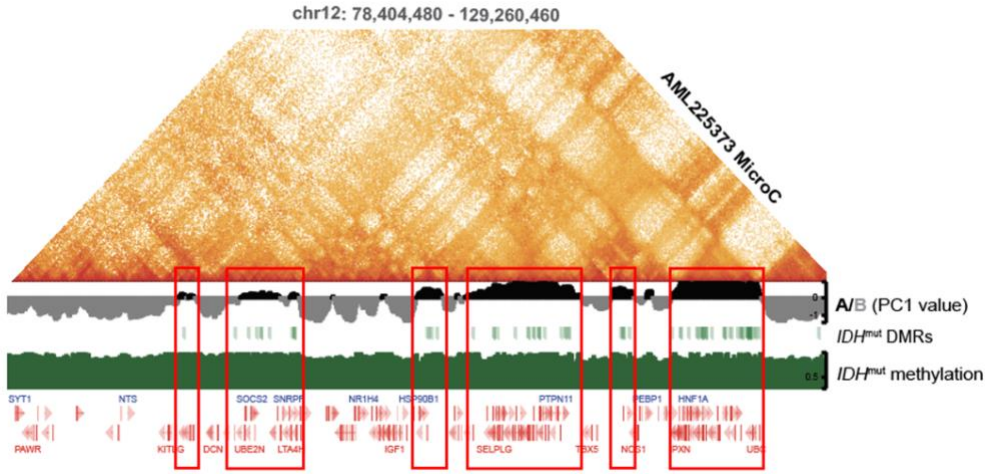


Figure 3.5. Association of *IDH*^{mut}-specific hypermethylation with the active genome compartment.

MicroC contact matrix for an *IDH* mutant AML sample at a representative region on chromosome 12 (top). PC1 values calculated for the same region of chromosome 12 where positive values (black) represent A compartments and grey (negative) values represent the B compartment (1st track of bottom panel). Location of commonly hypermethylated regions in AML and *IDH*^{mut}-specific DMRs (2nd and 4th tracks of the bottom panel, respectively). DNA methylation values for CD34+ normal cells and *IDH*^{mut} AML across the same representative region (3rd and 5th tracks of the bottom panel, respectively). Hg38 gene annotation track (bottom). Red boxes highlight the overlaps of *IDH*^{mut}-specific DMRs with the active compartment.

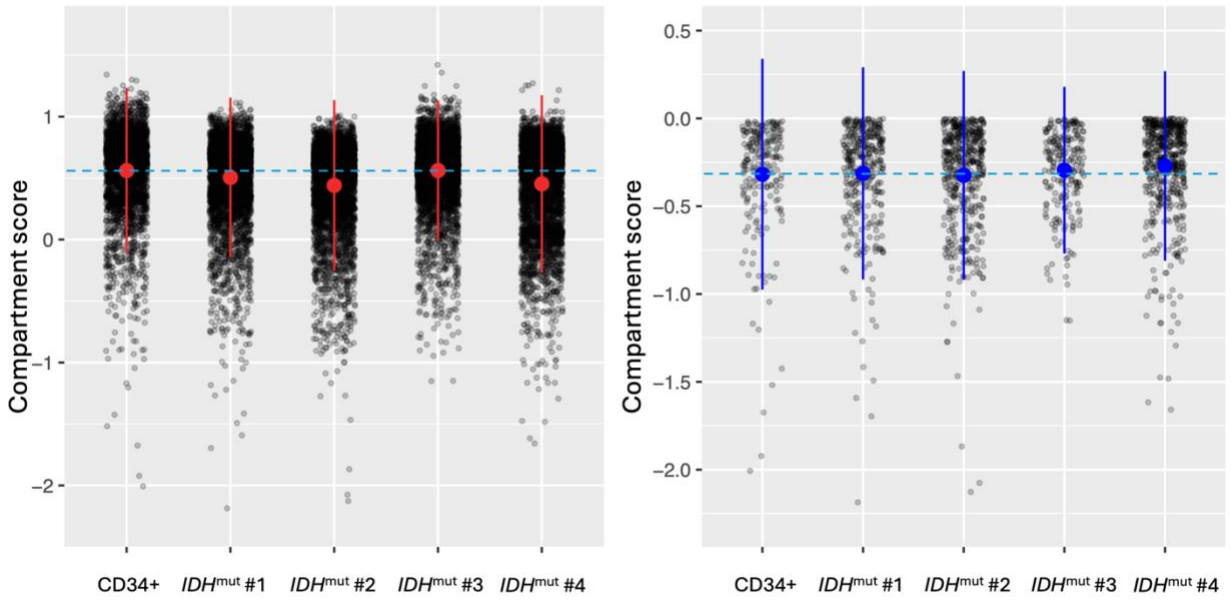


Figure 3.6. Distribution of compartment scores for all *IDH*^{mut}-specific DMR resident compartment bins in normal cells and *IDH*^{mut} samples.

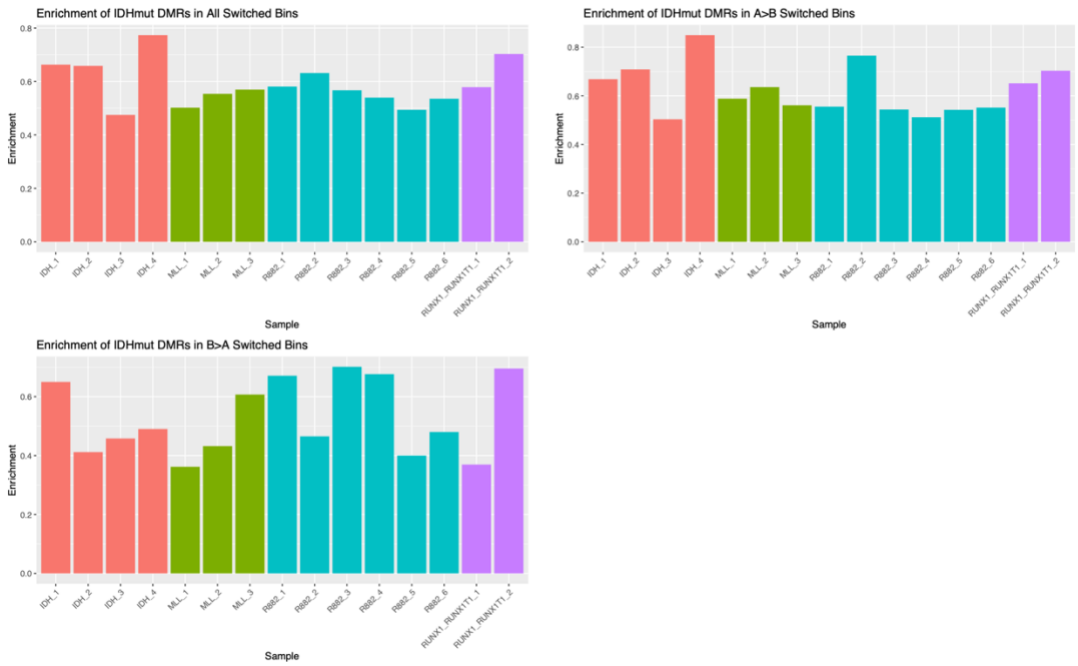


Figure 3.7. Enrichment of *IDH*^{mut}-specific DMRs for switched bins in individual samples.

Top Left: Enrichment calculated over the full set of switched bins irrespective of direction. Top Right: Enrichment calculated over the A > B switch regions only. Bottom Left: Enrichment

calculated over the B > A switch regions only. Enrichment was calculated as the quotient of the frequency of *IDH*^{mut} -DMRs intersecting with a switched bin divided by the frequency of common hypermethylation intersecting a switched bin.

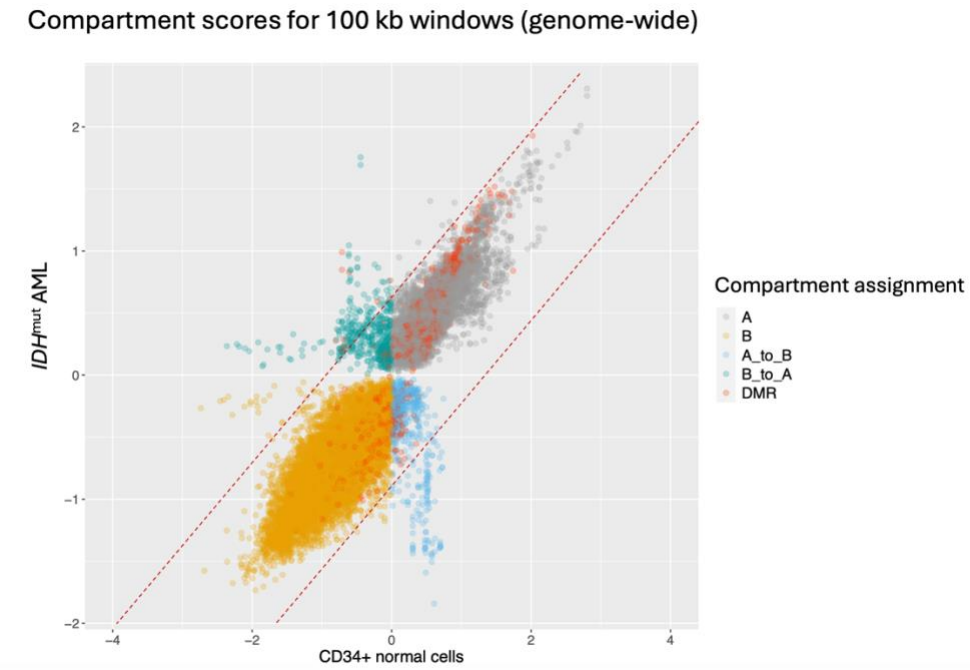


Figure 3.8. Comparison of PC1 scores for 50kb genomic windows between CD34+ normal cells and *IDH*^{mut} AML.

Each point shows the mean PC1 value for a 50kb genomic region from CD34 normal cells (n=2) and IDmut AML (n=4); the color of the point denotes compartment status. Points on the line y=x have similar compartment scores in CD34+ normal cells and *IDH*^{mut} AML (gray dots are stable A, yellow are stable B) while points falling off the y=x diagonal represent compartment switches (teal are A to B switches, blue are B to A switches). Points outside of the dotted red lines falling off the x=y axis represent bins with the strongest shifts in compartmentalization (weak A to strong B or weak B to strong A along the y-axis; strong A to weak A or B along the x-axis).

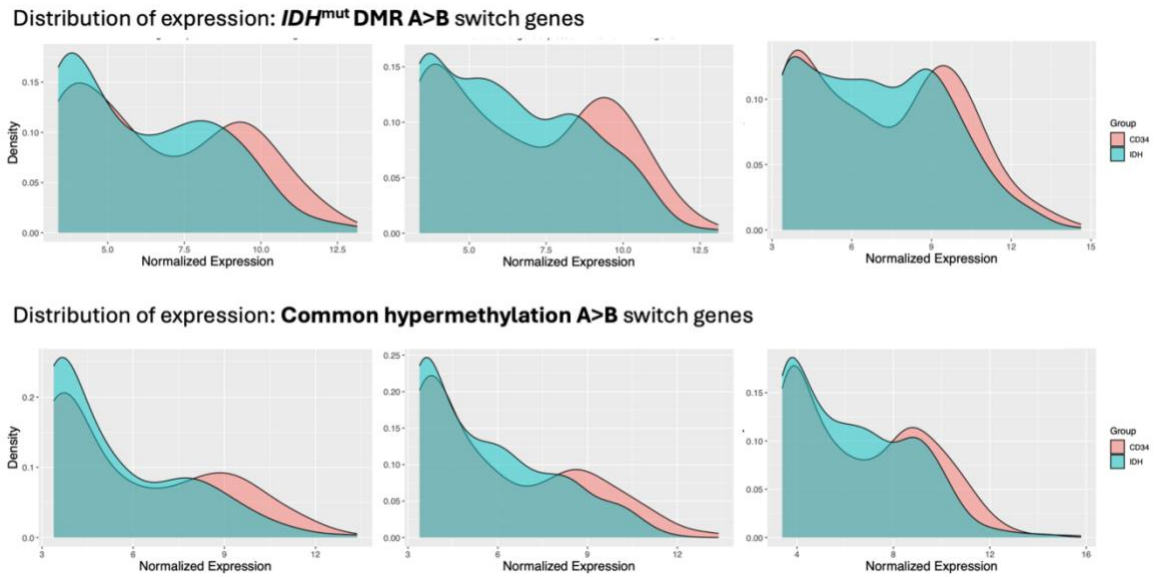


Figure 3.10. Distribution of expression values for genes within DMR-associated A to B compartment switched regions.

The distribution of normalized expression levels for switch compartment genes is plotted for three individual *IDH*^{mut} samples (left to right), with *IDH*^{mut} -specific DMR switch genes displayed in the top row and commonly hypermethylated switch genes displayed in the bottom row. Blue curves represent the distribution of normalized expression in the *IDH*^{mut} case while pink curves represent the expression levels of the same set of genes in normal donor CD34+ cells.

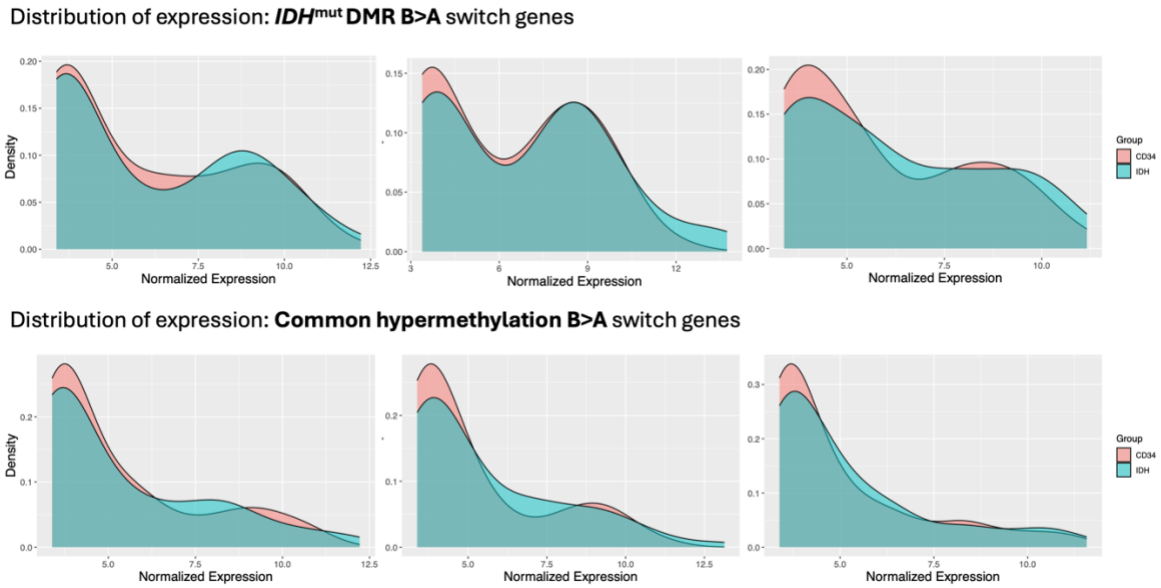


Figure 3.11. Distribution of gene expression within B to A switched DMR compartments.

The distribution of normalized expression levels for switch compartment genes is plotted for three individual *IDH*^{mut} samples (left to right), with *IDH*^{mut} -specific DMR switch genes displayed in the top row and commonly hypermethylated switch genes displayed in the bottom row. Blue curves represent the distribution of normalized expression in the *IDH*^{mut} case while pink curves represent the expression levels of the same set of genes in normal donor CD34+ cells.

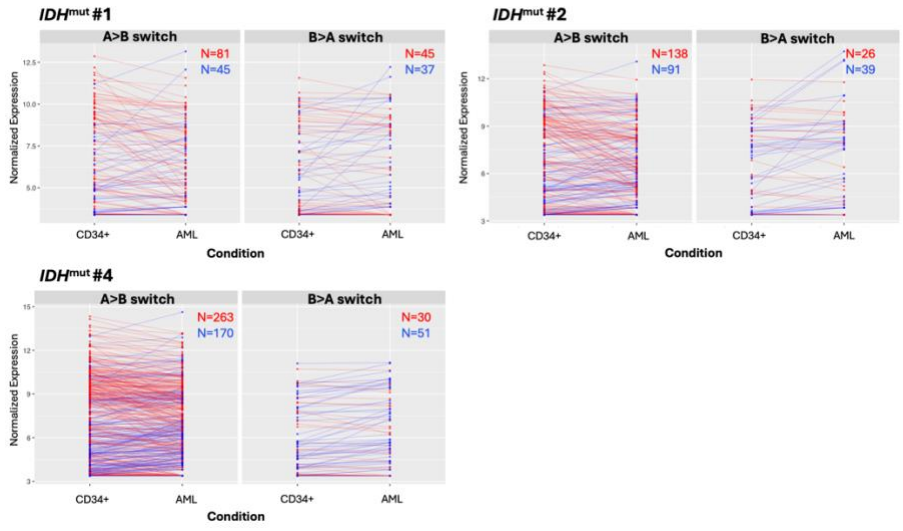


Figure 3.12. Gene expression changes in DMR-associated compartment switches.

Each panel shows the pairwise normalized expression values for genes in DMR resident bin switches for a given patient compared to normal (A to B switches on the right, B to A on the left). Blue lines denote an increase in expression compared to normal, while red lines denote a decrease in expression compared to normal cells.

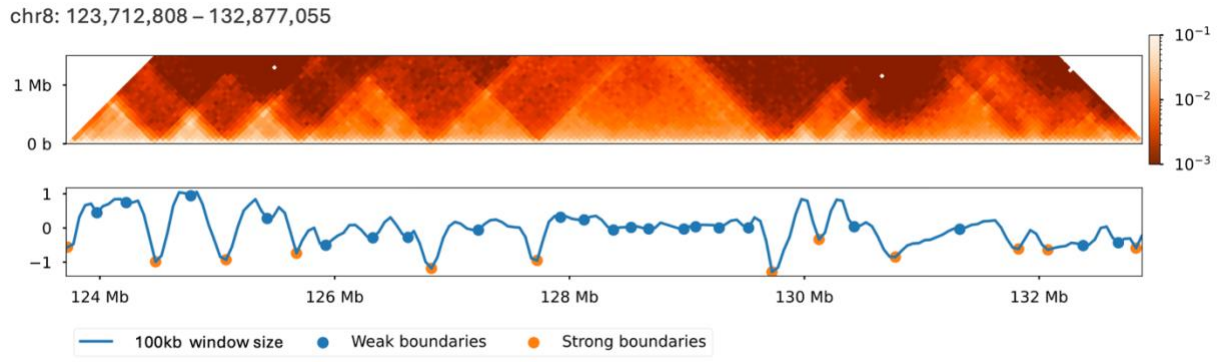


Figure 13. Example contact matrix with insulation profile for a region on chromosome 8.

Cooltools uses a sliding window to summarize interaction frequencies and inspect for local minima consistent with decreased contact between upstream and downstream loci. Insulation minima are annotated with boundary strength, shown here as weak boundaries in blue and strong boundaries in orange.

