UNIVERSITY OF CALIFORNIA SAN DIEGO

Role of DNA Methylation and *TET2* Loss in Myelodysplastic Syndrome Pathobiology and Azacitidine Response

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Biomedical Sciences

by

Brian Matthew Reilly

Committee in charge:

Professor Rafael Bejar, Chair
Professor Frank Furnari
Professor Renate Pilz
Professor Gene Yeo
Professor Kun Zhang

2019

The Dissertation of Brian Matthew Reilly is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2019

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

5-Aza	Azacitidine or 5-azacitidine or 5-azacytidine
5caC	5-carboxycytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
AML	Acute Myeloid Leukemia
ASXL1	*ASXL transcriptional regulator 1*
BER	Base excision repair
BSPP	Bisulfite Padlock Probes
CMP	Common myeloid progenitor
CMML	Chronic myelomonocytic leukemia
CLP	Common Lymphoid progenitor
CRISPR	Clustered Regularly Interspersed Palindromic Repeats
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic Acid
DNAm	Deoxyribonucleic Acid Methylation
EGFP	Enhanced Green Fluorescent Protein
EZH2	*Enhancer of zeste 2 polycomb repressive complex 2 subunit*
FACS	Fluorescence-Activated Cell Sorting
FBS	Fetal Bovine Serum
GEO	Gene Expression Omnibus
GFP	Green Fluorescent Protein
GMP	Granulocyte-monocyte progenitor
HSC	Hematopoietic Stem Cell
HSPC	Hematopoietic Stem and Progenitor Cells (marked by CD34+)
IPSS-R	Revised International Prognostic Scoring Scheme
LAML	Acute Myeloid Leukemia study cohort of The Cancer Genome Atlas
LMPP	Lymphoid-primed multipotent progenitor
M	Molar

MDS	Myelodysplastic Syndromes
mL	Milliliter
mM	Millimolar
MEP	Megakaryocyte-erythrocyte progenitor
MPP	Multipotent progenitor
mRNA	Messenger Ribonucleic Acid
NCBI	National Center for Biotechnology Information
nM	Nanomolar
NMF	Non-negative Matrix Factorization
PCR	Polymerase Chain Reaction
RNA	Ribonucleic Acid
RNA-Seq	Ribonucleic Acid purification followed by high throughput Sequencing
RPMI-1640	Roswell Park Memorial Institute cell culture medium formulation
RUNX1	*Runt related transcription factor 1*
SF3B1	*Splicing factor 3b subunit 1*
*sp*Cas9	*Streptococcus pyogenes* CRISPR Associated Protein 9
SRSF2	*Serine and arginine rich splicing factor 2*
TCGA	The Cancer Genome Atlas
TDG	*Thymine-DNA glycosylase*
TET2	*Tet methylcytosine dioxygenase 2*
TP53	*Tumor protein p53*
TPM	Transcripts Per Million
U2AF1	*U2 small nuclear RNA auxiliary factor 1*
uM	Micromolar

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I must first and foremost thank my wife, Carole, our two furry children, Gili and Luna, and my family for being my emotional support throughout the sometimes trying experience of pursuing a PhD. Carole, thank you for believing in me and putting up with my strange laboratory hours for so many years.

I would also like to thank my mentor Dr. Rafael Bejar for his support and for the opportunity to train as PhD candidate in his lab. Rafael, I feel lucky to have had the good fortune to work under an adviser of such intelligence, kindness and integrity. You have been an inspiration to me for how to balance all of the important things in life with the high stakes job of running a lab, and while maintaining an amazing degree of intellectual curiosity throughout your career.

I am also immensely grateful to the knowledgeable members of my thesis committee: Frank Furnari, Renate Pilz, Gene Yeo, and Kun Zhang. Their guidance immeasurably improved the scientific merits of the work described in this thesis, and I appreciated their kindness and candor throughout my doctoral training experience.

I would also like to acknowledge all of the members of the Bejar Lab, without whom my research would have been much less enjoyable, and doubtless less fruitful. I'd especially like to thank Dr. Tiffany Tanaka for her humor and patience in the long road we had to publish the work in Chapter 2. I'd also like to thank Tim Luger and Dr. Soo Park for their help in completing the work in Chapter 3, but more importantly for the good conversations and camaraderie we shared in

the lab and abroad. Finally, I can't forget to thank Drs. Albert Perez and Sigrid Katz for being great mentors (and great people) in the early years of my training, and introducing me to the ways of the lab.

Chapter 2, in part, has been submitted for publication of the material as it may appear in *Blood Advances*, 2019, Reilly, Brian; Tanaka, Tiffany N.; Diep, Dinh; Yeerna, Huwate; Tamayo, Pablo; Zhang, Kun; Bejar, Rafael. "DNA Methylation Identifies Genetically and Prognostically Distinct Subtypes of Myelodysplastic Syndromes". The dissertation author was one of two primary investigators and authors of this paper.

Chapter 3, in part is currently being prepared for submission for publication of the material. Reilly, Brian; Luger, Timothy; Park, Soo; Lio, Chan-Wang; González-Avalos, Edahí; Wheeler, Emily C.; Lee, Minjung; Diep, Dinh; Zhang, Kun, Huang, Yun; Rao, Anjana; Bejar, Rafael. "*TET2* Dependent Effects of 5-Azacitidine on DNA Methylation and Gene Expression Dynamics". The dissertation author was the primary investigator and author of this material.

VITA

2011	Bachelor of Science, California Polytechnic State University, San Luis Obispo
2012-2014	Student, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego
2014-2019	Research Assistant, University of California San Diego
2019	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

(1) Ali, Thibault B., et al. “Adverse Effects of Cholinesterase Inhibitors in Dementia, According to the Pharmacovigilance Databases of the United-States and Canada.” *PLoS One*, vol. 10, no. 12, 7 Dec. 2015, doi:10.1371/journal.pone.0144337.

(2) Reilly, Brian M., et al. “DNA Methylation Identifies Genetically and Prognostically Distinct Subtypes of Myelodysplastic Syndromes.” *Blood Advances*, 2019. *Under review.*

(3) Reilly BM, Luger T, Park S, Chan-Wang L, González-Avalos E, Wheeler EC, Lee M, Diep D, Zhang K, Huang Y, Rao A, Bejar R. “TET2 Dependent Effects of 5-Azacitidine on DNA methylation and Gene Expression.” *In preparation.*

FIELDS OF STUDY

Major Field: Biomedical Sciences

Studies in Genetics and Cancer Biology
Professor Rafael Bejar

ABSTRACT OF THE DISSERTATION

Role of DNA Methylation and *TET2* Loss in Myelodysplastic Syndrome Pathobiology and Azacitidine Response

by

Brian Matthew Reilly

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2019

Professor Rafael Bejar, Chair

Myelodysplastic syndromes (MDS) are hematologic malignancies characterized by impaired differentiation of hematopoietic stem and progenitor cells (HSPC) causing peripheral blood cytopenias and an increased propensity for developing acute myeloid leukemia. MDS arises as a consequence of genetic and/or epigenetic lesions in HSPCs. Dysfunctional epigenetic regulation is a hallmark of MDS, where a majority of MDS patients harbor mutations in genes regulating

DNA methylation (5mC) or chromatin modification. Aberrant 5mC patterns are common in MDS and may correlate with disease progression.

Disease progression in MDS is highly variable, making prognostication critical for decisions on treatment options and intensity. In Chapter 2 we interrogated 5mC patterns at ~500,000 CpGs across the genome in 141 MDS patient tumor samples and applied an unsupervised classification to define 5mC sub-types of MDS. We identified five patient clusters with distinct 5mC patterns, and these groups were enriched for distinct patterns of genetic lesions related to prognostic risk. The 5mC sub-types displayed differences in survival that were independent of all known prognostic variables, including genetics. Differentially methylated genes between clusters included those with known prognostic impact, as well as novel gene associations which were validated in external MDS cohorts. Our findings highlight the importance of 5mC in disease progression and its utility as a prognostic biomarker.

DNA methyltransferase inhibitors, such as 5-Azacitidine (5-Aza), are the only therapy approved for treating higher-risk MDS, yet only half of patients respond to therapy. Mutations in the 5mC regulator, *TET2*, are associated with increased response rates, however the mechanisms involved are unknown. In Chapter 3 we modeled *TET2* loss in isogenic erythroleukemia cell lines and studied differences in 5mC, 5-hydroxymethylcytosine (5hmC), and gene expression during 5-Aza exposure. We show that *TET2* loss particularly influences 5mC and 5hmC patterns at erythroid gene enhancers, and is associated with

down-regulation of gene expression. 5-Aza disproportionately induces expression of erythroid genes in TET2KO cells through 5mC reduction at erythroid enhancers. This work highlights the role of 5mC and 5hmC changes at enhancers in altering differentiation-associated gene expression signatures, and sheds light on how 5-Aza may be more effective in patients harboring *TET2* mutations.

Chapter 1: Introduction

1.1 Normal Hematopoiesis

Hematopoiesis is the process whereby a hematopoietic stem cell (HSC) in the bone marrow differentiates into the wide array of functional cell types which make up the blood and immune system. Starting with the early histopathologists studying leukemia, there there has existed an idea that there was a “stem cell” which gave rise to both the red and white cells of the blood (Pappenheim 1896; *Folia haematol (Leipz)* 1907). In the many years since this idea was suggested, many studies lead to the synthesis of a hierarchical model of hematopoiesis where HSCs differentiate step-wise into further and further differentiated precursors with varying self-renewal potential before finally differentiating into the terminally differentiated functional cells of the blood (Akashi et al. 2000; Kondo, Weissman, and Akashi 1997). As shown in Figure 1.1 this process involves HSCs (short and long-term HSCs which can self-renew) differentiating into multipotent progenitors (MPP) which retain multilineage differentiation potential but lose self-renewal capacity, followed by differentiation into the lineage restricted progenitors: the common myeloid progenitors (CMP) which go on to form all of the myeloid lineage cells, and the common lymphoid progenitors (CLP) which form the lymphoid lineage cells of the blood. This model originally synthesized from experiments based on cell surface markers and assays of clonogenic and differentiation potential, has been refined over the years as newer technologies have allowed greater specificity. First, the discovery of a lymphoid precursor cell which retained some capacity to generate cells of the granulocytic/monocytic lineage, prompting

the coining of a new class of multipotent progenitor, the lymphoid-primed multipotent progenitor (LMPP) (Adolfsson et al. 2005). Single-cell transplantation studies later demonstrated that HSCs (cells which retain self-renewing capacity) can be biased in terms of their ability to recapitulate all of the cells of the bone marrow (Yamamoto et al. 2013), the first indication that lineage commitment may bypass the MPP and CMP/CLP steps of the hierarchical model of differentiation. Finally, cutting edge studies using high-throughput single-cell gene expression coupled with cell-marker analysis have indicated that differentiation likely exists on a continuum in single-cells (Velten et al. 2017). In this model, individual HSCs can acquire lineage biases without passing through discrete hierarchically organized progenitor populations historically associated with that lineage (e.g. MPP → CMP → MEP → erythrocyte). Furthermore, HSCs can be primed for multiple different lineages at once, and depending on external stimuli and stochastic processes, become further committed along a particular lineage to produce differentiated cells. This model explains many of the earlier findings in which individual HSCs and further differentiated progenitors didn't behave according to the hierarchical model, and serves as the most modern understanding of the process of hematopoiesis. The hierarchical model remains a fairly accurate simplified model of hematopoiesis with the caveat that there is not likely to be true isolated committed progenitor populations (CMP, CLP, LMPP), but rather heterogeneous collections of lineage biased HSCs which have some probability for differentiation

along multiple lineages that narrow as differentiation progresses (Laurenti and Göttgens 2018).