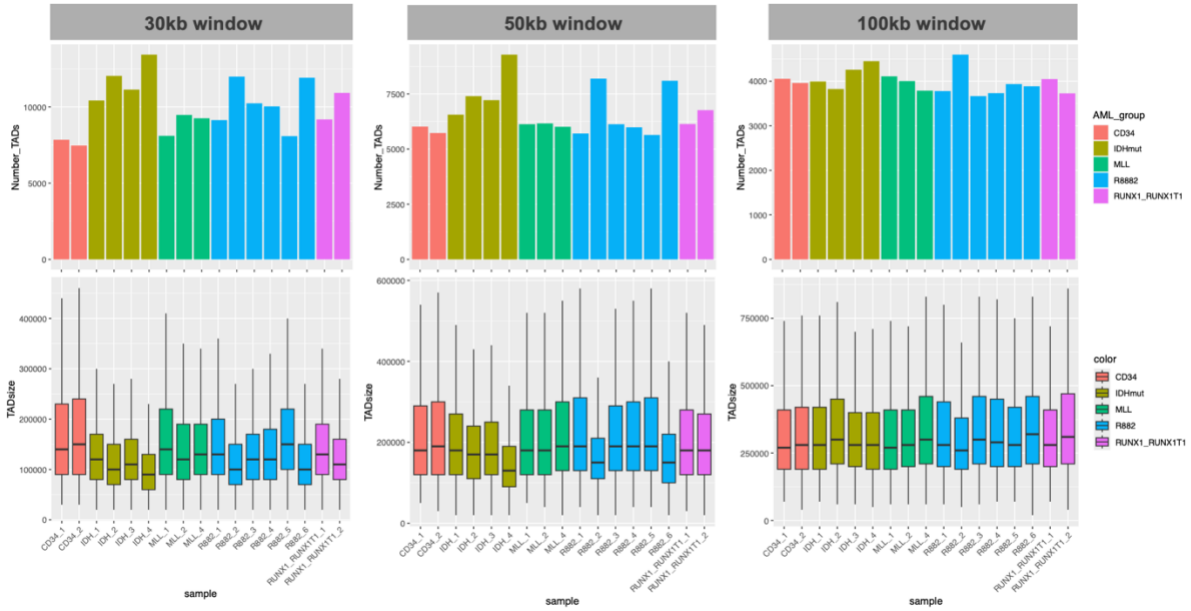


Figure 3.14. Sample level summary of genome-wide cooltools TAD boundaries assigned using different sliding window sizes (30kb, 50kb, 100kb) for 10kb resolution data.

Top: Total number of boundaries meeting the threshold cutoff for strong TAD classification. Bottom: Distribution of TAD sizes. Normal donor and AML samples are color-coded according to driver mutation subtypes.

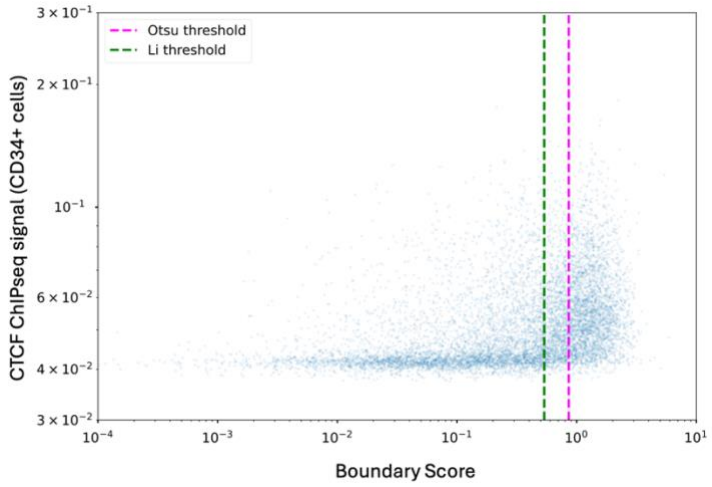


Figure 3.15. CTCF binding enrichment in strong TAD boundaries.

Boundary strengths assigned to all insulation minima called in CD34+ contact matrix are plotted against the CTCF ChIPseq signal calculated over the same region. Threshold parameters used to define the set of strong peaks in cooltools are plotted in pink (Otsu) and green (Li) dashed lines to show the level of enrichment of CTCF binding for strong peaks.

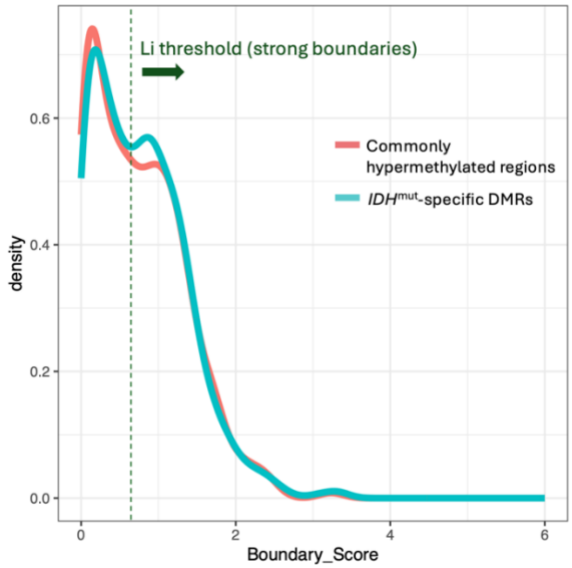


Figure 3.16. Distribution of scores for boundaries intersecting an *IDH*^{mut}-specific DMR (blue) or a commonly hypermethylated region (salmon) in CD34+ normal cells.

The Li threshold line demarcates the cutoff for strong boundaries.

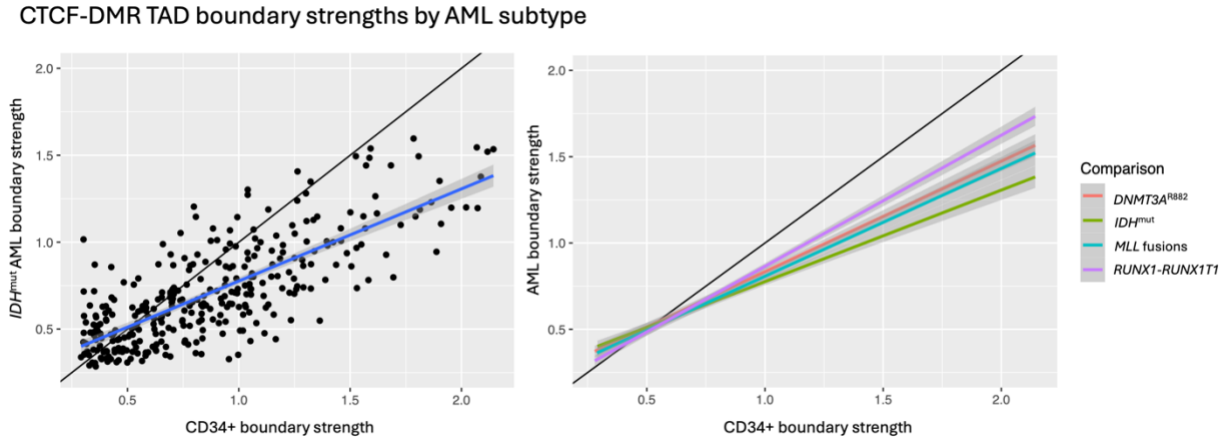


Figure 3.17. TAD boundary scores for CTCF-containing boundaries in *IDH*^{mut} (left) and *IDH*^{wt} AML samples at boundaries defined in CD34+ normal cells.

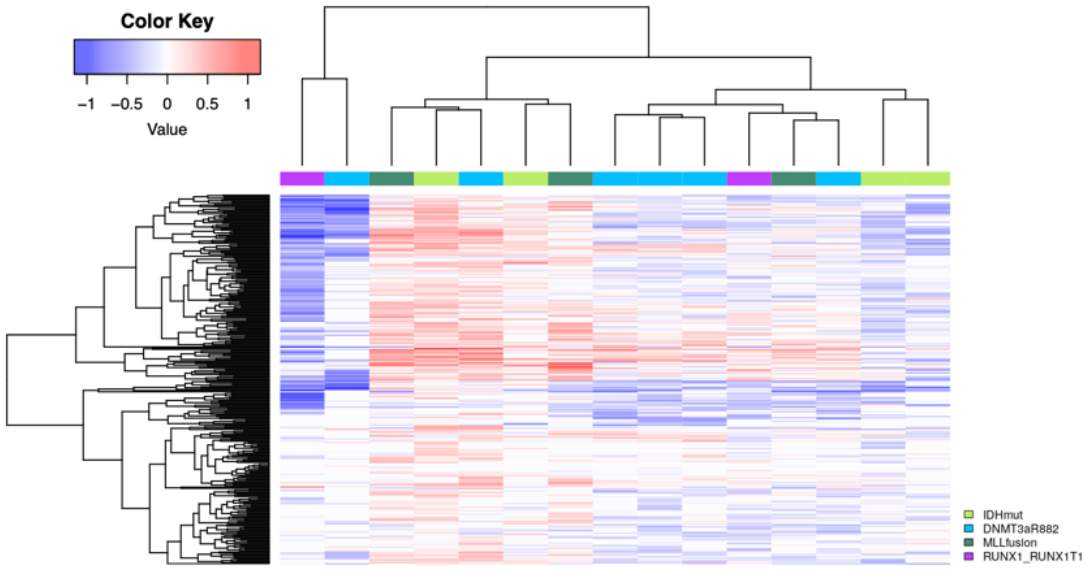


Figure 3.18. Hierarchical clustering of AML samples based on the difference from normal boundary scores at CTCF-DMR TAD boundaries as defined in CD34+ normal donor cells.

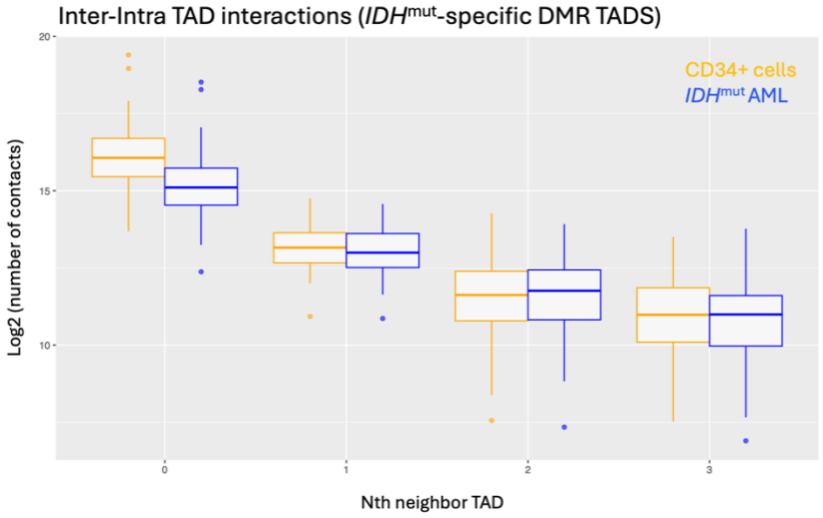


Figure 3.19. Intra- and inter-TAD interaction frequencies for the set of CTCF-DMR associated TADs in CD34+ normal cells (yellow) and *IDH* mutant AML (blue).

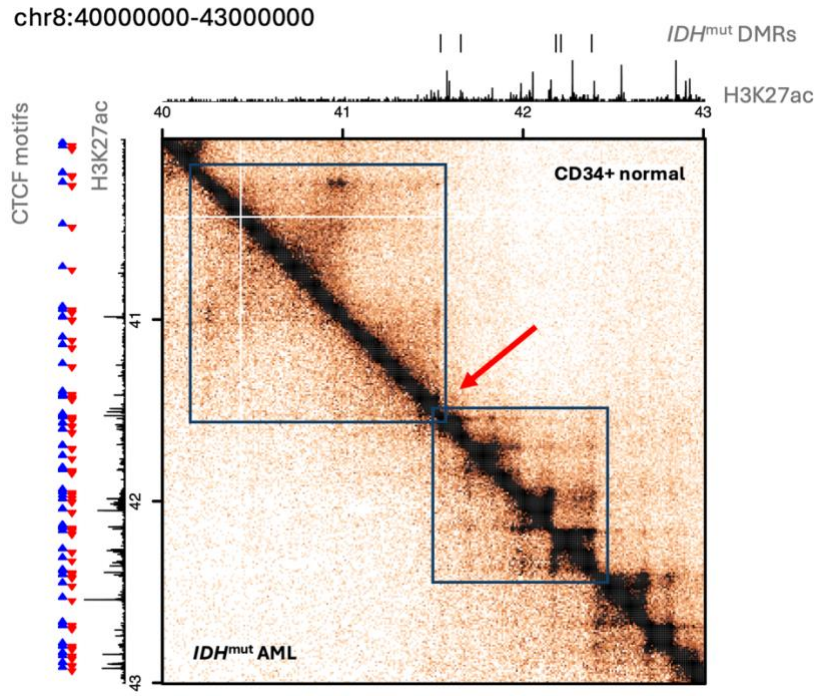


Figure 3.20. Example region on chromosome 8 of contact matrix demonstrating loss of TAD insulation around a CTCFdmr TAD boundary in *IDH*^{mut} AML (bottom corner matrix) compared to normal donor cells (top corner matrix).

The arrow indicates a boundary with reduced insulation in *IDH*^{mut} AML. Black boxes annotate regions with notable disruption of local topology. Top: Tracks annotating *IDH*^{mut} -specific DMR location and H3K27ac Cut&Tag signal from normal donor apheresis sample. Left: CTCF motifs called within CTCF ChIPseq peaks in primary AML samples.

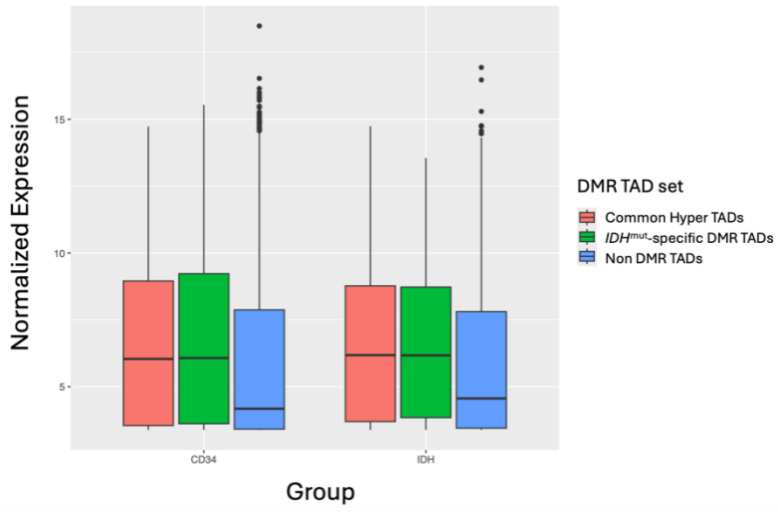


Figure 3.21. Normalized expression for all genes residing in a TAD with a CTCFdmr or CTCFcommon-hyper boundary compared with the set of all expressed genes.

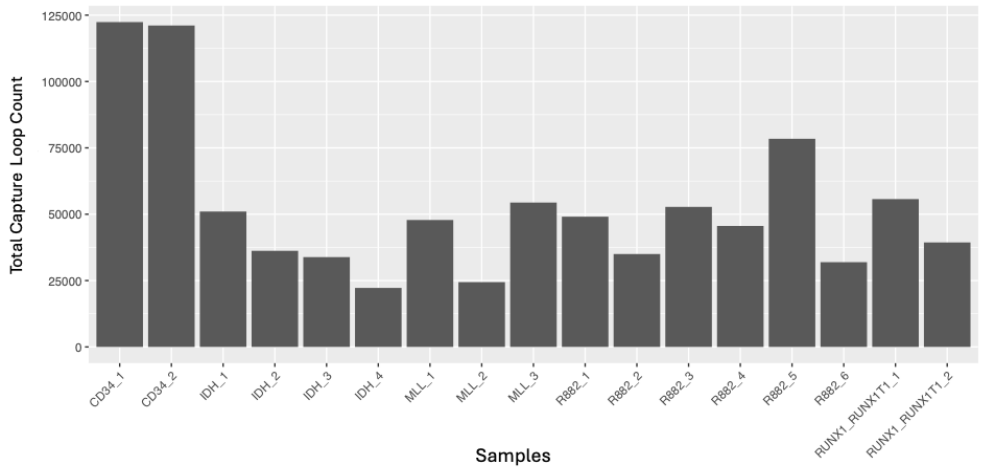


Figure 3.22. The total number of CHiCAGO loops clearing the statistical threshold (interaction score > 5) per sample.

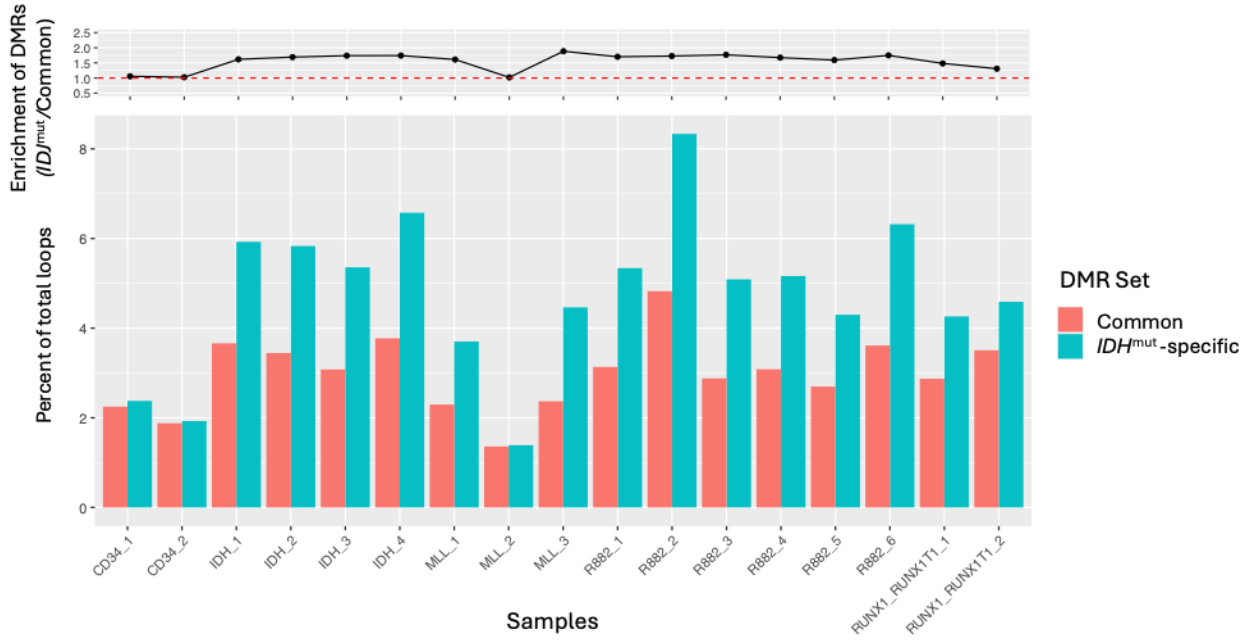


Figure 3.23. Intersection of loop interactions with DMRs. Bottom: Fraction of total loops intersecting the set of commonly hypermethylated regions (salmon bars) and *IDH*^{mut}-specific DMRs (blue bars) per sample.

Top: Enrichment for *IDH*^{mut}-specific DMRs in loop anchors compared to commonly hypermethylated regions.