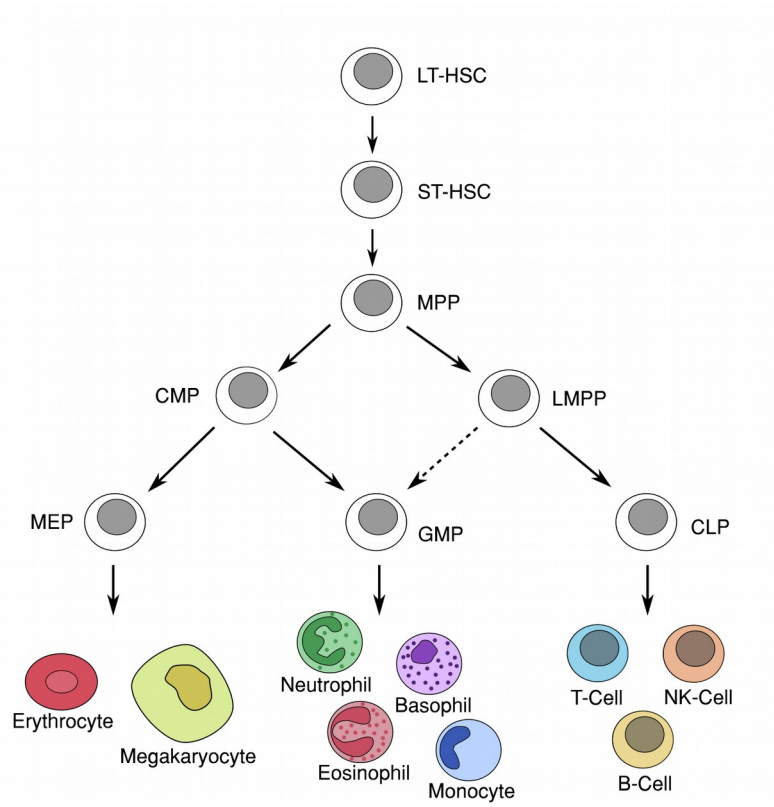


Figure 1.1 The hierarchical model of hematopoiesis. This simplified diagram shows the pathway for hematopoietic stem cell (HSC) to differentiate to produce all of the mature cells of the blood and immune system. LT-HSCs can sustain hematopoiesis for an individual's lifetime, whereas ST-HSCs are short-lived but can self-renew and can produce all mature cell types. More differentiated progenitors (MPP, CMP, LMPP, etc.) are incapable of self-renewal. LT-, ST-HSC; long-term- and short-term hematopoietic stem-cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; LMPP, lymphoid-primed multipotent progenitor; MEP, megakaryocyte-erythroid progenitor; GMP, granulocyte-monocyte progenitor; CLP, common lymphoid progenitor.

1.2 DNA Methylation

DNA methylation (DNAm) is one of the main forms of epigenetic regulation, which in the most general sense is the mechanism by which an individual's genome which is static across all cells, is able to produce the wide diversity of different cell type phenotypes that make up an organism (Waddington 1942). DNA methylation typically refers to an additional methyl group on the fifth position of

cytosine nucleotides in DNA, known as 5-methylcytosine (5mC). 5mC is a stable, heritable, epigenetic mark which is critically important in the maintenance of cellular identity and differentiation of stem cells (Smith and Meissner 2013). 5mC typically occurs in the context of CpG di-nucleotides which are symmetrical on either strand of DNA, and due to this symmetry, 5mC marks can be maintained after DNA replication, making them heritable (Ramsahoye et al. 2000; Ziller et al. 2011). Non-CpG methylation, referring to 5mC occurring outside the context of a CpG dinucleotide, is a rare form of 5mC which is less well understood (Ziller et al. 2011).

5mC has a complex role in the regulation of gene expression, where 5mC may serve different purposes in different contexts. Of the ~28 million CpGs in the human genome, the majority (60-80%) are constitutively methylated, while the less than 10% of CpGs occurring in CpG dense regions known as CpG islands (CGI's) and are typically resistant methylation (Saxonov, Berg, and Brutlag 2006; Smith and Meissner 2013). Thus, the majority of CpGs in the genome are static across tissues and throughout life, only changing in localized contexts as different cellular processes must be activated or repressed for cellular differentiation or in response to external factors (Smith and Meissner 2013). While the majority of CpGs are static, the non-static CpGs play a critical role in gene regulation. Generally when 5mC is located at non-CGI gene promoters it is associated with repression of transcription, while when located within gene bodies, it is associated with activation of transcription (Schultz et al. 2015a; Yang et al. 2014; Roadmap

Epigenomics Consortium et al. 2015). 5mC can also regulate gene expression by altering the binding of transcription factors or transcriptional insulator proteins (Hark et al. 2000; P.-Y. Chen et al. 2011; Prendergast and Ziff 1991). Transcriptional enhancers are regions of DNA/chromatin which are located distal to a gene transcription start site (TSS) which, through DNA/chromatin looping are able to enhance gene expression from a distance (Shlyueva, Stampfel, and Stark 2014). The role of 5mC at enhancers is less well established, however several studies have reported that hypomethylated enhancers induce higher transcriptional activity, indicating that maintenance of low methylation at enhancers is important for at least some genes (Lister et al. 2009; Schmidl et al. 2009; Andersson et al. 2014).

DNA methylation is a major determinant of cellular identity and cell fate during differentiation. This fact is especially clear in hematopoietic differentiation, where 5mC protects HSCs from premature differentiation (Álvarez-Errico et al. 2015; Bröske et al. 2009). 5mC is also a determining factor in commitment of HSCs to a lymphoid vs. myeloid cell fate and differentiated cell identity (Álvarez-Errico et al. 2015; Trowbridge et al. 2009). Notably differentiation-associated 5mC differences in both hematopoietic as well as other tissues seems to be particularly enriched at cis-acting transcriptional enhancers as opposed to gene promoters or gene bodies (Farlik et al. 2016; Sheaffer et al. 2014a).