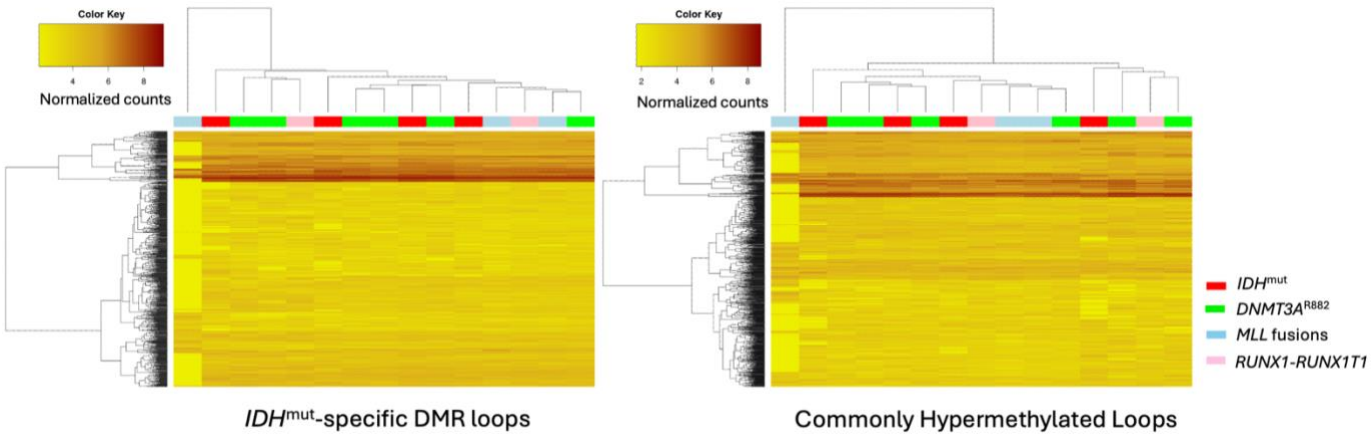


Figure 3.24. Hierarchical clustering of chromatin loop interaction frequencies for DMR loops (left) and commonly hypermethylated loops (right). Normalized counts are plotted in the

heatmap with light yellow bins representing lower interaction frequencies and darker red bins representing higher interaction frequencies.

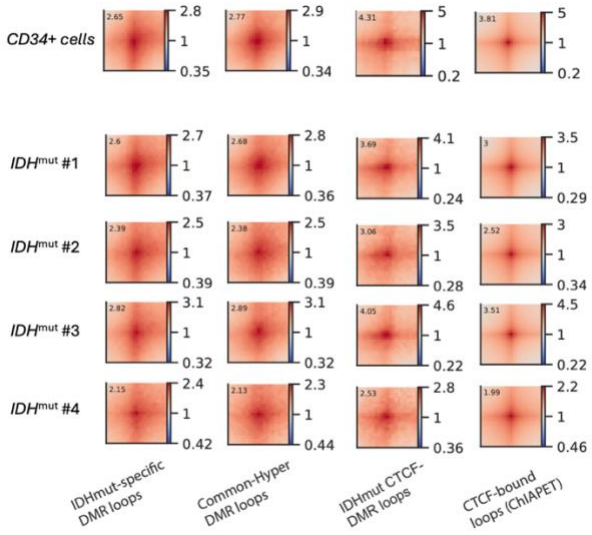


Figure 3.25. Aggregate peak analysis (APA) of MicroC data using different subsets of DMR capture loops in *IDH*^{mut} samples.

Columns left to right: Loops intersecting an *IDH*^{mut}-specific DMR in the anchor opposite the bait, loops intersecting a commonly hypermethylated region in the anchor opposite the bait, CD34+ normal cell CTCF-bound loops defined by ChIAPET assays, CTCF-bound *IDH*^{mut}-specific DMR loops. Rows: CD34+ normal donor and individual *IDH*^{mut} AML samples.

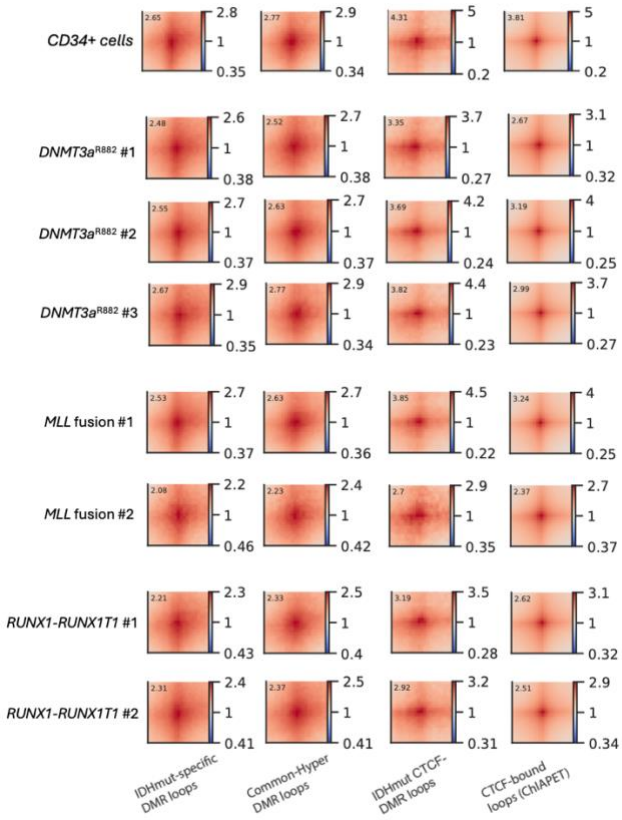


Figure 3.26. Aggregate peak analysis (APA) of MicroC data using different subsets of DMR capture loops in *IDH*^{wt} AML.

Columns left to right: Loops intersecting an *IDH*^{mut}-specific DMR in the anchor opposite the bait, loops intersecting a commonly hypermethylated region in the anchor opposite the bait, CD34+ normal cell CTCF-bound loops defined by ChIAPET assays, CTCF-bound *IDH*^{mut}-specific DMR loops. Rows: CD34+ normal donor and individual *IDH*^{wt} AML samples.

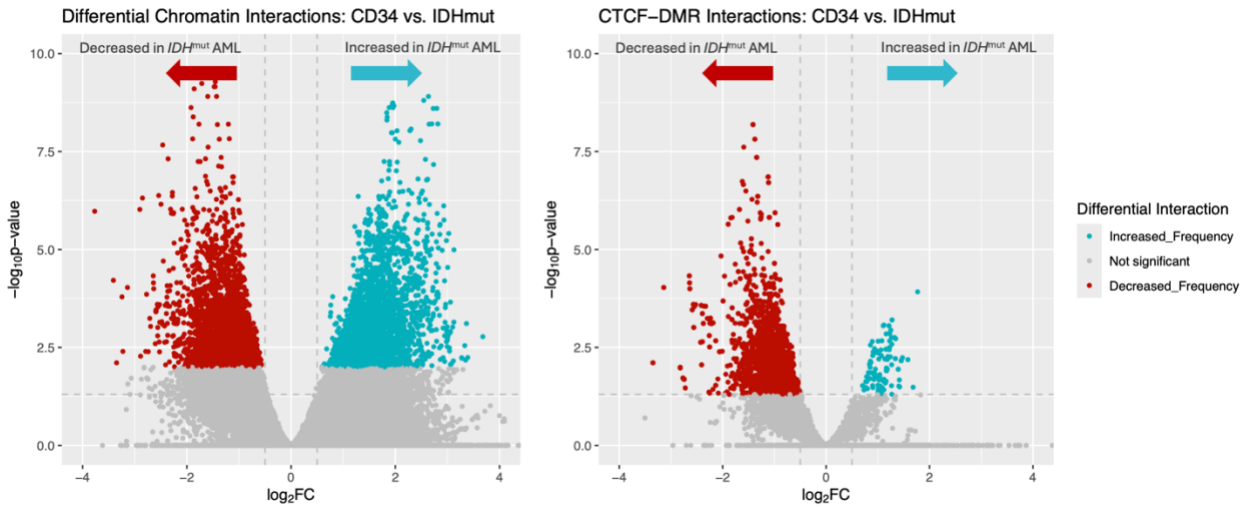


Figure 3.27. Differential chromatin interactions detected in a comparison between CD34+ normal cells and *IDH*-mutant AML.

Left: Volcano plot of the union set of interactions. Right: Volcano plot of the subset of interactions intersecting a CTCF-DMR. Interactions with statistically increased frequency in *IDH*-mutant AML are annotated in blue while those statistically decreased in frequency are annotated in red. Gray dots are not significantly different between groups.

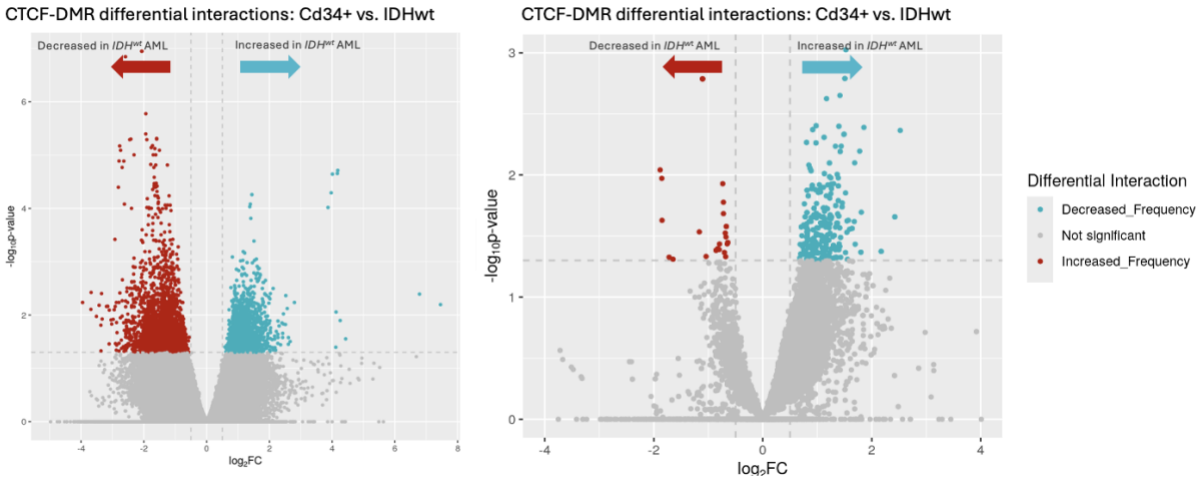


Figure 3.28. Differential chromatin interactions detected in a comparison between CD34+ normal cells and *IDH* wild-type AML. Left: Volcano plot of the union set of interactions.

Right: Volcano plot of the subset of interactions intersecting a CTCF-DMR. Interactions with statistically increased frequency in *IDH*-mutant AML are annotated in blue while those statistically decreased in frequency are annotated in red. Gray dots are not significantly different between groups.

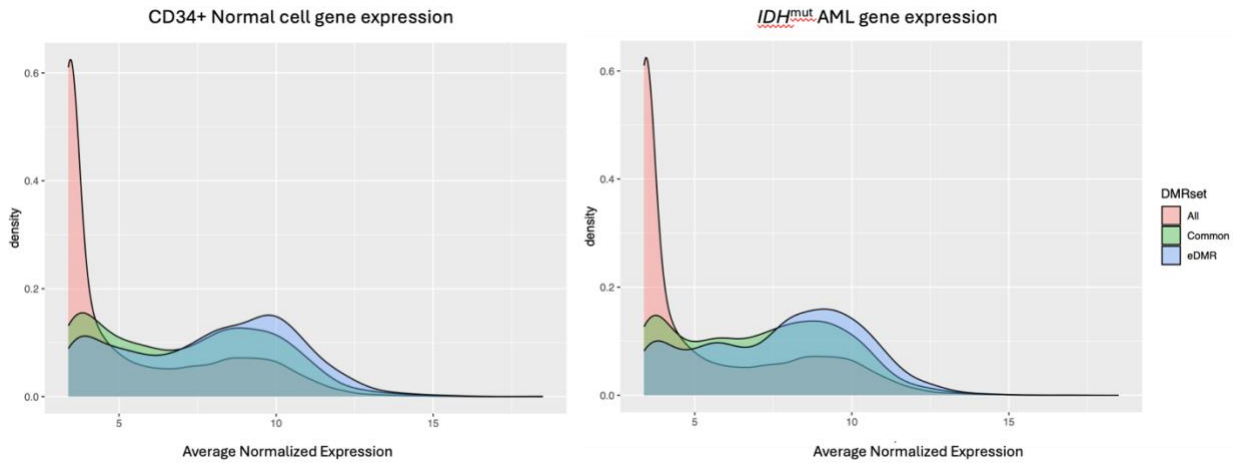


Figure 3.29. Distribution of gene expression in CD34+ normal samples (left) and *IDH*^{mut} AML samples (right).

Density curves represent different subsets of genes including those opposite commonly hypermethylated loop anchors (green), those opposite *IDH*^{mut}-specific DMR loop anchors, and the full set of expressed genes.

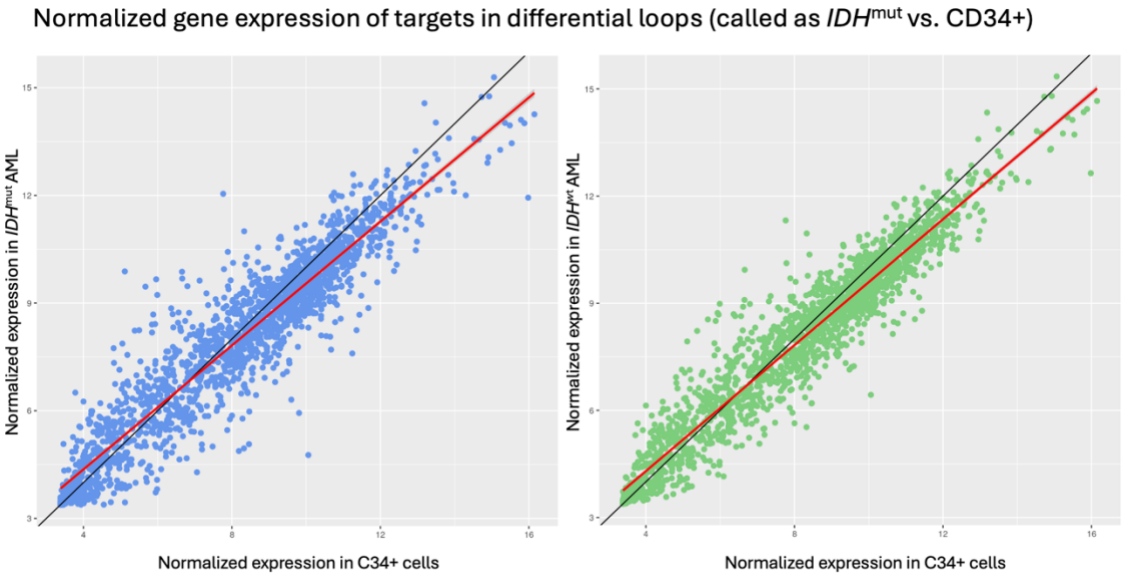


Figure 3.30. Expression of DMR target genes in differential loops.

Normalized expression of *IDH*^{mut}-specific DMR target genes in *IDH*^{mut} AML and (right) and *IDH*^{wt} AMLs (left) involved in the set of CD34+ vs. *IDH* mutant differential interactions.

3.5 References

1. Wilson, E. R. *et al.* Focal disruption of DNA methylation dynamics at enhancers in IDH-mutant AML cells. *Leukemia* **36**, 935–945 (2022).
2. Rahman, H. Dovetail® Micro-C Kit User Guide - Version 1.2.
3. Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods* **12**, 963–965 (2015).
4. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
5. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
6. Gaspar, J. M. Improved peak-calling with MACS2. Preprint at https://doi.org/10.1101/496521 (2018).
7. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
8. null null. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
9. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).

10. Cresswell, K. G., Stansfield, J. C. & Dozmorov, M. G. SpectralTAD: an R package for defining a hierarchy of topologically associated domains using spectral clustering. *BMC Bioinformatics* **21**, 319 (2020).
11. Open2C *et al.* Cooltools: enabling high-resolution Hi-C analysis in Python. 2022.10.31.514564 Preprint at https://doi.org/10.1101/2022.10.31.514564 (2022).
12. Flyamer, I. M., Illingworth, R. S. & Bickmore, W. A. Coolpup.py: versatile pile-up analysis of Hi-C data. *Bioinformatics* **36**, 2980–2985 (2020).
13. van der Weide, R. H. *et al.* Hi-C analyses with GENOVA: a case study with cohesin variants. *NAR Genomics Bioinforma.* **3**, lqab040 (2021).
14. Kerpedjiev, P. *et al.* HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 125 (2018).
15. Kruse, K., Hug, C. B. & Vaquerizas, J. M. FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. *Genome Biol.* **21**, 303 (2020).
16. Leidescher, S. *et al.* Spatial organization of transcribed eukaryotic genes. *Nat. Cell Biol.* **24**, 327–339 (2022).
17. Yoshida, A., Kato, J.-Y., Nakamae, I. & Yoneda-Kato, N. COP1 targets C/EBPα for degradation and induces acute myeloid leukemia via Trib1. *Blood* **122**, 1750–1760 (2013).
18. Li, Z. *et al.* FTO Plays an Oncogenic Role in Acute Myeloid Leukemia as a *N*6-Methyladenosine RNA Demethylase. *Cancer Cell* **31**, 127–141 (2017).
19. Guo, J.-R. *et al.* Hepatocyte growth factor promotes proliferation, invasion, and metastasis of myeloid leukemia cells through PI3K-AKT and MAPK/ERK signaling pathway. *Am. J. Transl. Res.* **8**, 3630–3644 (2016).

20. Nepstad, I., Hatfield, K. J., Grønningsæter, I. S. & Reikvam, H. The PI3K-Akt-mTOR Signaling Pathway in Human Acute Myeloid Leukemia (AML) Cells. *Int. J. Mol. Sci.* **21**, 2907 (2020).

21. Heuser, M. *et al.* MN1 overexpression induces acute myeloid leukemia in mice and predicts ATRA resistance in patients with AML. *Blood* **110**, 1639–1647 (2007).

22. Bradbury, C. A. *et al.* Histone deacetylases in acute myeloid leukaemia show a distinctive pattern of expression that changes selectively in response to deacetylase inhibitors. *Leukemia* **19**, 1751–1759 (2005).

23. Liu, Y. & Dekker, J. CTCF–CTCF loops and intra-TAD interactions show differential dependence on cohesin ring integrity. *Nat. Cell Biol.* **24**, 1516–1527 (2022).

24. Dekker, J. *et al.* Spatial and temporal organization of the genome: Current state and future aims of the 4D nucleome project. *Mol. Cell* **83**, 2624–2640 (2023).

25. Narendra, V., Bulajié, M., Dekker, J., Mazzoni, E. O. & Reinberg, D. CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev.* **30**, 2657–2662 (2016).

26. Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).

27. Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).

28. Welcome to Micro-C documentation — Micro-C 0.1 documentation. https://micro-c.readthedocs.io/en/latest/.

29. Holwerda, S. J. B. & de Laat, W. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20120369 (2013).

30. Haarhuis, J. H. I. *et al.* A Mediator-cohesin axis controls heterochromatin domain formation. *Nat. Commun.* **13**, 754 (2022).