1.3 Regulation of 5mC

The methyl group of 5mC is added to cytosines by several members of the DNA methyltransferase (DNMT) gene family (*DNMT1*, *DNMT3A*, *DNMT3B*), which each have unique roles in regulation of 5mC. *DNMT1* is known as the "maintenance" methyltransferase, as it recognizes and binds to hemi-methylated CpGs (CpGs which are methylated on only one strand of DNA) with higher affinity, and catalyzes the addition of a methyl group to the cytosine of the unmethylated DNA strand (Hermann, Goyal, and Jeltsch 2004). In this way, *DNMT1* is responsible for maintenance of epigenetic identity during cellular division, for without *DNMT1* maintenance of 5mC, it would become successively diluted with increasing cellular divisions. *DNMT3A* and *DNMT3B* are known as *de novo* methyltransferases, as they can catalyze the addition of new methyl groups to cytosine in the absence of hemi-methylated DNA (Okano et al. 1999). *De novo* DNMTs are therefore especially important during cellular differentiation which requires new methylation marks for the establishment of new cell lineages from stem cells such as HSCs. In fact, *DNMT1*, *3A*, and *3B* are all indispensible for normal mammalian development (Okano et al. 1999; Li, Bestor, and Jaenisch 1992).

DNA methylation can be removed by both passive and active mechanisms. Passive loss of 5mC occurs when *DNMT1* levels are insufficient to maintain all hemi-methylated CpGs during successive rounds of DNA replication. Mechanisms for the active removal of 5mC have been postulated to exist for many years

(Mayer et al. 2000) and have only recently been elucidated with the discovery of the function of the tet methylcytosine dioxygenase enzymes (TET 1, 2, and 3). TET proteins can catalyze the oxidation of 5mC to 5-hydroxymethylcytosine (5hmC) (Tahiliani et al. 2009a; Ito et al. 2010a) and to further oxidized derivatives, 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) (Ito et al. 2011; He et al. 2011a). 5caC and 5fC are both recognized and excised by thymine-DNA glycosylase (TDG) creating abasic sites which are restored with unmodified cytosines through base excision repair (BER) (Maiti and Drohat 2011a; Weber et al. 2016), thus completing the active demethylation pathway. Recent work indicates that TET proteins and 5hmC likely have functions outside the demethylation pathway, which will be enumerated in a section below.

1.4 Myelodysplastic Syndromes

Myelodysplastic syndromes (MDS) in the simplest terms are a group of disorders in which hematopoietic stem and progenitor cells (HSPC) are unable to properly differentiate to form functional mature hematopoietic cells, often accompanied by an accumulation of immature and dysfunctional HSPCs in the bone marrow (Figure 1.2). The disorder is associated with clonal hematopoiesis and an increased risk of transformation to acute myeloid leukemia (AML) (Tefferi and Vardiman 2009; Adès, Itzykson, and Fenaux 2014). Patients with MDS typically exhibit a hypercellular bone marrow with dysplasias (abnormal cell morphology) and an excess of immature cells (blasts), as well as peripheral blood cytopenias in one or more hematopoietic lineages (Vardiman et al. 2009). MDS is a disease of

aging, with a median age at diagnosis of 65-70 years and increasing incidence with age (Adès, Itzykson, and Fenaux 2014).

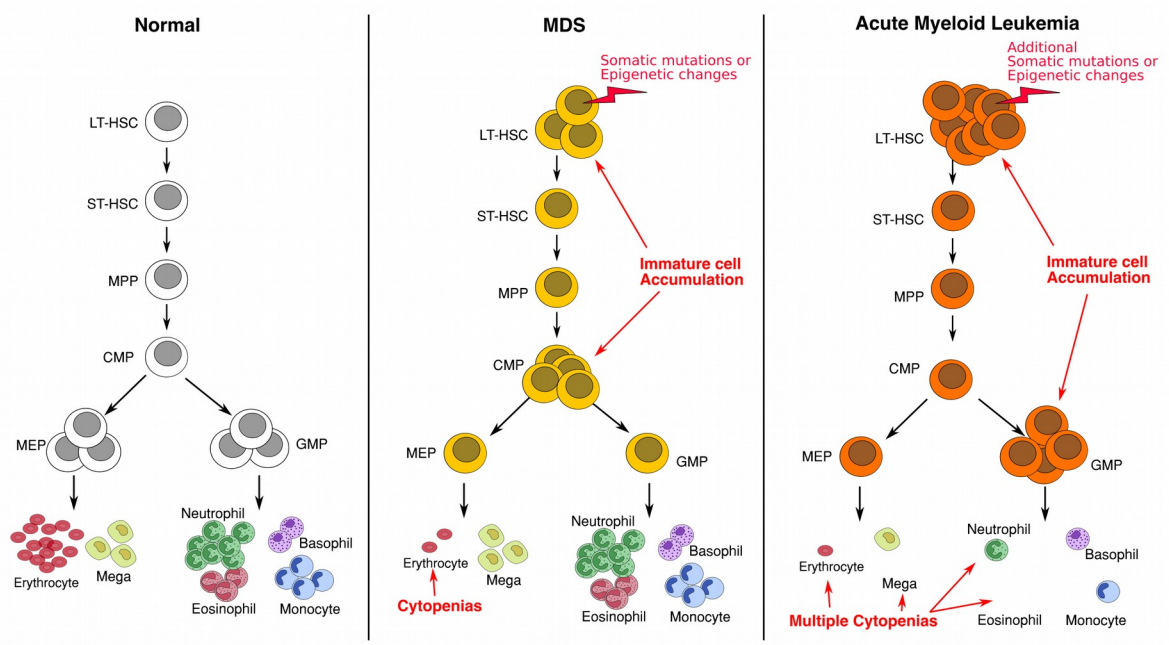


Figure 1.2 Normal and malignant myelopoiesis in MDS during disease progression. Normal myelopoiesis produces sufficient quantity of functional mature blood lineages, while in MDS, genetic and epigenetic lesions lead to an accumulation of immature progenitor (HSC and CMP) cells and uni- or multi- lineage cytopenias in the peripheral blood that worsen as the disease progresses towards high-risk disease or secondary AML. Adapted from Shastri A., Will B., Steidl U., Verma A. *Blood.* 2017.

1.5 Molecular Basis of MDS

The majority MDS cases occur without a known cause, while a minority of patients may acquire MDS secondary to exposure to chemotherapeutic agents or environmental exposures, and there is some evidence for an autoimmune component to the disorder (Adès, Itzykson, and Fenaux 2014). The molecular basis for MDS is complex, but there is now strong evidence that cytogenetic and genetic lesions are at the heart of the disorder (Ogawa 2019). While roughly half of MDS patients exhibit clear cytogenetic alterations, virtually all patients harbor somatic

mutations, and the number of somatic mutations increases with disease progression (Low-risk < high-risk < secondary-AML) (Ogawa 2019). A key finding from genetic studies of MDS was the high frequency of somatic mutations in genes involved in epigenetic regulation (*DNMT3A*, *TET2*, *ASXL1*, *EZH2*) (Bejar et al. 2011; Haferlach et al. 2014). This finding helped explain other studies which found MDS and AML patients had a deregulated epigenome, and also provides a plausible explanation for the efficacy of DNA methyltransferase inhibitors in treating the disorder. Notably, recent work has shown that some healthy individuals exhibit clonal hematopoiesis with mutations in epigenetic regulators, and these patients are at an increased risk of developing myeloid malignancies (Jaiswal et al. 2014; Kwok et al. 2015). This finding suggests that epigenetic regulator mutations are likely an initiating event on the path towards myeloid malignancy that may take many years to present (Sperling, Gibson, and Ebert 2017).

1.6 Prognostication in MDS

The clinical course of MDS is highly variable. While some patients may survive on supportive care (blood transfusions and hematopoietic growth factors) alone for the remainder of their lives, other patients may rapidly progress and die within one year (Greenberg et al. 2012). For this reason, identifying subsets of patients who may require more aggressive treatments is paramount (Platzbecker 2019). The most current guidelines for prognostication in MDS are laid out in the Revised International Prognostic Scoring System for myelodysplastic syndromes

(IPSS-R) (Greenberg et al. 2012), which scores patients on a variety of prognostic factors including degree of cytopenias, percentage of bone marrow blasts, as well as the presence of cytogenetic abnormalities present in bone marrow disease cells. Adverse cytogenetic risk abnormalities include deletions of chromosome 7 and chromosome 3 abnormalities, as well as complex karyotype. Certain cytogenetic changes such as loss of the Y chromosome or long (q) arm of 11 confer no added risk, however in general the greater the number of cytogenetic abnormalities the higher the disease risk.

While the IPSS-R remains the gold-standard for prognostication in MDS, in the years since its publishing a number of groups have studied the value of somatic genetics as a means of refining the risk groups laid out in the IPSS-R, and these are increasingly being incorporated into clinical decision making (Platzbecker 2019; Bejar et al. 2011; Haferlach et al. 2014). Bejar et al. were the first to demonstrate that the presence or absence of specific somatic mutations as well as the number of mutations were predictive of survival even among patients with identical IPSS risk scores (Bejar et al. 2011). Later studies have refined these observations as well as identified new relationships between somatic genetics and survival after HSC transplantation (Haferlach et al. 2014; Bejar, Stevenson, et al. 2014).

Given the high frequency of epigenetic regulator mutations in MDS, alterations in 5mC have also been studied for their value as prognostic biomarkers. Jiang et al. Were the first to show that the degree of aberrant

hypermethylation at the *FZD9* locus, among others, was predictive of IPSS risk scores, and correlated with overall survival in a cohort of 184 patients (Ying Jiang et al. 2009). Other studies using higher-throughput methods used unsupervised classification approaches to identify 5mC subtypes in AML, and discovered that while several of these subtypes corresponded to known genetic or cytogenetic risk groups, multiple subtypes were novel and independently associated with survival (Figueroa, Lugthart, et al. 2010a). No such study has yet been performed in MDS, until now with the work we lay out in Chapter 2.

1.7 Treatment of MDS

The therapeutic strategy in MDS is largely guided by prognostic risk (Platzbecker 2019). Many low risk patients will die of causes other than MDS, so often these patients will receive supportive care alone. With higher risk patients the goal is always to modify the course of the disease, avoid progression to acute myeloid leukemia and to prolong survival (Adès, Itzykson, and Fenaux 2014; Platzbecker 2019). The only curative treatment for MDS is hematopoietic stem-cell transplantation, however transplantation is not available for the majority MDS patients due to the advanced age and likely comorbidities present in the patient population (Witte et al. 2017; Platzbecker 2019). For many low-risk patients, erythropoiesis-stimulating agents are effective, however resistance can build over time at which point these patients will require more intensive treatments as for higher-risk patients. Patients with isolated deletion within the q arm of chromosome 5 benefit from lenalidomide treatment which achieves transfusion

independence in two-thirds of patients, however resistance can also arise with this drug (Giagounidis et al. 2012).

In most higher-risk patients who are not eligible for HSC transplant, DNA methyltransferase inhibitors (DNMTIs) such as azacitidine and decitabine are the first-line therapy choice as they confer a survival advantage, delay the time to transformation to AML, and help some patients achieve red blood cell transfusion independence (Fenaux et al. 2009; Adès, Itzykson, and Fenaux 2014). Responses to DNMTIs are delayed in many patients, with a median time to response of 2-3 months, while some patients first response may be as long as 6 months after the start of treatment (Silverman et al. 2011). DNMTIs remain the most effective treatment for higher-risk patients, however less than 50% of patients respond to therapy, and there are no reliable predictors of response (Fenaux and Ades 2009; Bejar, Lord, et al. 2014b). Due to the delayed response times, many nonresponding patients lose precious time during treatment while their disease may progress and while they could otherwise be pursuing alternative or experimental therapies. For the above reasons, biomarkers of response to DNMTIs are sorely needed.