31. Richter, W. F., Nayak, S., Iwasa, J. & Taatjes, D. J. The Mediator complex as a master regulator of transcription by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **23**, 732–749 (2022).

32. Marsh, J. C. W. *et al.* Heterozygous RTEL1 variants in bone marrow failure and myeloid neoplasms. *Blood Adv.* **2**, 36–48 (2018).

33. Artigas-Baleri, A. *et al.* RAS Mutations in Adult Acute Myeloid Leukemia (AML). Frequency, Mutational Spectrum, and Identification of a Comutation Bias for KRASK117 (TET2/ASXL1). *Blood* **140**, 11821–11823 (2022).

Chapter 4: Characterizing the activity of candidate enhancer DMRs to better understand consequences of hypermethylation: *MYC* BENC and *ARID1a* enhancer.

Elisabeth R. Wilson, Mohamed Mahgoub, Heidi Struthers, and David H. Spencer

Declarations

Conception and design: ERW, DHS. Acquisition of data: ERW, HS. Data processing, statistical and computational analysis: ERW. Administrative, technical, or material support: HS, DHS. Study supervision: DHS.

4.1 Introduction

Our observation that focal hypermethylation in *IDH* mutant primary acute myeloid leukemia samples affects enhancers suggests these regions could be manipulated to affect gene expression. While the optimal experiment would be to induce a hypermethylation phenotype in biologically relevant models, the available tools for remodeling epigenetic patterns in a targeted, site-specific manner remain challenging to employ. Therefore, we first wanted to characterize the baseline activity of candidate eDMR regions to understand how they influence gene expression in their normal state. Since much of the work described in this chapter was conducted in parallel with our analysis of Micro-C data from AML patients, our choice of candidate loci was informed by the lower-resolution HiC data available at the time for a limited number of AML samples and cell lines. Future studies will benefit from the higher-resolution MicroC data recently generated from a large set of AML cases when comprehensively investigating the nature of eDMR-target gene interactions.

Several hematopoietic gene candidates (as well as many less well-studied genes) were identified in our preliminary analysis as targets of enhancers uniquely hypermethylated in *IDH*^{mut} samples including *MYC, ARID1A, RUNX3, ASXL1, DOT1L, EGFRA, KDM2A, NOTCH,* and *SRSF3*. As previously discussed, the genes forming interactions with *IDH*^{mut}-specific eDMRs tend to be highly expressed compared to the distribution of all expressed genes in AML cells, suggesting that hypermethylation may occur in regions where regulatory activity is increased to maintain elevated target gene expression. Many of the genes in our selection are known to be active during development, including *MYC*. HiC data from primary AML samples and AML cell lines show robust interactions between the *MYC* promoter and a region that is approximately 1.7 mb downstream of the transcription start site containing a discrete locus that is hypermethylated in *IDH*^{mut} AML samples and unmethylated in *IDH*^{wt} samples and normal hematopoietic progenitor cells (Figure 1).

Previous studies of the *MYC* locus have characterized an extended enhancer region containing our DMR of interest. Termed the blood enhancer cluster (BENC), this region contains evolutionarily conserved regulatory elements that are critical for *MYC* expression in blood cells[1]. The BENC consists of multiple modules, each of which is thought to have varying importance for *MYC* expression depending on the stage of hematopoietic development. To date, the most comprehensive studies of the BENC have been performed in mouse bone marrow, where the authors systematically knocked out individual BENC modules or the whole cluster and measured the consequences on *MYC* expression and cell differentiation. The authors reported that knocking out the entire cluster led to complete loss of *MYC* expression in hematopoietic stem and progenitor cells and accumulation of developmentally arrested myeloid progenitors, while changes resulting from individual module deletion were restricted to different developmental

states. Interestingly, deletion of the module bearing homology to our eDMR region of interest was the most deleterious for *MYC* expression in mouse hematopoietic stem cells, suggesting a potential role for the eDMR region in regulating *MYC* expression in normal human hematopoiesis and AML.

Mechanistic studies of the *MYC* BENC in human cells have been limited, however structural variations including tandem duplications within the BENC region have been observed in ~3% of primary AML samples and are thought to introduce additional binding sites for BRG1 and other TFs that may alter or amplify the regulatory capacity of the region[2]. Studies in K562 cell lines have interrogated the BENC for units of active regulation by tiling dCas9-KRAB binding across the locus in a silencer screen. Though their gRNAs span the full length of the BENC and identified critical regions for *MYC* expression, their screen did not specifically test the regulatory capacity of our eDMR region. Other groups have indirectly investigated independent modules of the BENC during their studies of BET-inhibitor therapy and resistance, reporting altered binding of BRG1 to the *MYC* super-enhancer and consequent decreases in *MYC* expression and cell viability, however, definitive examples of activity for the eDMR region are still outstanding. Given the well-established role of *MYC* expression in hematopoiesis and the evidence for robust 3D chromatin architecture encapsulating the region, we decided to prioritize the *MYC* locus as a model for investigating the relationship between DNA hypermethylation and enhancer activity.

In parallel, we also wanted to develop models to investigate the relationship of *ARID1a* with its cognate eDMR. *ARID1a* participates in epigenetic regulation as a component of SWI/SNF chromatin remodeling complexes and is documented to be the most frequently mutated subunit across human cancers. *ARID1a* is thought to play an important role in both tumor initiation and tumor suppression depending on the biological context. For example, primary endometrial

cancers carry WT *ARID1a* prior to developing deleterious mutations upon metastasis, suggesting a role for *ARID1a* in a narrow window of early disease transformation[3], however, the majority of data supports a tumor-suppressive function for the protein. Recurrent loss of function mutations in ovarian carcinoma, endometrial adenocarcinomas, colon cancer, and gastric cancer have been linked to altered processes including tumor formation, cell proliferation, cell cycle, and DNA damage repair.

Although mutations in *ARID1a* are less common in hematologic malignancies, altered regulation and/or expression of *ARID1a* may still be relevant for leukemogenesis or processes related to disease maintenance and progression. For example, *ARID1a* has been shown to target SWI/SNF complexes to enhancers thereby cooperating with transcription factors to promote chromatin accessibility and gene expression[4]. *In vitro* models of *ARID1a* deficiency in hepatocellular carcinoma and colon cancer have shown correlations between augmented histone H3K27ac at known enhancers and changes in transcription of the nearby genes[4]. In colorectal cancers, loss of *ARID1a* showed the inactivation of *ARID1a* -bound enhancers and a concomitant decrease in target gene expression[5]. Therefore, it is possible that changes in *ARID1a* expression resulting from methylation-related changes at its enhancers could affect transcriptional and/or signal pathways in favor of a leukemia phenotype. Our goal for this set of experiments was simply to determine if interactions between the *ARID1a* promoter and its cognate eDMR are necessary for normal expression in a hematopoietic model.

As mentioned above, the goal of future experiments will be to perform enhancer modification studies where methylation is targeted to individual enhancement units, thereby recapitulating the *IDH*^{mut}-specific enhancer hypermethylation phenotype in a forward approach to assess the consequences on gene expression and 3D architecture. To inform future hypotheses, we started

by testing the dependency of candidate gene expression on the native enhancer context by knocking out the eDMR enhancer regions and measuring the effect on target gene regulation. CRISPR-Cas9 targeted editing of AML cell lines was used to delete specific regions identified as being hypermethylated and *IDH*^{mut} AML and RNAseq was performed to measure resultant gene expression.

4.2 Methods

CRISPR-Cas9 cell editing

Ribonuclear complexes of Cas9 protein coupled with guide RNAs were electroporated into AML cell lines using the NEON electroporator. RNAs were designed to flank the eDMR regions of interest. Single-cell sorting of treated cells followed by expansion in culture, generated clonal knockout lines that were validated by PCR sequencing.

RNA-seq analysis

RNA-seq libraries were generated on whole-cell RNA. Sequenced data were processed using standard read alignment and feature counting methods. The DESeq2 Bioconductor package was used to normalize gene counts and perform differential analysis.

4.3 Results

4.3.1 Generation of clonal eDMR knockout cell lines

Our previous work profiling methylation phenotypes in *IDH*^{mut} AML identified several focally hypermethylated regions enriched for active enhancer modifications. Given the observed association of hypomethylation with putative regulatory activity, we wanted to characterize the relationship of the association using a few candidate loci including a distinct region coincident with a portion of the BENC, a highly conserved multi-modular superenhancer of *MYC* located 1.7 mb downstream of the genes' promoter[1]. Additional annotation of these regions identified

transcription factor binding motifs including BRD4, BRG1, and YY1, further supporting their putative activity and suggesting that aberrant methylation may alter the epigenetic capacity of the locus to promote transcription.

To functionally validate the dependence of AML cells on the *MYC* eDMR locus, we generated clonal knockouts of the eDMR region in AML cell lines. Using a CRISPR-Cas9 RNP editing approach, we used flanking guide RNAs to target two individual regions of the *MYC* enhancer for deletion in OCI-AML3 and MOLM13 cells. Neighboring regions were identified as having active enhancer marks; one region spanning ~1.1 kb overlapped a known binding site of BRD4 in other AML cell lines (enhancer A or "EnhA"), while the other region spanning ~1.4 kb with *IDH*^{mut}-specific focal hypermethylation harbored a YY1 binding site specifically detected in the OCI-AML3 cell line (enhancer B or "EnhB") (Figure 1). Interactions between the *MYC* promoter and these distal enhancer regions are enriched as shown in the MicroC contact matrices from AML samples and normal cells, forming distinct loops consistent with corner TAD insulation (Figure 2). We also targeted the adjacent regions for deletion in the same cell to measure the impact of combinatorial enhancer loss. We elected to use OCI-AML3 cells as an initial model because they have low methylation in these enhancer regions similar to normal hematopoietic stem/progenitor cells and AML samples without *IDH* mutations. Of note, OCI-AML3 cells are triploid for the *MYC* gene and enhancer (+chr8), which requires careful interpretation of the deletion clone data. At the onset of our experiments, we did not have genome-wide methylation data from MOLM13 cells to adequately assess the methylation status of the eDMR locus, but we wanted an orthogonal AML model that was diploid for the *MYC* locus to help interpret the consequences of *MYC*-eDMR deletion.

After performing CRISPR-Cas9 nucleofection with dual guides, single cells were sorted and cultured for 2-3 weeks to allow for clonal expansion. Expanded clones were then passaged 3-4 times before screening for deletions—we noticed a delayed doubling time for certain clones during the first and second passages, however, after an additional two passages the cells all appeared to double at a normal rate. eDMR deletions were screened using PCR primers that spanned the deletion site and promising clones were sent for follow-up sequencing to confirm the size of enhancer deletions. We obtained multiple deletions in the OCI-AML3 cell line while MOLM13 cells grew poorly after editing. Single-cell-derived MOLM13 clones were frozen back for future experiments, while OCI-AML3 clones were carried forward for phenotypic analysis. Of note, we were only able to successfully generate EnhB KO clones in OCI-AML3 and MOLM13 clones. While we had multiple clones grow out from our EnhA and double Enh KO experiments, none of them were confirmed to have the appropriate deletions, suggesting that loss of EnhA may be incompatible with the viability or growth of OCI-AML3 and MOLM13 cells.

Twelve OCI-AML3 clones were identified as having PCR bands consistent with the EnhB eDMR deletion. These were predominantly heterozygous at the locus (or contained at least one WT allele; Figure 4). Targeted analysis confirmed deletions of roughly the same size across clones consistent with the targeting sites of our guide RNA (Figure 5). We selected 7 representative EnhB KO clones with clean deletion signals based on sequencing, 6 heterozygous clones, and one probable homozygous clone, and performed RNAseq to measure changes in *MYC* expression. Whole-genome sequencing was not performed on these clones but will be necessary in the future to rule out any off-target editing that may affect the *MYC* locus.

Deletion of the *ARID1a* eDMR was carried out using a similar approach. gRNAs were designed to flank the region of differential hypermethylation. Transfected cells were again single-cell

sorted and given time to grow out before screening. Although *ARID1a* eDMR deletion was less efficient in OCI-AML3 cells compared with the *MYC* locus, we did identify three clones with PCR deletion bands (Figure 6). PCR screening suggested successful deletion in all three clones, one clear heterozygous deletion, and two compound heterozygous clones. Targeted sequencing confidently identified deletions in two of the clones (Figure 7), while the other was inconclusive due to sequencing artifacts, however, we proceeded with further phenotypic characterization for all three deletions with the caveat of interpreting data from the inconclusive clone with caution. Whole-genome sequencing was not performed on these clones but will be necessary in the future to rule out any off-target editing that may affect the *ARID1a* locus.

4.3.2 Deletion of the *MYC* eDMR correlates with a decrease in *MYC* expression.

As a readout of enhancer function, we measured changes in *MYC* expression after the deletion of eDMR regions. RNA was collected from individual clonal deletion lines and RNAseq libraries were generated. Sequencing data was processed and aligned with standard tools and normalized exon counts were used to compare gene expression levels across seven EnhB KO clones and two OCI-AML3 WT cells. *MYC* expression was very similar all samples; however, KO clones were variable and displayed a trend towards decreased expression compared with WT cells (average 12.7 vs. 12.5, respectively; Figure 8). This decrease in expression did not reach a level of statistical significance (p-value = 0.142; Figure 10) and will require additional RNAseq data from WT cells to adequately assess whether the decrease observed in KO clones is a consequence of enhancer deletion, or consistent with intra-group heterogeneity that we aren't seeing in the WT group due to a small sample size. If the correlation between EnhB deletion and decreased expression were to hold up in a larger sample size, our results may suggest a role for the eDMR region in regulating normal *MYC* expression.

One of the anticipated challenges of interpreting expression changes in *MYC* is the presence of a third allele in OCI-AML3 cells that likely contributes to the overall transcriptional output of the locus. Most of our clones were heterozygous for enhancer deletion (i.e. had at least one intact WT allele based on PCR results), and it's unclear whether loss of enhancer activity from one or even two alleles would have a measurable impact on overall expression levels of *MYC*. To address this challenge, we tried to determine if altered *MYC* expression could be observed at the level of allele-specific expression. Leveraging common variation in the genome, we wanted to assign RNAseq reads to individual alleles and look for any bias or skewing consistent with the deletion of a *cis*-acting enhancer. We were able to locate two single nucleotide polymorphisms (SNPs) in *MYC*, one at the 3' end of the first exon and one in the middle of the second exon. While sequencing coverage in a small fraction of clones was optimal for measuring the allelic ratios at both SNPs, the second exon SNP was reproducibly covered across all samples and allowed us to assess allelic expression for all KO clones.

We found that three of seven clones showed some evidence of allelic skewing based on SNP ratios. RNAseq data from WT OCI-AML3 cells showed a 2:1 ratio at both SNP sites consistent with *MYC* expression from all three alleles. Similarly, we observed a 2:1 SNP ratio for four of our EnhB KO clones, which is likely the result of expression from all three alleles (Figure 9, top). Three KO clones had a 1:1 SNP ratio which suggests that expression is restricted to two of three alleles (Figure xx, bottom). These data alone don't rule out the possibility of altered eDMR enhancer activity, as other compensatory methods may help maintain expression from the *MYC* locus in the absence of this downstream enhancer module but do seem to suggest that stable expression of *MYC* from all alleles is the preferred state in OCI-AML3 cells. It will be necessary to follow up with Micro-C studies on these deletion clones to determine if there are altered

looping dynamics between the *MYC* promoter and the downstream enhancer region upon eDMR deletion. These data could be used not only to phase the expression data with the deletion allele but also to indicate whether loss of the enhancer is reflected in reduced BENC interactions and/or increased interactions with alternative enhancer regions.

In line with this result, we also want to acknowledge that cancer cells are known to be addicted to *MYC* in certain cases, meaning they either die or find a way to maintain *MYC* expression against whatever regulatory barrier is put in the way. This may be true of OCI-AML3 cells, and if so, may be the reason we don't detect any viable clones with significantly decreased *MYC* expression. There is evidence in the literature describing the complexity of *MYC* regulation in cancer cells including examples of enhancer modularity with potential combinatorial activity, enhancer switching (use of the *PVT1* enhancer region during BETi-dependent loss of BRD4 binding in the BENC)[6,7], and other compensatory methods used for maintaining elevated expression. While our results are suggestive that these AML cells may utilize the EnhB eDMR for optimal *MYC* expression, loss of a single allele, or two alleles in the case of a triploid cell, it's not enough to significantly reduce cellular levels of MYC protein. However, it is certainly possible that further disruption of the broader enhancer region would compromise *MYC* regulation. Evidence supporting this hypothesis is the lack of success we had in generating EnhA deletion or dual enhancer deletion clones. The adjacent EnhA eDMR may be necessary for *MYC* regulation, in which case loss of the region either individually or in combination with the EnhB eDMR would be deleterious for cell growth and survival.

4.3.3 Deletion of the *ARID1a* eDMR correlates with a significant increase in *ARID1a* expression.

ARID1a eDMR KO experiments were overall less efficient, however, we were able to generate three clones with successful eDMR deletions. RNAseq data was generated and analyzed similarly to the *MYC* eDMR knockout data. We found that normalized *ARID1a* expression was not only increased in the KO clones compared with OCI-AML3 wild-type cells but also noticeably consistent amongst the biological replicates (average 13.2 vs. 11.5, respectively; p-value = 0.002; Figure 10). Though this result was opposite of what we hypothesized in our original model of enhancer hypermethylation, we did leave room for the possibility of enhancer hijacking or the use of alternative enhancers where eDMR deletion may favor rearrangement of the 3D interaction landscape. There are well-documented examples of alternative looping at imprinting loci, where hypermethylation of an imprinting control region facilitates interactions between the imprinted gene promoter and a distal enhancer element to drive expression. Although regulation of the *ARID1a* locus doesn't fall into the category of imprinting, it's reasonable to think that hypermethylation of regulatory elements could direct alternative looping interactions with the *ARID1a* promoter in favor of increased expression.

Another explanation for the increased expression of *ARID1a* is the possible loss of a suppressor-promoter interaction upon eDMR deletion. Although we annotated the *ARID1a* eDMR region as having active enhancer marks in primary AML samples and OCI-AML3 cells, it's possible that H3K27ac is helping to recruit suppressive factors to the locus that act as a "regulatory break" to keep expression within a specific range. Loss of such a region may result in a loss of balance or un-checked stimulatory factors that result in increased expression. Without 3D chromatin conformation data, it is difficult to say if altered looping dynamics are contributing to the increased expression of *ARID1a* in eDMR KO clones. CTCF ChIPseq data from OCI-AML3 WT

cells does indicate potential binding of the insulator protein, suggesting that hypermethylation could deplete the locus of CTCF binding and favor the use of an alternative enhancer. Analysis of CUT&TAG data from these clones will be useful for identifying nearby active enhancer regions that could feasibly loop onto the *ARID1a* promoter and drive expression.

4.4 Discussion and Future Directions

We believe these KO lines provide novel information about the function of the *MYC* and *ARID1a* eDMR elements and test two models that could explain their hypermethylation in AML cells. In the first model, the enhancer is not essential for regulating the expression gene expression and passively accumulates methylation during AML development. In the second, the eDMR does influence target gene expression to some degree, and its activity is altered by hypermethylation. If methylation accumulates at non-essential enhancers, we would have expected deletions of the eDMR regions to have no impact on gene expression. However, we observed changes in expression for both *MYC* and *ARID1a* after eDMR deletion, suggesting some regulatory functions of the eDMR regions and supporting the hypothesis that aberrant methylation contributes to the dysregulation of critical regulatory elements.