1.8 DNA Methyltransferase Inhibitors

DNMTIs are cytosine nucleotide analogs initially developed as traditional cytotoxic chemotherapeutics in the 1960s (Sorm et al. 1964). It wasn't until 1980 that they were discovered to induce cellular differentiation *in vitro* as well as DNA hypomethylation (Jones and Taylor 1980). The mechanism by which DNMTIs

inhibit DNA methylation is now known to be through the irreversible inhibition of DNA methyltransferase enzymes (*DNMT1*, *DNMT3A*, and *DNMT3B*). DNMTIs must be incorporated into DNA during DNA replication in order to act as a substrate for DNMTs leading to irreversible covalent binding and subsequent proteasomal degradation (Santi, Norment, and Garrett 1984). In the absence of sufficient DNMT enzymes, passive DNA demethylation occurs during subsequent cell divisions (Stresemann and Lyko 2008). DNMTIs require active cell cycling to be incorporated into DNA, yet they also cause DNA damage themselves by forming a covalent bond between enzymes and DNA, therefore dosage optimization was critical in finding a balance between inhibiting DNA methylation and stopping cell cycling. For this reason, current dosage schedules typically call for 5 to 7 sequential days of low-dose DNMTI infusion in MDS patients (Jabbour et al. 2017).

DNMTIs were first applied in MDS due to their known role in inducing cellular differentiation, as well as the knowledge that MDS patients exhibit aberrant hypermethylation of tumor suppressor gene promoters(Mund, Brueckner, and Lyko 2006). In theory DNMTIs were thought to exert their efficacy through demethylation of aberrantly methylated tumor suppressor genes, and studies have shown that this does indeed occur during treatment (Stresemann and Lyko 2008), however there may be other mechanisms which contribute to response to these agents.

As mentioned above, the main challenge with DNMTIs is predicting which patients will respond to therapy. A number of studies have examined dynamic DNA methylation changes at different time-points during treatment in the hopes of identifying a signal associated with treatment response at an earlier time-point (Welch et al. 2016a; Merlevede et al. 2016; Meldi et al. 2015; Shen et al. 2010a). One question at the forefront of the quandary was whether nonresponding patients were resistant to therapy due to accelerated drug metabolism and/or incomplete inhibition of DNA methylation. From studies by Welch et al. we now know that both responding and nonresponding patients display similar levels of demethylation at an acute time point after treatment (24hrs after final dose), so response is not dictated by the amount of demethylation after treatment. We also now know that the amount of demethylation at the promoters of tumor suppressor genes (Shen et al. 2010a), as well as genome wide (Merlevede et al. 2016), after 4 to 6 monthly cycles of treatment does indeed correlate with treatment response, with nonresponding patients exhibiting no change or gains in 5mC, and responding patients displaying profound hypomethylation. These studies together suggest that the *rate* of remethylation may be an important governing factor in determining treatment response.

Finally, a number of studies have examined genetic predictors of response. Varying degrees of association have been observed between mutations in several genes and response, however the only mutation to be validated as a predictor of response in multiple studies were those in *TET2* (Bejar, Lord, et al. 2014b; Traina

et al. 2014; Itzykson et al. 2011). Unfortunately, mutation in *TET2* is an imperfect predictor, as it is one of the most frequently mutated genes in MDS and is often comutated with a variety of genes which likely affect its ability to confer sensitivity to DNMTIs. For example Bejar et al observed that TET2 mutant patients who did not harbor a mutation in ASXL1 were more likely to respond than those who did. A better understanding of the processes which confer sensitivity to DNMTIs in *TET2* mutant MDS is needed.

1.9 TET2 in MDS

TET2 is one of the most frequently mutated genes in MDS and other myeloid malignancies, and has been extensively studied in this context in recent years (Ogawa 2019; Bejar 2017). The vast majority of TET2 mutations in MDS and other hematologic malignancies are inactivating mutations which inhibit or abolish its catalytic activity leading to globally reduced 5hmC (Ko et al. 2010). down-regulation of TET proteins also appears common in hematologic malignancy, as tumor cells from MDS and AML patients wild-type for TET proteins may also exhibit global reduction in 5hmC content (Ko et al. 2010). *TET2* inactivation in myeloid malignancies and animal models was shown to bias hematopoiesis in favor of myeloid differentiation at the expense of other lineages, and also enhanced HSC expansion and repopulating capacity (Delhommeau et al. 2009; Ko et al. 2011). In the absence of other cooperating mutations, *TET2* inactivation rarely induces frank leukemias or MDS in animal models, which supports its role

as an early or initiating lesion in the disorder (Kasper Dindler Rasmussen and Helin 2016).

Previously, we highlighted the role of 5mC in differentiation, however 5hmC and active demethylation appear equally important in this process. TET2 loss appears to induce differentiation defects in a variety of stem cells including and especially in hematopoietic tissues (Orlanski et al. 2016; Hon et al. 2014; Moran-Crusio et al. 2011). The differentiation defects have been linked to aberrant hypermethylation of lineage-specific enhancer and other regulatory elements, which provides a plausible connection to the efficacy of DNMTI-induced hypomethylation in the TET2-mutant context (Yamazaki et al. 2015; Kasper D Rasmussen et al. 2015; Hon et al. 2014).

The specific role of *TET2* inactivation in the context of DNMTI treatment response is relatively unknown. In competetive bone marrow transplantation experiments in mice, Tet2-null cells displayed significantly decreased peripheral blood chimerism after DNMTI treatment, while Tet2-wild-type cells showed no significant change (Bejar, Lord, et al. 2014b). In two studies which examined disease cell clonal dynamics of MDS and CMML patients during DNMTI treatment it was shown that while certain disease clones can be cleared by DNMTIs during treatment response, founding clones, and particularly those carrying mutations in epigenetic regulators like *TET2* did not show any major differences in clone size during treatment response (recovery from peripheral blood cytopenias), suggestive

of an epigenetic mechanism of response (Merlevede et al. 2016; Unnikrishnan et al. 2017a).