One reason for the subtle effects we observed in expression changes could be due to our selection of gene targets, which may be regulated by multiple enhancers. Other genes with fewer regulatory inputs may be more amenable to single-locus manipulation of functional regulatory elements to understand whether hypermethylation at the *IDH*^{mut}-specific DMRs in AML affects expression. Future experiments could include enhancers for additional genes that we found in our studies, including (but not limited to) *ASXL1*, *DOT1L*, *EGFRA*, *KDM2A*, *NOTCH*, and *SRSF3*, all of which are considered to play a role in hematopoietic development and form interactions

with regions that display differential hypermethylation specific to *IDH*^{mut} AML based on preliminary findings.

We would also like to generate Micro-C libraries for our *MYC* and *ARID1a* enhancer deletion clones and perform targeted capture experiments to measure changes in loop interactions converging at the *MYC* and *ARID1a* promoter. While our current data set measures the functional consequence of enhancer deletion on target expression, it doesn't capture any potential changes in the surrounding interaction landscape. Initial analysis will interrogate interactions gained or lost with the *MYC* eDMR or *ARID1a* eDMR respectively and follow up by surveying for new loops that may suggest alternative enhancer usage or regulatory compensation.

For these initial experiments, we opted to use OCI-AML3 cells as our model for enhancer knockout but believe performing knockout experimenters in other AML cell lines will add important insights. For example, the THP1 cell line is wild-type for the *IDH* gene and has a single copy of chromosome 8, which may simplify our interpretation of BENC eDMR deletions. We've opted to use AML cell lines for these experiments given that they are easier to manipulate and more homogenous than primary cells, however, we are aware that these cells tend to have significantly altered genomes and may not behave the same as primary cells. As an alternative approach in future experiments, we will consider using embryonic stem cells (ESCs) for enhancer knockout experiments in the future, as they may provide a more comparable or representative genome equivalent to primary AML samples. Using our hematopoietic differentiation protocol, we would generate hematopoietic progenitors for targeted enhancer knockout experiments and test gene expression and local 3D architectural changes accordingly.

Finally, in attempts to increase the throughput of eDMR enhancer activity, we have also considered the use of a CRISPR interference (CRISPRi) screen where Cas9 is fused to a KRAB repressive domain that when targeted to an enhancer region, could shut down or block the activity of the locus. Paired with single-cell RNAseq, the ability to correlate individual gRNAs with gene expression changes would allow us to narrow in on potential enhancers that we may have overlooked in epigenetic correlation approaches.

4.5 Acknowledgements

A special thank you to Heidi Struthers for performing all the RNAseq experiments reported in this chapter.

4.6 Figures

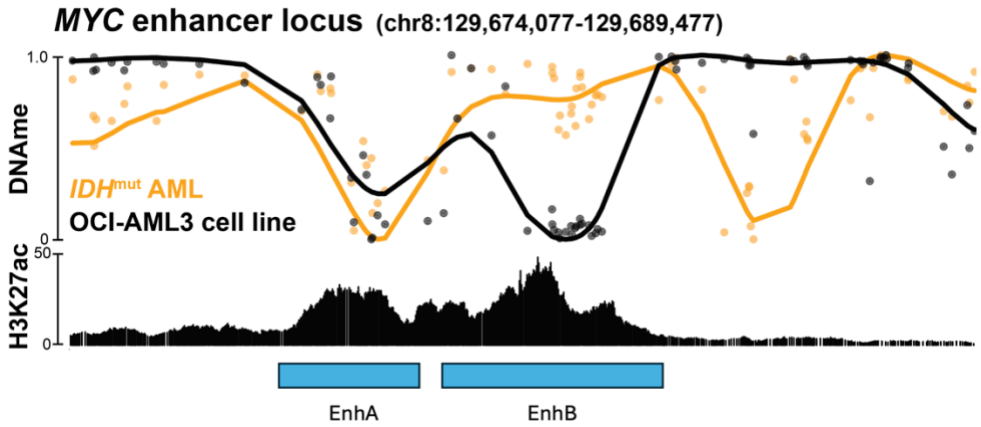


Figure 4.1. DNA methylation in *IDH*^{mut} AML and OCI-AML3 cell line at the *MYC* eDMR. Top: methylation tracks for *IDH*^{mut} AML cells (orange) and OCI-AML3 WT cell (black). Individual CpG values are smoothed across the locus for visualization. Bottom: H3K27ac signal from OCI-AML3 WT cells with neighboring peaks designated as EnhA and EnhB.

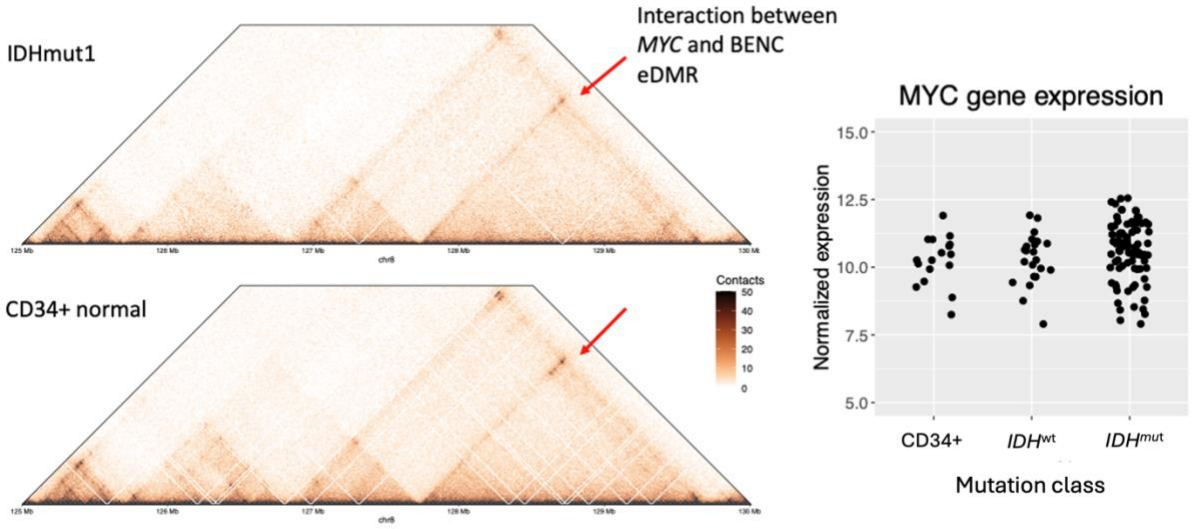


Figure 4.2. Regulation of *MYC*. Left: Interaction matrix surrounding the *MYC* locus in *IDH*^{mut} AML and CD34+ samples.

Darker pixels in the matrix represent increased interaction frequencies compared to background levels. Red arrows indicate point interactions between the *MYC* promoter and the eDMR regions in the BENC. Right: Normalized *MYC* expression in CD34+ cells, *IDH*^{mut} AML, and *IDH*^{wt} AML.

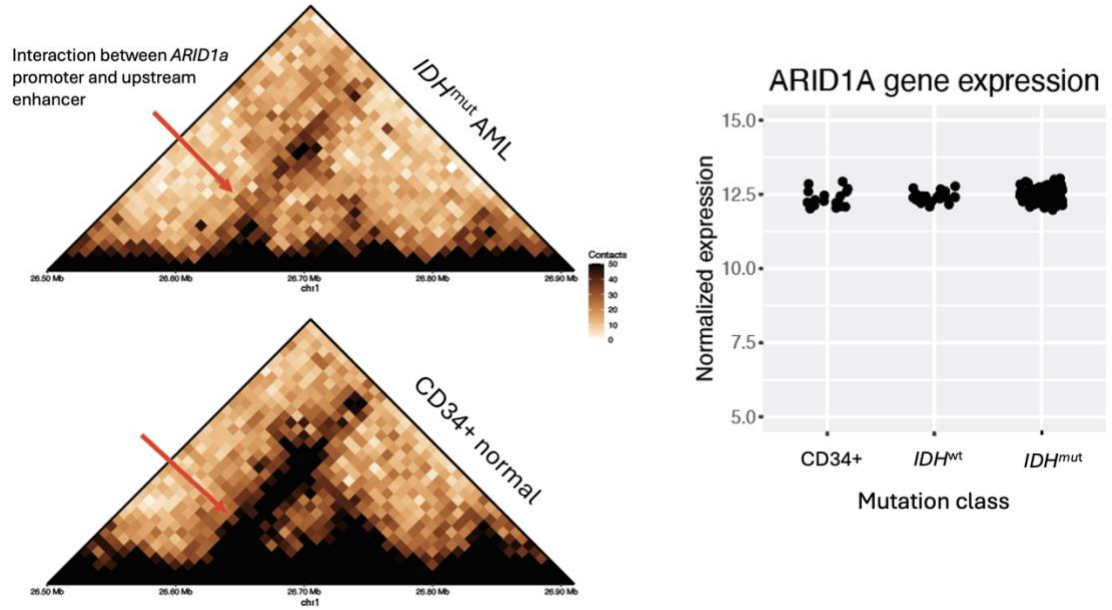


Figure 4.3. Regulation of *ARID1a*.

Left: Interaction matrix surrounding the *ARID1a* locus in *IDH*^{mut} AML and CD34+ samples. Darker pixels in the matrix represent increased interaction frequencies compared to background

levels. Red arrows indicate point interactions between the *ARID1a* promoter and the eDMR regions in the BENC. Right: Normalized *ARID1a* expression in CD34+ cells, *IDH*^{mut} AML, and *IDH*^{wt} AML.

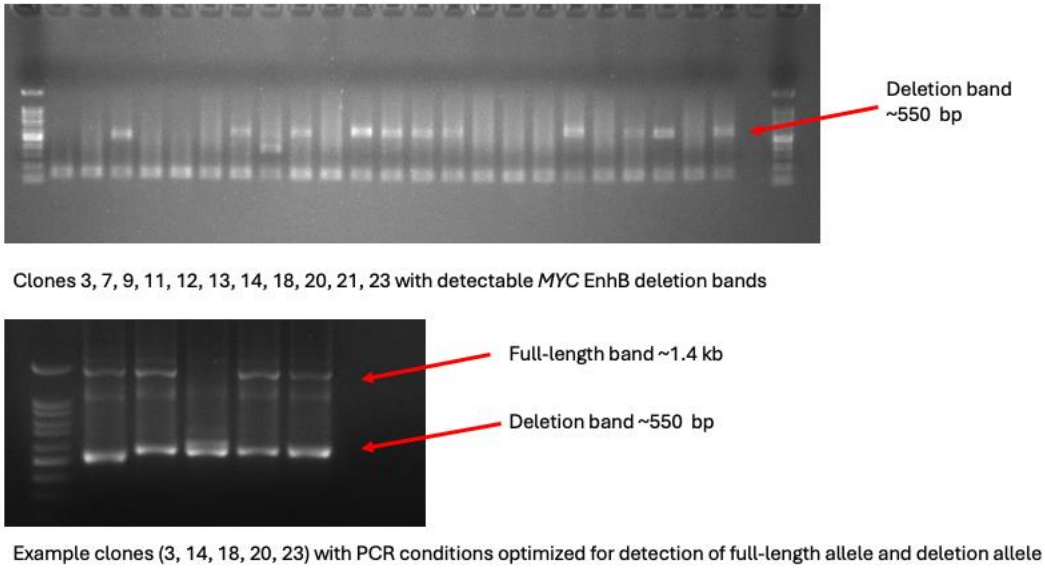


Figure 4.4. PCR genotyping for the *MYC* EnhB DMR deletion clones.

The size of deletion bands and full-length bands are annotated by red arrows.

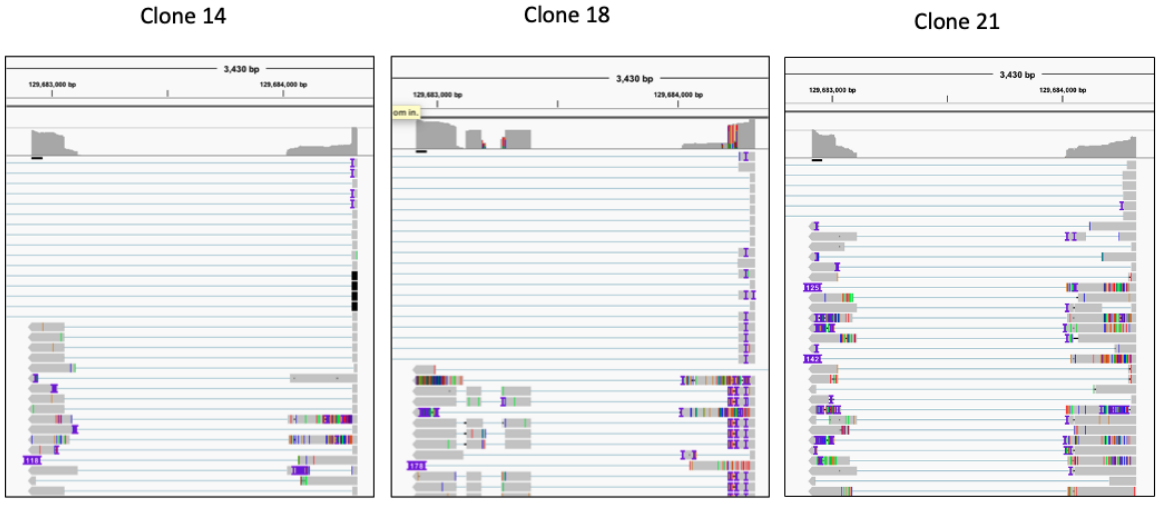


Figure 4.5. NGS sequencing reads of PCR amplification for the *MYC* EnhB deletion allele.

Three representative clones are shown to demonstrate the size of the deletions.

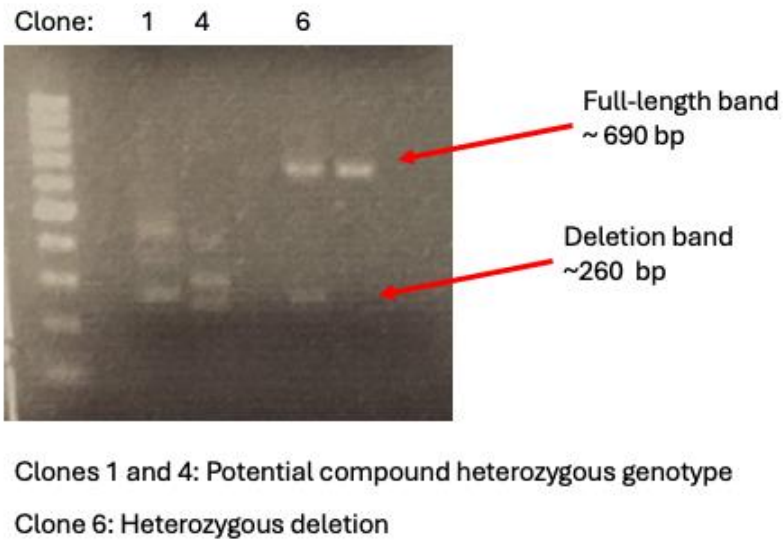


Figure 4.6. PCR genotyping for the *ARID1a* DMR deletions.

The size of deletion alleles and full-length alleles are annotated by red arrows.

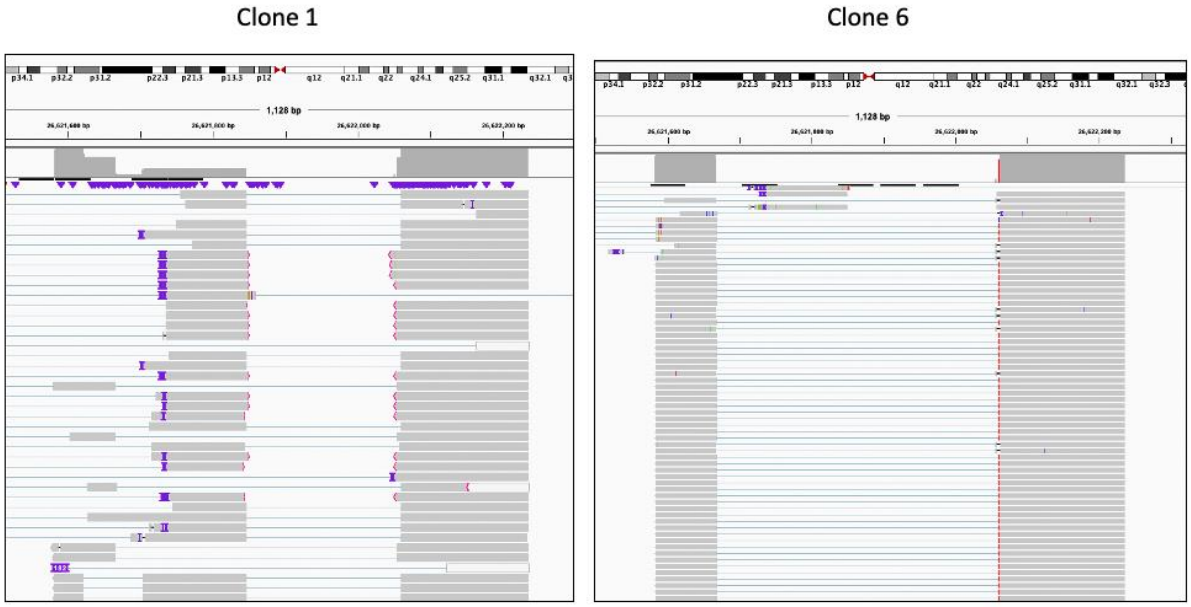


Figure 4.7. NGS sequencing reads of PCR amplification for the *ARID1a* eDMR deletion allele.

Three representative clones are shown to demonstrate the size of the deletions.

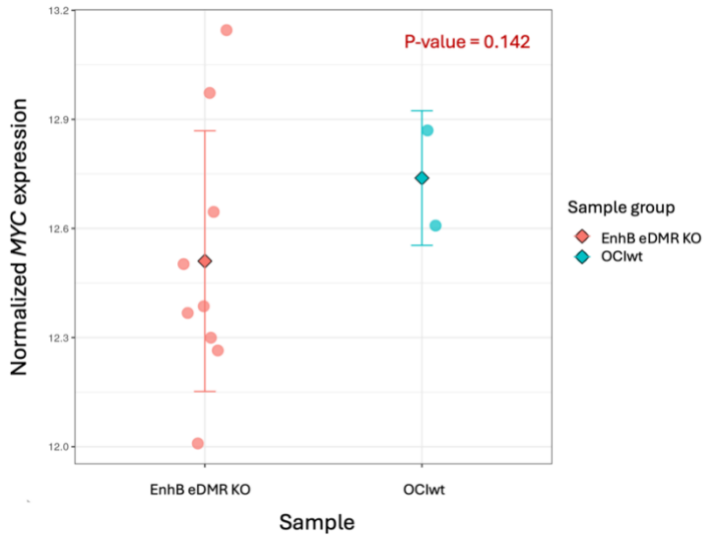
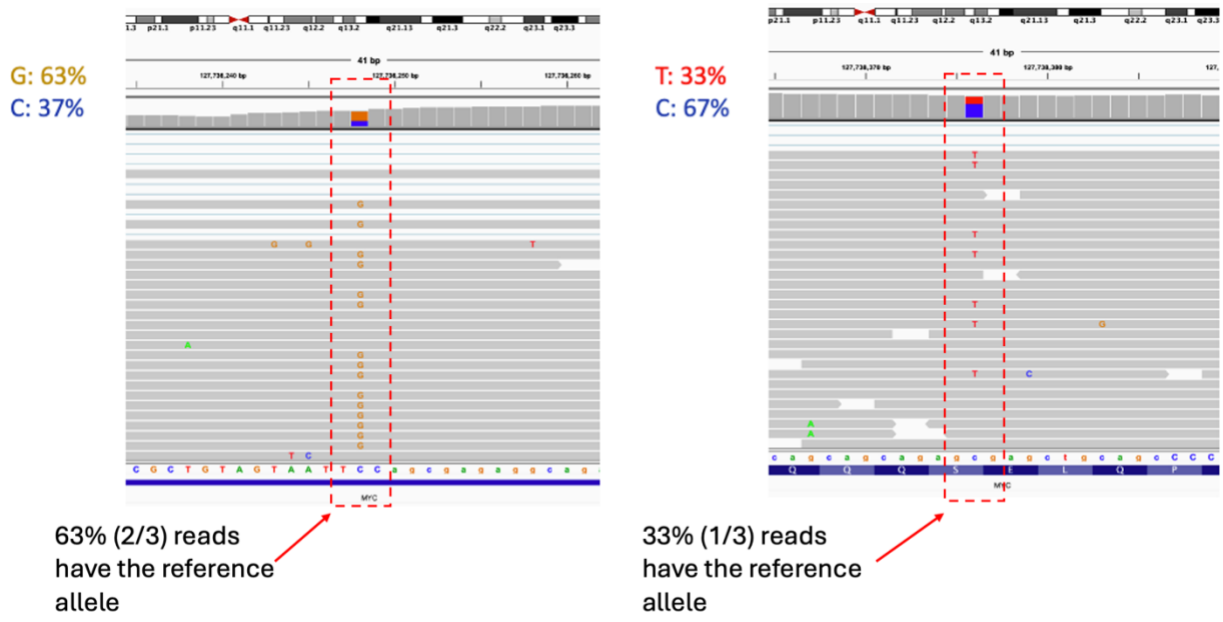


Figure 4.8. Normalized *MYC* expression in EnhB eDMR KO clones and OCI-AML3 WT comparators.

KO clones have moderately decreased expression of *MYC* compared to the WT control (p-value = 0.142).



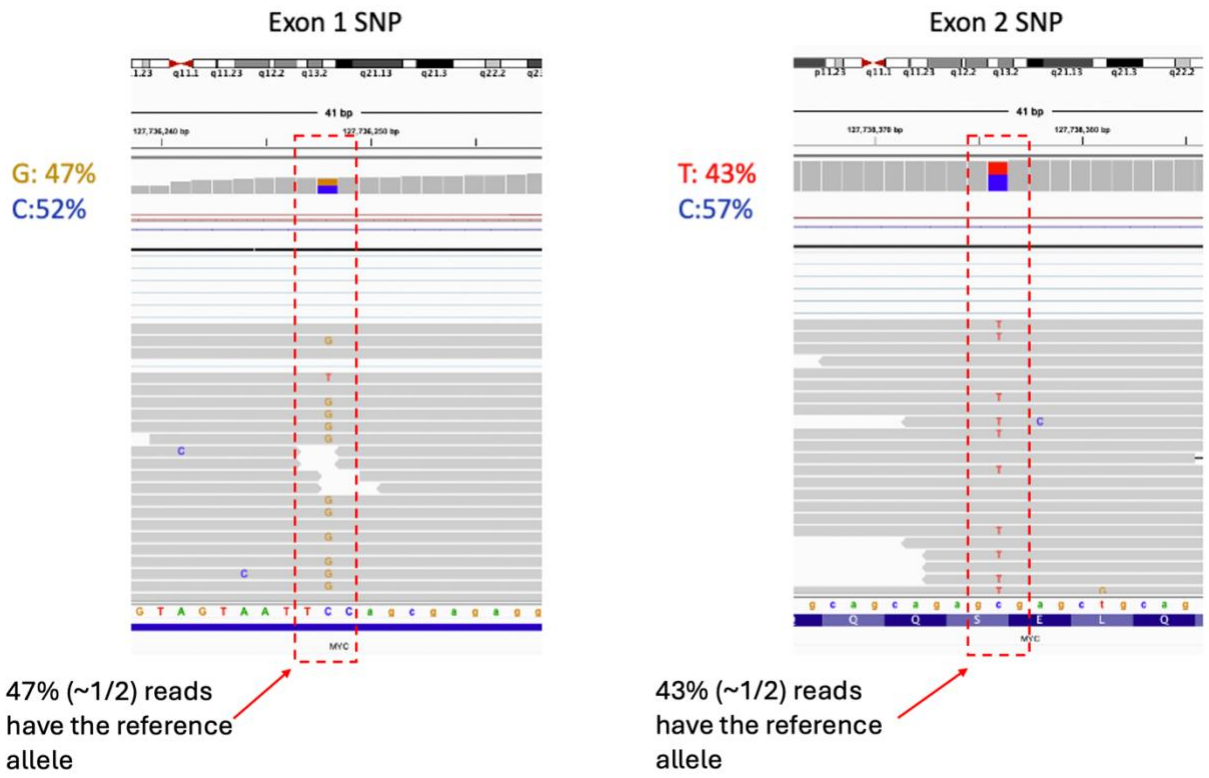


Figure 4.9. SNP analysis in exon 1 and exon 2 of *MYC*.

IGV browser views of RNAseq reads aligned at individual SNP loci in the first and second exons of *MYC*. Alternate bases (SNP bases) are designated by color. Top: Example clone with 2:3/1:3 SNP ratios. Bottom: Example clone with 1:1 SNP ratios. Red boxes annotate the SNP. Red boxes highlight the SNP locus.

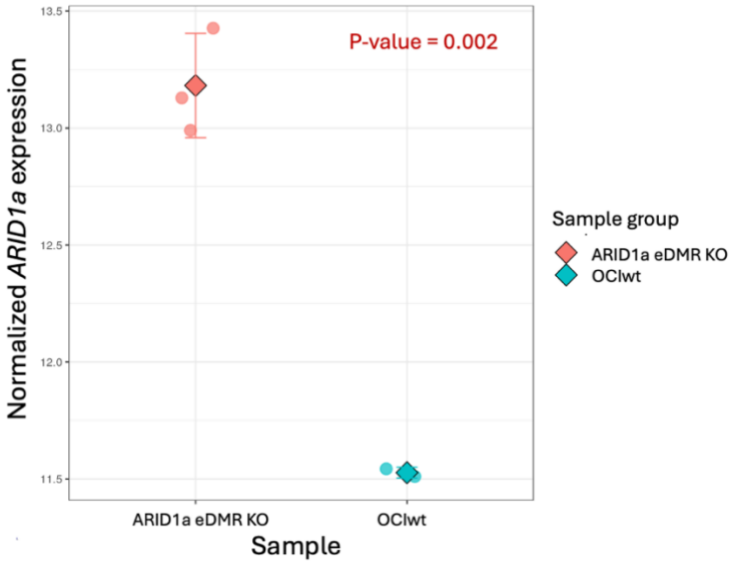


Figure 4.10. Normalized *ARID1a* expression in eDMR KO clones and OCI-AML3 WT comparators.

KO clones have significantly *ARID1a* increased expression compared to the WT control (p-value 0.002).

4.8 References

1. Bahr, C. *et al.* A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature* **553**, 515–520 (2018).
2. Shi, J. *et al.* Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes Dev.* **27**, 2648–2662 (2013).
3. Cimmino, L. *et al.* Restoration of TET2 Function Blocks Aberrant Self-Renewal and Leukemia Progression. *Cell* **170**, 1079-1095.e20 (2017).
4. Han, L. *et al.* Chromatin remodeling mediated by ARID1A is indispensable for normal hematopoiesis in mice. *Leukemia* **33**, 2291–2305 (2019).

5. Kamori, T. *et al.* The effects of ARID1A mutations on colorectal cancer and associations with PD-L1 expression by stromal cells. *Cancer Rep*. **5**, e1420 (2021).
6. Parolia, A., Cieślik, M. & Chinnaiyan, A. M. Competing for enhancers: PVT1 fine-tunes MYC expression. *Cell Res.* **28**, 785–786 (2018).
7. Schuijers, J. *et al.* Transcriptional Dysregulation of *MYC* Reveals Common Enhancer-Docking Mechanism. *Cell Rep.* **23**, 349–360 (2018).

Chapter 5: Conclusions and Future Directions

Elisabeth R. Wilson and David H. Spencer

IDH1 and *IDH2* are two of the most mutated genes in AML patients, accounting for roughly 20% of newly diagnosed cases. It is well established that recurrent gain-of-function mutations in these genes increase DNA methylation in AML through the production of the oncometabolite 2-HG that inhibits TET2 demethylation activity. However, the precise genomic locations and functional consequences of these changes have not previously been clearly defined. Previous studies have surveyed the hypermethylation landscape of IDH^{mut} AML using array-based methods and reduced representation bisulfite sequencing, which capture only a fraction of the genome known to be enriched in CpG dinucleotides (i.e. CpG islands within gene promoters). Although these studies identified a clear hypermethylation phenotype, the functional consequences have been difficult to define given the lack of correlation with changes in gene expression.

Our comprehensive genome-wide study of methylation changes in primary AML samples with *IDH* mutations expands upon previous findings by highlighting that the most profound methylation changes occur at regulatory enhancers. We were able to leverage our comprehensive whole-genome bisulfite sequencing datasets to detect what we believe to be truly IDH^{mut}-specific DNA methylation changes. We found that these changes were not diffusely spread through the genome but were rather concentrated as focal changes marked by hypermethylation compared to both normal CD34+ cells and AML cells without *IDH* mutations. Moreover, differentially

methylated regions were enriched for active enhancer modifications in normal CD34+ progenitor cells and primary AML samples, suggesting they may play important regulatory roles in hematopoiesis. DNA methylation is generally thought to have a repressive effect on genome function. We sought to better understand the functional consequences of enhancer hypermethylation in *IDH*^{mut} AML and gain insights into why enhancers are uniquely affected by these methylation changes.

In our survey of subtype-specific methylation patterns, we were surprised to find that canonical mutations in the two *IDH* genes had slightly different hypermethylation signatures. Notably, *IDH2*^{mut} AMLs in our dataset had less pronounced hypermethylation than those with *IDH1* mutations, even though both were hypermethylated at a highly overlapping set of loci. The reason for a more subtle *IDH2* mutant methylation phenotype is still unclear. However, patients with *IDH2* mutations tend to have better survival outcomes than those with *IDH1* mutations, suggesting this genomic difference may have important biological and clinical implications. One possible explanation for this phenomenon is the difference in subcellular localization of the two mutant enzymes—IDH1 participates in the citric acid cycle in the cytosol while IDH2 is active in the mitochondria. If the oncometabolite 2HG originating in the cytosol diffuses into the nucleus more readily, then *IDH1* mutations could have a more potent effect on TET enzymes. Our analysis revealed that AMLs with biallelic inactivating *TET2* mutations exhibited an even less pronounced methylation pattern compared to *IDH* mutant AML, despite the fact that many of the hypermethylated DMRs identified in these samples overlapped *IDH*^{mut} -specific DMRs. Additionally, our analysis using oxidative bisulfite sequencing demonstrated elevated levels of 5-hydroxymethylcytosine (5hmC) in these regions in AML samples that were wild-type for *TET2*, *IDH1*, and *IDH2*. Conversely, 5hmC levels were notably lower in samples with mutations in

IDH or *TET2* in these regions. Together, these data necessitate revisions to previous assumptions in the field that *IDH1*, *IDH2*, and *TET2* mutations all have the same effects. While our data supports previous models where *IDH* mutations induce increased DNA methylation by interfering with TET-mediated DNA demethylation, it also suggests there may be *TET2*-independent factors that contribute to the *IDH* mutant hypermethylation phenotype.

Beyond defining subtype-specific methylation patterns, we also gained insight into the dynamics of *de novo* DNA methylation and active demethylation pathways in normal hematopoietic cells and AML. The fact that IDH^{mut} associated hypermethylation occurs at regions with low levels of DNA methylation in normal CD34+ cells suggests that *de novo* DNA methylation and TET-mediated demethylation are both active in these regions, despite their low steady-state methylation levels. This is supported by the observation that AML samples with co-occurring *IDH* and $DNMT3A^{R882}$ mutations show significantly attenuated hypermethylation and that IDH^{mut}-specific DMRs have high levels of 5hmC, which is produced from 5mC as a substrate. Remodeling of DNA methylation by these processes in specific regions has been reported previously in studies of embryonic stem cells, which have shown that methylation and active demethylation are in equilibrium at many loci and may be maintained by the occupancy of methylation and demethylation complexes. Our analysis suggests this equilibrium also exists in normal hematopoietic stem/progenitor cells and is disrupted in the presence of mutant *IDH* alleles, leaving *de novo* DNA methylation unopposed. The focal nature of IDH^{mut} associated hypermethylation implies that the activity (or occupancy) of DNMT3A and TET enzymes is not diffuse and may instead be targeted to specific genomic regions. The genomic or epigenetic features directing this activity are unclear, but the enrichment of IDH^{mut} DMRs in active enhancers suggests that components of active chromatin may recruit methylation and

demethylation machinery. The convergence of these processes at enhancers could provide clues as to why mutations with opposite effects on DNA methylation both contribute to AML development via dysregulation of common target genes.

In addition to investigating the factors that predispose regions to *IDH*^{mut}-specific hypermethylation, we also wanted to characterize the potential functional consequences of these changes. One of the possible consequences of altered enhancer methylation could be a change in the expression of target genes. Our initial analysis of 3D genome interactions using Hi-C data from a limited set of normal CD34+ cells and AML cases identified a set of genes that form interactions with *IDH*^{mut}-specific DMRs. In normal CD34+ cells and *IDH*^{mut} AML, these genes were enriched for highly expressed genes compared with the full distribution of expressed genes and those opposite a commonly hypermethylated region. Although increased methylation at these loci was not associated with repressed chromatin or lower gene expression in *IDH*^{mut} AML samples, this finding suggested that *IDH*^{mut} associated hypermethylation affects the regulatory sequences of genes that may contribute to AML pathogenesis. While none of our Hi-C samples had *IDH* mutations, this analysis was key for identifying generalized features of eDMR target genes. What we lacked was an ability to correlate enhancer methylation status with the strength of interactions, or even the adaption of novel interactions.

To follow up on these preliminary findings, we performed Micro-C on 15 AML samples and 2 CD34+ normal donor samples (*IDH*^{mut}, n=4; *DNMT3a*^{R882}, n=6; *MLL* fusions, n=3; *RUNX1-RUNX1T1*, n=2). We additionally performed a hybridization-capture of our Micro-C libraries to enrich the interactions converging on gene promoters across the genome. Importantly, these Micro-C experiments improved upon our previous data set by increasing the overall resolution at which we could identify chromatin interactions (particularly enhancer-promoter interactions) and

by extending chromatin conformation data across *IDH*^{mut} and wild-type cases. While this dataset contains a wealth of information, we focused on defining and characterizing the intersection of chromatin interactions with *IDH*^{mut}-specific DMRs. Specifically, we used the Micro-C data to call genome compartments, topologically associated domains (TADs), and chromatin loops in all AML and normal samples. Each set of features was assessed for overlap with the DMRs and summarized based on methylation phenotype (*IDH*^{mut} vs. CD34+ normal and *IDH*^{wt}). At all levels of nuclear organization, we identified unique features of chromatin organization in *IDH*^{mut} AML and/or the set of interactions involving *IDH*^{mut}-specific DMRs, suggesting that hypermethylation occurs at regions relevant to genome organization.

Our analysis of compartments showed significant enrichment of *IDH*^{mut}-specific DMRs in the active genome compartment compared with the control set of commonly hypermethylated regions, consistent with our hypothesis that these regions preferentially interact with transcriptionally active loci. An unsupervised analysis also demonstrated unique compartmentalization in *IDH*^{mut} AML, stratifying these samples separately from *IDH*^{wt} AML CD34+ and normal cells. When we looked at the intersection of DMRs with TAD boundaries, we found a slight overall decrease in boundary strength in *IDH*^{mut} AML compared with wild-type samples and normal cells. Similarly, at the level of looping interactions, we found that differential interactions between *IDH*^{mut} and normal cells were heavily skewed towards having decreased frequency in AML when they intersected a DMR. Many of the interactions coincided with CTCF binding sites, regions where this methylation-sensitive protein helps orchestrate looping, providing a potential explanation for the reduction in interaction frequency. Additional measurements of CTCF binding in these primary samples will be necessary to determine the impact of methylation on CTCF-mediated looping dynamics.

To follow up on observations from our previous studies of whole-genome bisulfite sequencing and 3D chromatin architecture data, we selected candidate hypermethylated enhancers for mechanistic study. We focused on the *MYC* locus which had robust interactions with a previously identified superenhancer region containing focal hypermethylation specific to *IDH* mutant AML. Three regions within the superenhancer showed H3K27ac signal coinciding with DMRs, two of which we attempted to knock out using CRISPR-Cas9 based editing strategies. To measure the consequences of enhancer DMR knockout, we performed RNAseq to measure any coincident changes in *MYC* expression and/or histone modifications surrounding the regulatory region.

These experiments were designed to test two models that could describe the effect of hypermethylation on the *MYC* eDMR in human AML. In our first model, the enhancer region has regulatory input on *MYC* expression, and its activity is altered by the hypermethylation we observed in IDH^{mut} AML. In this case, loss of the enhancer may cause *MYC* expression to decrease. Although the influence of the cluster of enhancers at this locus is complex and some level of redundancy is expected, previous studies have shown that knockout of even a single enhancer can have measurable activity. Nonetheless, we anticipated that our experiments in transformed cell lines might have only modest changes in *MYC* expression. A second hypothesis is that the eDMR is completely inactive and does not regulate *MYC* expression at all. Hypermethylation in this scenario accumulates due to this inactivity. In this case, no changes in *MYC* expressions would be expected upon deletion of the eDMR.

We generated seven clones with deletions of the eDMR and observed that *MYC* expression trended lower compared with unmodified cells. The deletions all spanned roughly the same region and there was no apparent correlation between deletion size and expression difference.

One of the challenges we experienced in interpreting expression results from OCI-AML3 eDMR KO cells was an additional allele of *MYC* on the +8 chromosome. Since clones were genotyped using PCR and sequencing, we did not have the exact genotype for each clone, making it difficult to determine how many alleles were successfully edited. We looked for patterns of allele skewing in the RNAseq data that suggested allele-specific expression due to loss of the cis-acting enhancer region. Although we observed some evidence of skewing in three of seven clones, the pattern was inconsistent such that definitive correlation between the presence of a deletion and expression remained unclear.