1.10 Summary and Rationale:

MDS is highly heterogeneous both in terms of disease presentation as well as disease progression and outcome. Genetic studies in MDS have improved our understanding of the pathophysiology of the disease as well as improved outcome prediction and helped resolve some of the heterogeneity of these disorders. 5mC patterns may integrate cues not only from genetic lesions, but also from the microenvironment, the differentiation states of cells, and age-related changes, making 5mC an attractive biomarker in MDS. Given the above reasons as well as the important role of DNA methylation in MDS pathobiology, **in Chapter 2 we hypothesized** that examining MDS patient DNA methylomes in an unsupervised manner may improve our understanding of MDS pathobiology and further resolve some of the remaining heterogeneity in these disorders.

DNMTIs are the first-line treatment of higher-risk MDS patients, yet less than 50% of patients will respond to therapy, and responses can take as long as 6 months to become apparent, during which time a non-responding patient will suffer adverse affects and increased risks associated with treatment. It is of dire importance that we gain a better understanding of the mechanisms of response to DNMTIs in MDS. Previous work studying the role of 5mC in treatment response has highlighted the importance of rate of 5mC recovery as an important factor governing treatment responses. Previous work in our lab discovered that TET2-

mutant MDS patients were at an increased likelihood of responding to DNMTIs, yet this relationship was complicated by the patterns of comutations in these patients. Hypermethylation of lineage-specific enhancers leading to differentiation alterations in the hematopoietic system have been observed in TET2-mutant animal models and MDS patient samples. Given the role of TET2 in regulating 5mC and 5hmC and the importance of 5mC dynamics in determining DNMTI responses, **in Chapter 3 we hypothesized** that dynamic 5mC, 5hmC, and expression changes during treatment may reveal important pathways governing response, and we study this in carefully modeled isogenic cell lines to isolate the effects driven by *TET2* loss.

Chapter 2: DNA Methylation Identifies Genetically and Prognostically Distinct Subtypes of MDS

2.1 Abstract

Recurrent mutations implicate several epigenetic regulators in the early molecular pathobiology of myelodysplastic syndromes (MDS). We hypothesized that MDS subtypes defined by DNA methylation (DNAm) patterns could enhance our understanding of MDS disease biology and identify patients with convergent epigenetic profiles. Bisulfite padlock probe sequencing (BSPP) was used to measure DNAm of ~500,000 unique CpGs covering 140,749 non-overlapping regulatory regions across the genome in bone marrow DNA samples from 141 patients with MDS. Application of a non-negative matrix factorization (NMF) based decomposition of DNAm profiles identified five consensus clusters described by five NMF components as the most stable grouping solution. Each of the five NMF components identified by this approach correlated with specific genetic abnormalities and categorized patients into five distinct methylation clusters, each largely defined by a single NMF component. Methylation clusters displayed unique differentially methylated regulatory loci enriched for active and bivalent promoters and enhancers. Two clusters were enriched for samples with complex karyotypes although only one had an increased number of *TP53* mutations. Each of the three most frequently mutated splicing factors, *SF3B1*, *U2AF1*, and *SRSF2*, was enriched in different clusters. Mutations of *ASXL1*, *EZH2* and *RUNX1* were co-enriched in the *SRSF2* containing cluster. In multivariate analysis, methylation cluster membership remained independently associated with overall survival. Targeted DNAm profiles identify clinically relevant subtypes of MDS not otherwise

distinguished by mutations or clinical features. Patients with diverse genetic lesions can converge on common DNAm states with shared pathogenic mechanisms and clinical outcomes.

2.2 Introduction

Myelodysplastic syndromes (MDS) are a group of diverse hematologic malignancies caused by accumulation of somatic driver mutations in clonally expanded hematopoietic stem and progenitor cells (HSPCs) (Bejar et al. 2011; Haferlach et al. 2014; Papaemmanuil et al. 2013).The progeny of these cells demonstrate impaired myeloid differentiation, resulting in peripheral blood cytopenias, progressive bone marrow failure and potential progression to acute myeloid leukemia (AML). Genetic lesions are identified in the vast majority of patients with MDS, however, no single mutation profile describes the typical MDS patient (Woll et al. 2014; Walter et al. 2013). Instead, a wide variety of mutation profiles appear capable of generating morphologic and clinical features typical of MDS (Cazzola, Della Porta, and Malcovati 2013) . Epigenetic marks, such as DNA methylation (DNAm), are critical for mammalian development and are frequently implicated in oncogenesis (Gu et al. 2018; Westers et al. 2012; Klutstein et al. 2016). Somatic mutations in epigenetic regulators are prevalent in more than half of all MDS cases and include genes involved in maintenance of DNAm (e.g., *TET2* and *DNMT3A*) (Walter et al. 2011; Abdel-Wahab et al. 2009) and histone modification (e.g., *ASXL1 and EZH2*) (Gelsi-Boyer et al. 2009; Nikoloski et al. 2010). Epigenetic mutations frequently occur early in MDS pathogenesis (Corces-

Zimmerman et al. 2014; Malcovati et al. 2017) and aberrant DNAm has been associated with progression to AML (Ying Jiang et al. 2009; Figueroa et al. 2009; Spencer et al. 2017). MDS patients *without* mutations in epigenetic regulators can show alterations in their epigenome suggesting a convergent pathogenic mechanism (J. P. J. Issa 2013; Ko et al. 2010). Finally, DNA methyltransferase inhibitors are standard of care treatments for certain MDS subtypes as their effects can improve hematopoiesis and prolong overall survival (OS), with associated changes in DNAm and gene expression in responding patients (Fenaux et al. 2009; Lübbert et al. 2011; Tsai et al. 2012; Merlevede et al. 2016; Meldi et al. 2015; Shen et al. 2010b). As DNAm patterns may be influenced by microenvironmental cues and genetic perturbations, we hypothesized that convergent oncogenic states defined by DNAm may represent a useful tool to understand the different pathobiological mechanisms active in MDS.