Interestingly we found an opposite result when we deleted the eDMR region contacting the *ARID1a* promoter in OCIAML3 cells. Rather than observing a decrease in expression, which we might expect if enhancer hypermethylation inhibits its activity, we observed a consistent increase in *ARID1a* expression for all three of our knockout clones compared to OCIML3 wild-type cells. Similar to *MYC*, *ARID1a* is encoded on a triploid chromosome in OCIML3 cells, complicating our interpretation of expression results. Although additional sequencing is required to precisely genotype these cells, PCR results suggested that two of the three clones were compound heterozygous for eDMR deletions and lacked a wild-type allele, while the third clone had one copy knocked out leaving at least one intact wild-type allele. The fact that all three clones had similarly increased expression suggests the loss of even one copy of the eDMR is enough to change overall *AIRD1a* levels. The loss of additional alleles did not appear to have additive effects, perhaps due to feedback mechanisms that prevent expression from exceeding a certain level.

The eDMRs involving *MYC* and *ARID1a* represent a very small fraction of the total regions impacted by IDH^{mut}-specific hypermethylation. However, these regions are consistent with our

overall observations that *IDH*^{mut} associated hypermethylation occurs in regions that may regulate highly expressed genes in AML. They also serve as two different contexts in which hypermethylation may have consequences on gene expression. Decreased gene expression upon enhancer deletion may be the result one predicts *a priori* based on our general understanding of methylation/function relationships. However, the *ARID1a* locus provides an alternative example, where hypermethylation may have the opposite effect. While far from conclusive, these two examples suggest that a diversity of underlying mechanisms may be relevant when characterizing the impact of hypermethylation on genome regulation and gene expression.

We envision future experiments to study other eDMR regions identified in our genome-wide analysis. One approach is a methylation reporter assay that could efficiently measure the intrinsic activity of the DNA sequences of a large fraction of eDMRs. This would allow for the prioritization of interesting 'hits' for subsequent CRISPR editing approaches to either delete or modify their native context. We initiated such an approach using a published reporter plasmid that allows putative cis-acting regulatory sequences to be inserted upstream of a barcoded reporter gene. RNA sequencing of the reporter gene with the barcode would therefore provide a digital readout of the influence of the regulatory element on the expression level. This construct was also selected for its lack of CpGs in the backbone, providing an optimal vector for *in vitro* treatment with a DNA methyltransferase enzyme without having consequences on critical regulatory elements (other than insert itself). Methylated and unmethylated libraries could then be transfected into cells where cellular processes can act on the reporter plasmids and drive expression accordingly. Collecting and measuring RNA and DNA in tandem would then be used to define enhancer sequence activity as a function of methylation status. We have selected 46 eDMR elements for this experiment, which were prioritized based on consistent

hypermethylation signal across *IDH*^{mut} AML samples and CpG density to provide ample substrate for methylation. Initial cloning attempts proved challenging, and we had minimal success getting a diverse representation of species in our reporter library. Assembly approaches and ligation-based methods were both attempted using varying ratios of vector to insert, all of which failed to produce a successful library. Future experiments will be important for optimizing the cloning procedure to efficiently express all 46 eDMR elements selected (or any additional elements of interest). We believe these experiments will not only identify additional candidate enhancers where hypermethylation may be consequential to the regulatory landscape and/or gene expression but will also provide a more systematic approach to identifying characteristics of regulatory elements that are susceptible to hypermethylation. It will be particularly interesting to look for correlations between baseline enhancer activity, degree of hypermethylation in AML, and the consequences on reporter expression.

While reporter assays provide a massively parallel approach for screening eDMR activity, they fail to recapitulate the chromatin context within which these enhancers are operating. To address the direct consequences of focal hypermethylation at regulatory sequences residing in their native context, we have started to develop a system that pairs the localization of a catalytically dead Cas9 enzyme (dCas9) with the methyltransferase activity of DNMT3A to deposit methylation at candidate eDMRs. We currently have multiple clonal AML cell lines that express an inducible SunTag/DNMT3A system. However, additional experiments are necessary to determine the on-target efficiency and specificity of the system in each clonal line before testing any hypotheses on the effects of hypermethylation. Using RNAseq and Capture-C we believe this system will help to better characterize the role of focal hypermethylation at critical regulatory elements in their native context. We foresee the use of RNAseq and Micro-C as

important read-out assays for measuring the consequences of enhancer hypermethylation on cognate gene expression (and potential downstream targets), as well as changes in the looping interactions with the candidate target (loss of previously observed loops, neo-loops, etc.).

Our systematic studies of genome-wide methylation and 3D chromatin architecture suggest a potential role for altered methylation in regulating genome organization and gene expression in AML. We acknowledge that methylation is only a part of the larger regulatory story and that other factors contribute to gene regulation at these loci. We hypothesize that DNA methylation may be responsible for the 'fine-tuning' of gene expression, which may be especially important at certain developmental phases during hematopoiesis. Although the direct effects of *IDH*^{mut}-specific hypermethylation may be difficult to ascertain in AML cells, they may signal DNA elements that are critical for proper gene regulation at a specific point in the path between normal stems and full transformed leukemia.

Supplemental Figures

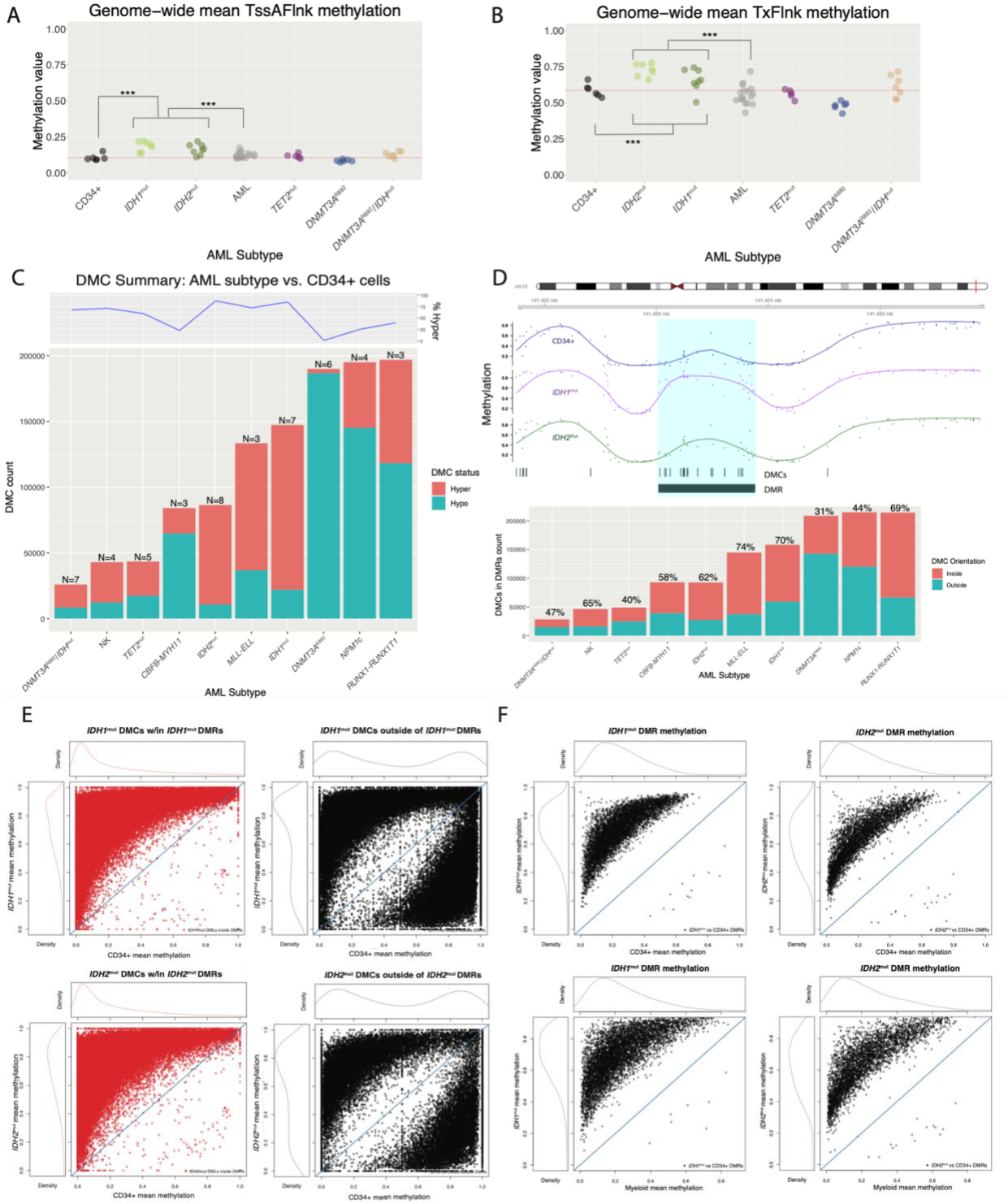


Figure S1. *IDH*^{mut} AMLs exhibit the greatest proportion of focal increases in methylation compared with CD34+ cells. A. DMC summary for individual AML mutational subtypes compared with CD34+ cells. DMC differential methylation status is classified as either hypermethylated (orange bars) or hypomethylated (teal bars) with respect to CD34+ cells. B. Top panel: Example locus with a focal enrichment of dmCpGs where the highlighted cytosines were identified as constituents of a DMR with statistically increased methylation in *IDH*^{mut} samples compared with CD34+ cells. Bottom panel: Summary of DMCs localizing within/outside DMRs identified in individual subtypes vs. CD34+ cells. C-D. Average methylation levels across TssAFlnk (C) and TxFlnk (D) ChromHMM regions in CD34+ cells (N=6) and AML subtypes (*IDH1*^{mut} or *IDH2*^{mut}, n=15; *TET2*^{mut}, n=5; *DNMT3A*^{R882}, n=6; *DNMT3A*^{R882}/*IDH*^{mut}, n=7; normal karyotype with *NPM1c* and wild-type *IDH1*, *IDH2*, *TET2*, and *DNMT3A,* n=4; Normal karyotype with wild-type *NPM1*, *IDH1*, *IDH2*, *TET2*, and *DNMT3A*, n=4; *CBFB-MYH11*, n=3; *KMT2A-ELL*, n=3; *RUNX1-RUNX1T1*, n=3). E. Distribution of mean methylation values across *IDH*^{mut} DMCs localizing within (left panels) or outside (right panels) *IDH1*^{mut} (top panels) and *IDH2*^{mut} (bottom panels) DMRs in *IDH*^{mut} AML samples vs. CD34+ HSPCs. F. Distribution of mean methylation values across *IDH1*^{mut} (top panels) and *IDH2*^{mut} (bottom panels) DMRs in *IDH*^{mut} AML samples vs. normal myeloid cells (n=3 promyelocyte samples; n=3 polymorphonuclear leukocyte samples; n=2 monocyte samples).

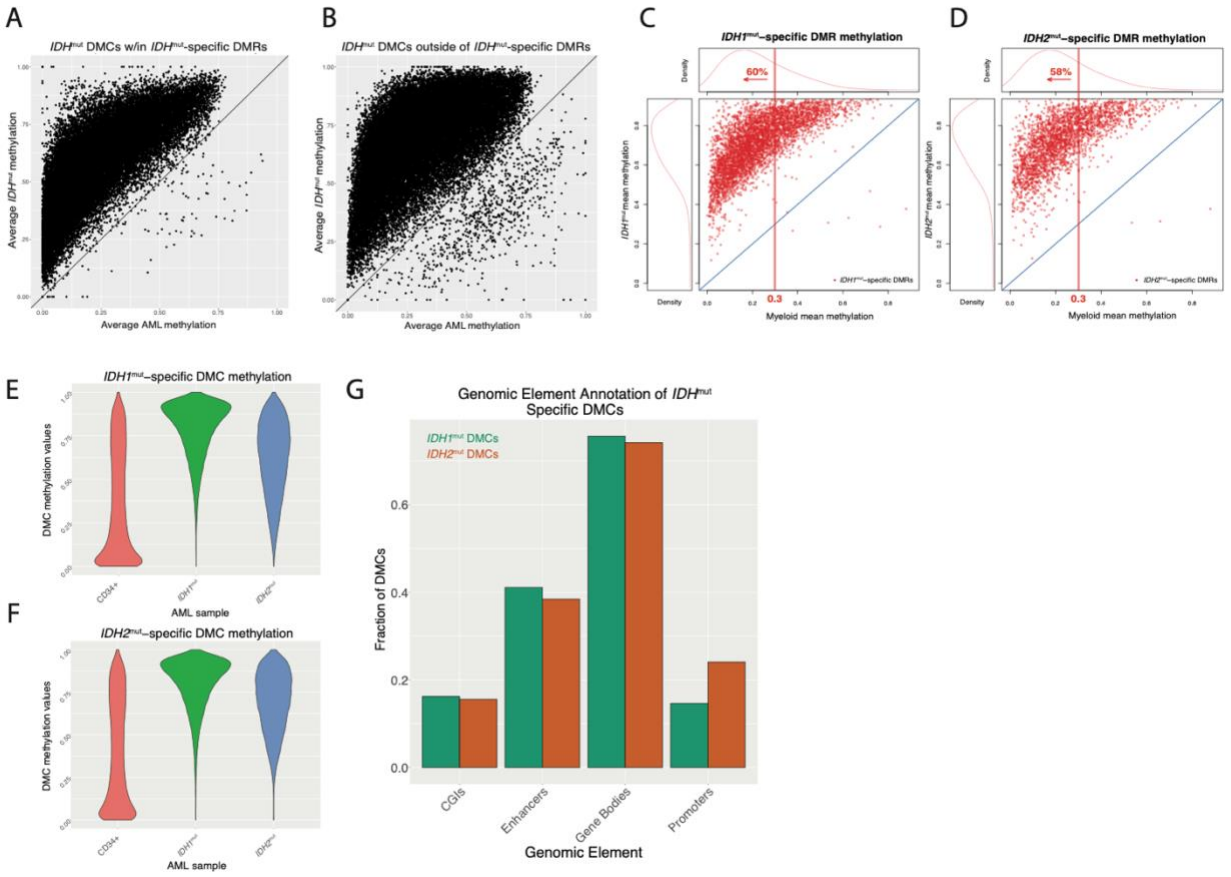
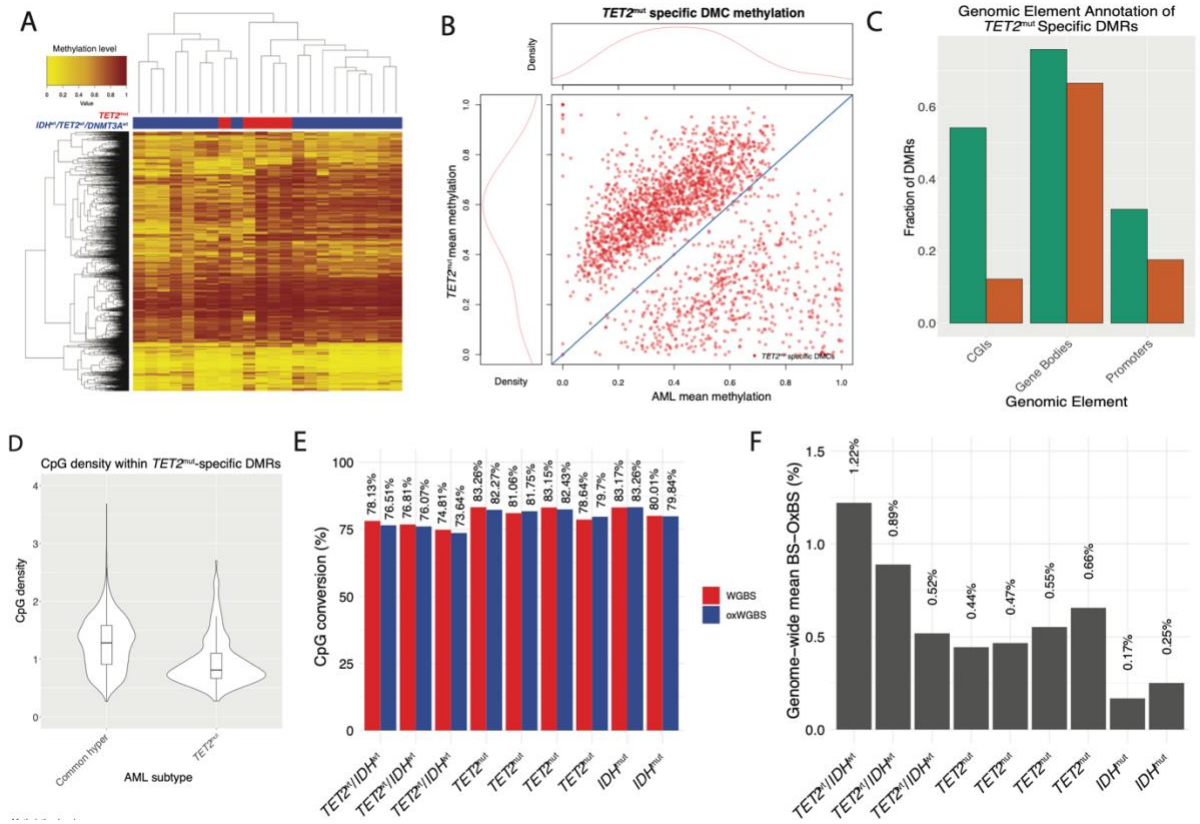


Figure S2. *IDH*^{mut}-specific DMCs exhibit similar methylation patterns and genomic annotations as *IDH*^{mut}-specific DMRs. A-B. Distribution of mean methylation across *IDH*^{mut} DMCs localizing within (A) or outside of (B) *IDH*^{mut}-specific DMRs in *IDH*^{mut} samples vs. all other *IDH*^{wt}/*TET2*^{wt}/*DNMT3A*^{wt} AMLs. C-D. Distribution of mean methylation values across *IDH1*^{mut}-specific (C) and *IDH2*^{mut}-specific (D) DMRs in *IDH*^{mut} AML samples vs. normal myeloid cells (n=3 promyelocyte samples; n=3 polymorphonuclear leukocyte samples; n=2 monocyte samples). E-F. Violin plots of methylation levels across all *IDH1*^{mut}-specific (E) and *IDH2*^{mut}-specific (F) DMCs in CD34+ cells, *IDH1*^{mut} samples, and *IDH2*^{mut} samples. G. Fraction of *IDH1*^{mut}-specific DMCs and *IDH2*^{mut}-specific DMCs overlapping functional genomic elements including CpG islands (CGIs), enhancers, gene bodies, and promoters.



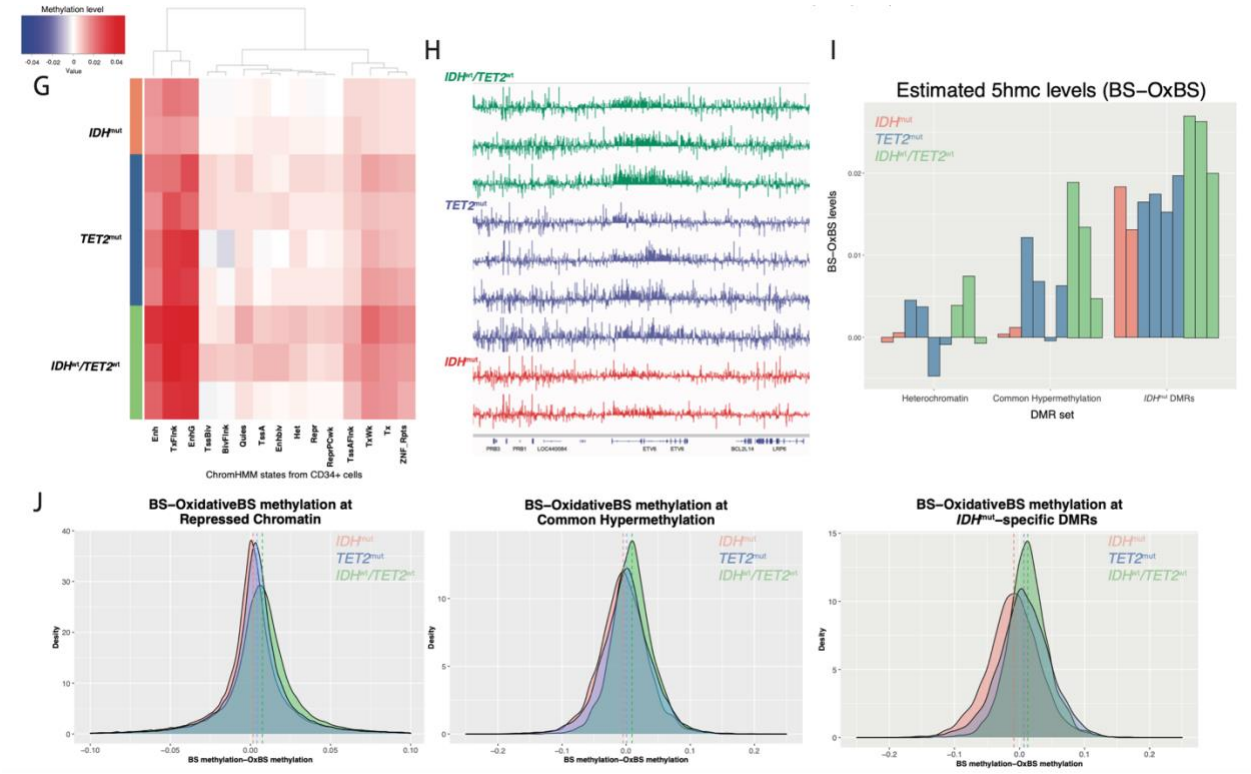


Figure S3. *TET2*^{mut}-associated hypermethylation is distinct from canonical CpG-island hypermethylation and is consistent with regions of increased TET2 hydroxymethylation activity in *TET2*^{wt} cells**.** A. Hierarchical clustering of *TET2*^{mut} AMLs with all other *IDH*^{wt}, *TET2*^{wt}, and *DNMT3A*^{wt} AMLs at the set of *TET2*^{mut} vs. CD34+ DMRs. Column annotation represents mutational classification by sample. B. Distribution of *TET2*^{mut}-specific DMC mean methylation in *TET2*^{mut} samples vs. the set of *IDH*^{wt}/*TET2*^{wt}/*DNMT3A*^{wt} AMLs. C. Percent overlap of generic AML-associated hypermethylation and *TET2*^{mut}- specific DMRs with defined genomic annotations. D. Distribution of CpG density across the set of commonly hypermethylated CpG islands and *TET2*^{mut}- specific DMRs. E. CpG conversion rate of 9 paired whole-genome bisulfite and oxidative bisulfite prepared libraries (n=2 for *IDH*^{mut} ; n=4 for *TET2*^{mut}; n=3 for *IDH*^{wt}/*TET2*^{wt}). F. Genome-wide average 5hmc levels across ~10.6 million CpGs with > 10x coverage in each of the paired samples, as calculated by subtracting oxidative bisulfite levels from bisulfite levels. G. Hierarchical clustering of mean 5-hmC levels across 15 chromatin states defined in CD34+ cells for individual patient samples. Columns represent chromatin states and rows represent individual patient samples. Vertical color blocks indicate *IDH* and *TET2* mutation status of individual samples. H. Example locus encompassing the *ETV6* gene with diminished 5hmc levels in two *IDH*^{mut} samples (red tracks) and four *TET2*^{mut} samples (blue tracks) compared with three *IDH*^{wt}/*TET2*^{wt} samples (green track). I. Mean 5-hmC levels in *IDH*^{mut}, *TET2*^{mut}, and *IDH*^{wt}/*TET2*^{wt} patient samples at heterochromatic regions, commonly hypermethylated regions, and IDHmut-specific DMRs. J. Distribution of estimated 5hmc levels across 4008 *IDH*^{mut} – specific DMRs, 4586 commonly hypermethylated regions in AML, and ~105,500 heterochromatic regions summarized by mutation.

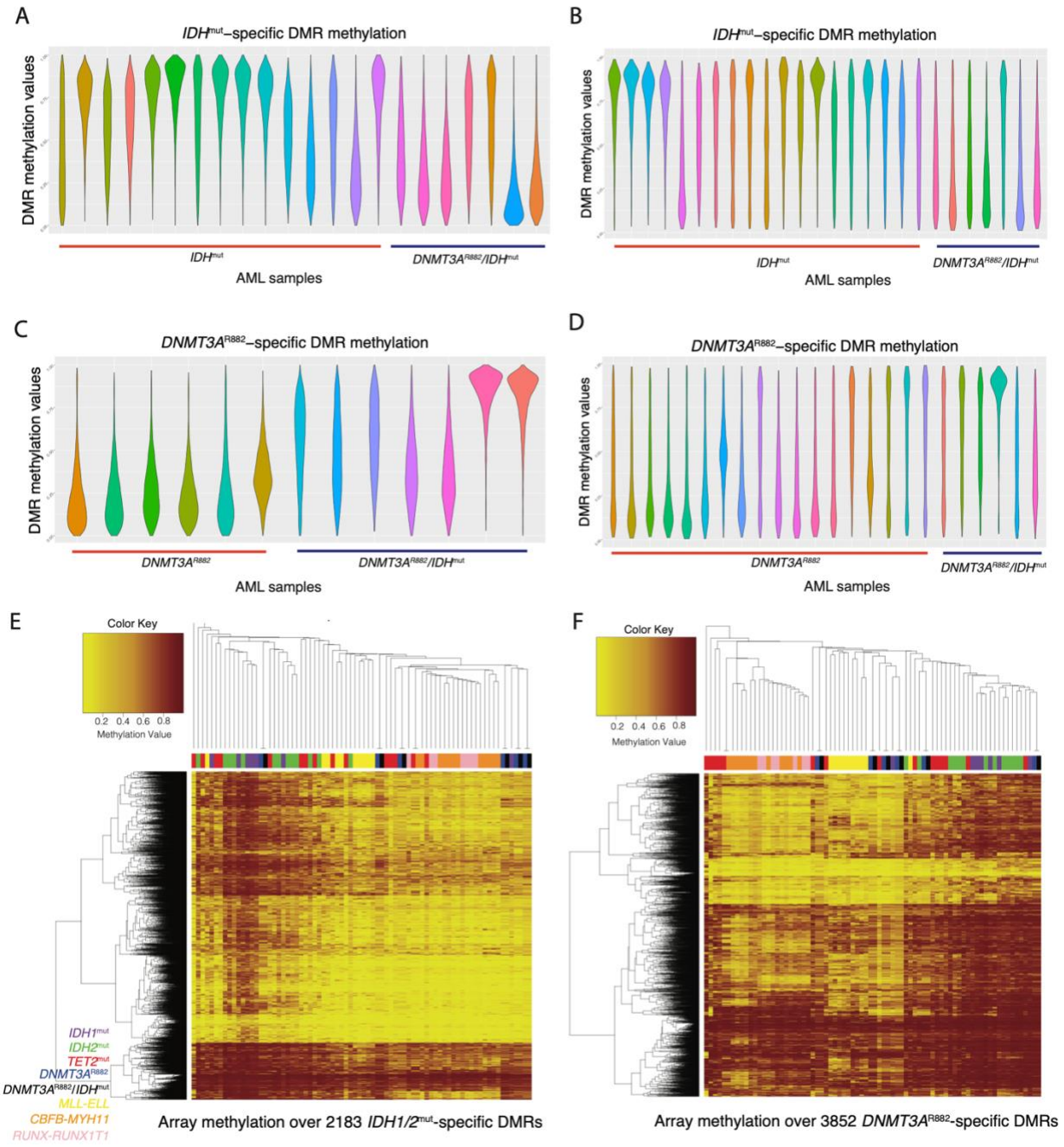


Figure S4. Individual samples with *IDH* mutations alone and in combination with *DNMT3A-R882* exhibit group level methylation trends at *IDH*^{mut}-specific and *DNMT3A-R882* DMRs. A. Methylation values across *IDH*^{mut}-specific DMRs in a set of 15 *IDH*^{mut} samples (red underline) and 7 *DNMT3A*^{R882}/*IDH* doubly mutant samples (blue underline) assayed with WGBS. B. Methylation values across *IDH*^{mut}-specific DMRs in a set of 20 *IDH*^{mut} samples (red underline) and 6 *DNMT3A*^{R882}/*IDH* double mutant samples (blue underline) assayed with methylation array. C. Methylation value across *DNMT3A*^{R882} DMRs in a set of 6 *DNMT3A*^{R882} samples (red underline) and 7 *DNMT3A*^{R882}/*IDH* doubly mutant samples (blue underline) assayed with

WGBS. D. Methylation values across *DNMT3A*^{R882} DMRs in a set of 18 *DNMT3A*^{R882} samples (red underline) and 6 *DNMT3A*^{R882}*/IDH* double mutant samples (blue underline) assayed with methylation arrays. E. Hierarchical clustering of CpG methylation values contained within 2183 *IDH*^{mut}-specific DMRs in primary AML samples with *IDH1* (n=7), *IDH2* (n=13), *TET2* (n=15), and *DNMT3A*^{R882} (n=6), and co-occurring *DNMT3A*^{R882}*/IDH* (n=6), and also *MLL-ELL* (n=11), *CBFB-MYH11* (n=12), and *RUNX1-RUNX1T1* (n=7) fusions. F. Hierarchical clustering of CpG methylation values contained within 3852 *DNMT3A*^{R882} DMRs in primary AML samples with *IDH1* (n=7), *IDH2* (n=13), *TET2* (n=15), *DNMT3A*^{R882} (n=6), and co-occurring *DNMT3A*^{R882}*/IDH* mutations (n=7), and also *MLL-ELL* (n=11), *CBFB-MYH11* (n=12), and *RUNX1-RUNX1T1* (n=7) fusions.

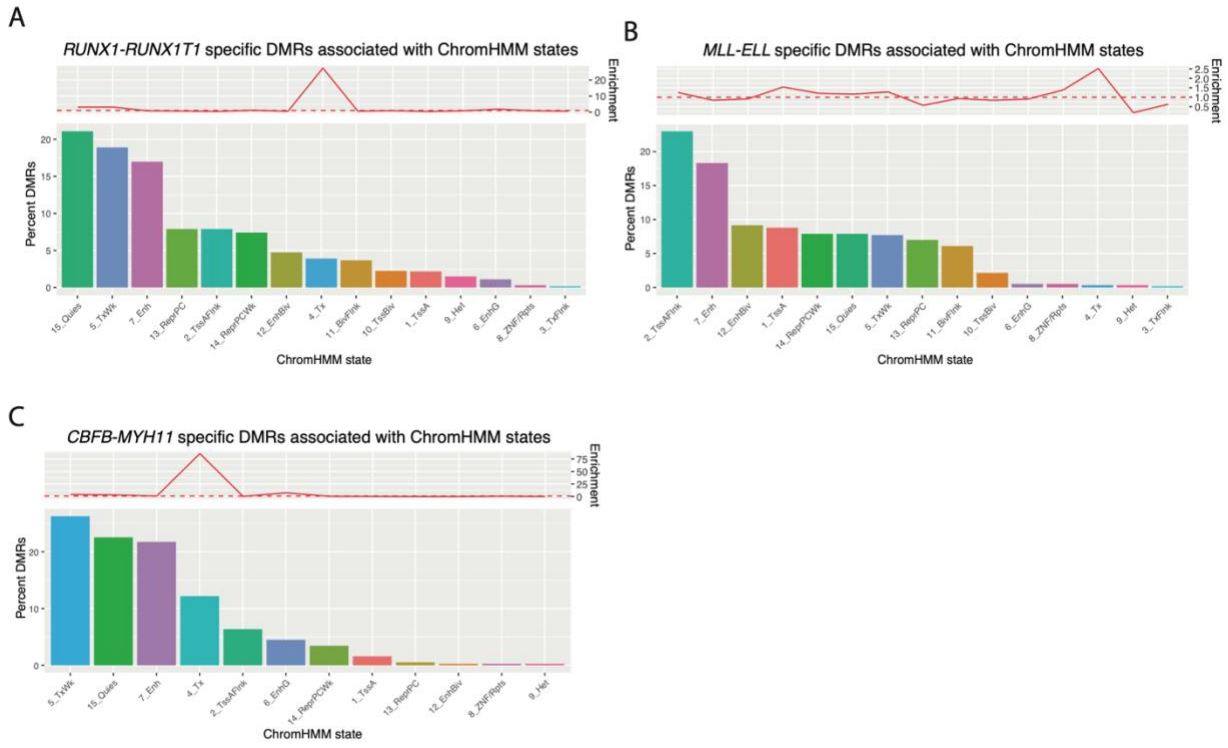


Figure S5. ChromHMM states are unique in subtype-specific DMRs for AMLs with canonical fusions and *DNMT3A*^{R882} mutations compared to *IDH*^{mut} AML. A. Percent overlap of 1921 RUNX1-RUNX1T1 DMRs with 15 ChromHMM chromatin states. B. Percent overlap of 276 *MLL-ELL* DMRs with 15 ChromHMM chromatin states. C. Percent overlap of 309 *CBFB-MYH11* DMRs with 15 ChromHMM chromatin states.

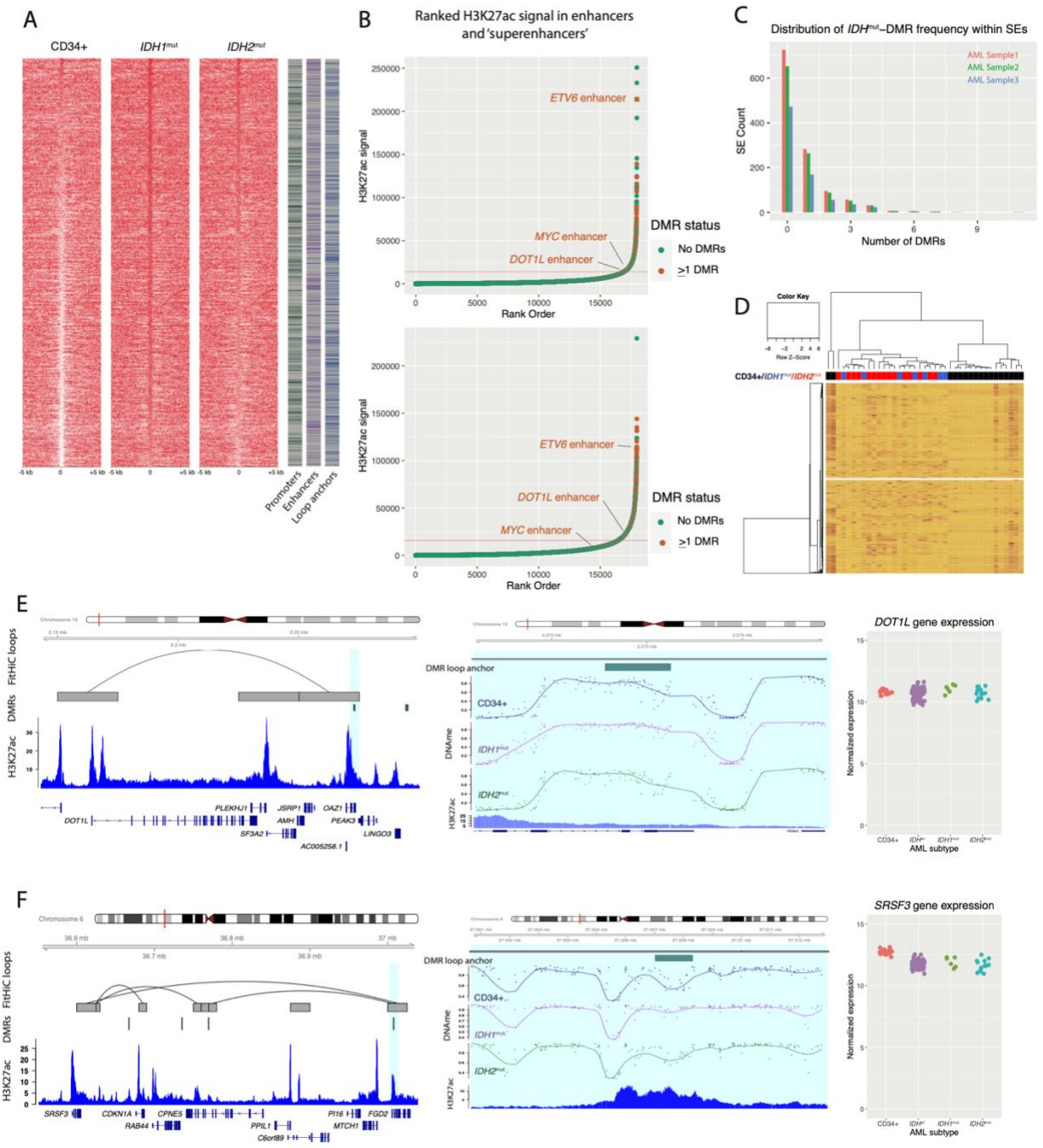


Figure S6. *IDH*^{mut}-specific enhancer DMRs are enriched in 'superenhancers' and contact highly expressed genes in AML. A. Locus heatmap of mean subtype methylation across *IDH*^{mut}-specific DMRs, including annotated overlaps with gene promoters (green), putative active enhancers (purple), and FitHiC loop anchors (blue). B. Representative rank-ordered analysis of H3K27ac marked enhancers in two *IDH*^{wt} AML samples annotated by enhancer and super-enhancer overlap with *IDH*^{mut}-specific DMRs. C. Distribution of number of *IDH*^{mut}-specific DMRs overlapping computationally defined 'superenhancers' in 3 *IDH*^{mut} AML samples. D. Hierarchical clustering of *IDH*^{mut} -eDMR target gene expression in *IDH1* (n=6), *IDH2* (n=14), and normal

CD34+ cord blood cells (N=17, GSE48846). E. Example *IDH*^{mut} -eDMR locus displaying robust interactions with the *DOT1L* promoter. A zoomed in view of the locus demonstrates focal enhancer hypermethylation in *IDH1*^{mut} (purple) and *IDH2*^{mut} (green) samples compared with CD34+ cells (blue). Normalized *DOT1L* expression is shown for 17 CD34+ samples, 6 and 14 *IDH1*^{mut} and *IDH2*^{mut} samples, and 91 *IDH*^{wt} samples. F. Example of *IDH*^{mut} -eDMR locus displaying robust interactions with the *SRSF3* promoter. A zoomed in view of the locus demonstrates focal enhancer hypermethylation in *IDH1*^{mut} (purple) and *IDH2*^{mut} (green) samples compared with CD34+ cells (blue). Normalized *SRSF3* expression is shown for 17 CD34+ samples, 6 and 14 *IDH1*^{mut} and *IDH2*^{mut} samples, and 91 *IDH*^{wt} samples.