Prior studies have attributed altered gene expression to DNA promoter hypomethylation in MDS, (Shen et al. 2010b; Fandy et al. 2009; Grövdal et al. 2014) and more recently, to aberrant methylation at more distal regulatory regions such as enhancers (Meldi et al. 2015; Hasegawa et al. 2017; Kasper D Rasmussen et al. 2015; Yamazaki et al. 2015; Yang et al. 2014). Methylation assays that interrogate genomic regions beyond promoters may better inform which regulatory regions possess biologic importance in the development and progression of MDS. Since the functional relevance of methylation in these regions is not well understood, computational methods that agnostically evaluate DNAm datasets

may better classify MDS patients without bias. To characterize biologically relevant features of the MDS methylome, we utilized a targeted bisulfite padlock probe (BSPP) sequencing method that captures specific non-overlapping regulatory regions, including differentially methylated regions (DMRs) related to genes involved in pluripotency, differentiation and cancer, (Irizarry et al. 2009; Doi et al. 2009; Lister et al. 2009; Figueroa, Lugthart, et al. 2010b) all known promoters for human NCBI Reference Sequence genes as well as all microRNA genes, CTCF binding sites and DNase I hypersensitive regions (Diep et al. 2012; Deng et al. 2009). We employed a computational approach called *Onco*Genic Positioning System (*Onco*-GPS) that uses non-negative matric factorization (NMF) to decompose a high dimensional methylation matrix into components which serve as inputs for consensus clustering, resulting in classification of patients by shared methylation signatures (Kim et al. 2017). One advantage of this approach over standard clustering of raw methylation values is that standard hierarchical clustering may disproportionately weight the influence of CpG dense regions with highly correlated methylation levels such as CpG islands (CGIs), whereas NMF first summarizes highly correlated parts of the data (such as CpGs in CGIs) into components. This gives equal weighting for regions of lower CpG density, such as enhancers, that may have comparable biologic importance.

Here, we report an epigenomic analysis using the BSPP methylation platform and *Onco*-GPS computational approach of 141 genetically characterized MDS patient samples (Bejar et al. 2011; Bejar, Levine, and Ebert 2011). We

identify distinct methylation states enriched for specific patterns of somatic mutations, cytogenetic abnormalities and clinical outcomes including differences in OS. Our findings suggest that DNAm signatures, similar to genetic mutations and gene expression profiles, may refine our ability to classify and predict clinical outcomes in MDS.

2.3 Results

Patient Demographics and Clinical Features

Of the 154 patient samples selected for methylation profiling, 10 were removed due to insufficient coverage and 3 were removed after being identified as extreme outliers, leaving 141 samples in the final cohort. Demographic and clinical information is provided in Table 2.1 and Supplementary Table 2.1. The median age was 72 years with 103 (73%) males. 48 patients (34%) were known to subsequently receive hypomethylating agent (HMA) therapy. HMA-treated patients had higher IPSS-R risk, higher bone marrow blast percentages, and lower hemoglobin levels. Demographic characteristics of the final cohort were similar to those in previously published studies from which most of these samples were selected (Bejar et al. 2011; Bejar, Levine, and Ebert 2011).

Table 2.1 Patient characteristics.

	N (%)
Number of Cases	141
Gender	
Male	103 (73)
Female	38 (27)
Age at Diagnosis	
< 70 years	55 (39)
>= 70 years	86 (61)
FAB	
RA	47 (33)
RARS	27 (19)
RAEB	48 (34)
RAEBT	5 (4)
RCUD/MD	4 (3)
CMML	7 (5)
Other	3 (2)
IPSS-R[*]	
Very Low	23 (21)
Low	31 (28)
Intermediate	20 (18)
High	24 (21)
Very High	14 (13)
Cytogenetics[†]	
Normal	89 (63)
Abnormal, Not Complex	41 (29)
Complex	11 (8)
Blood Counts	
WBC (x10^9/L), median (range)	3.7 (0.9–95.2)
ANC (x10^9/L), median (range)	1.6 (0.1–28)
Hgb (g/dL), median (range)	9.7 (5.8–15.2)
Platelet (x10^9/L), median (range)	102 (6–987)
BM Blast (%), median (range)[‡]	1 (0–28)
Treatment during follow-up	
Azacitidine	28 (20)
Decitabine[§]	20 (14)

[*]29 patients were missing data for one or more variables in the IPSS-R and could not be classified. [†]Several patients had incomplete cytogenetic information so we stratified them into known groups.[‡]20 patients had missing data for BM Blast %, these patients were excluded from survival modeling.[§]One DAC treated patient stopped treatment after one cycle.

Identification of Genetically Distinct DNA Methylation Subtypes of MDS

We employed the unsupervised *Onco*-GPS clustering approach to identify coherent subgroups defined by DNAm profiles (schema shown in Supplementary Figure 2.1). Repeated NMF decompositions with random seeding found the greatest stability with five NMF components, indicated by a peak in the cophenetic correlation coefficient (Supplementary Figure 2.2 A). Consensus clustering of patients by the amplitudes of these five NMF components resulted in stable clustering solutions (as indicated by the cophenetic correlation) for four and five patient clusters, however, only the five cluster solution provided the most distinct clusters as measured by the average silhouette width (Supplementary Figure 2.2 C-F). Subdivisions of less than five clusters yielded groups comprised of patient samples with heterogeneous component amplitudes (Supplementary Figure 2.2 G-H). The five cluster solution yielded clusters comprised of samples with high amplitudes for a single component, where each cluster was largely defined by a separate component (Figures 2.1 A-B and Supplementary Figure 2.2). For example, patient samples in Cluster A were significantly associated with component 1, but not with any other component. For comparison, we also performed consensus clustering of raw methylation data for the most variably methylated CpG tiles and observed that five clusters again provided a similarly stable solution as indicated by the delta-area plot which shows the change in the area under the cumulative distribution function (Supplementary Figures 2.3).

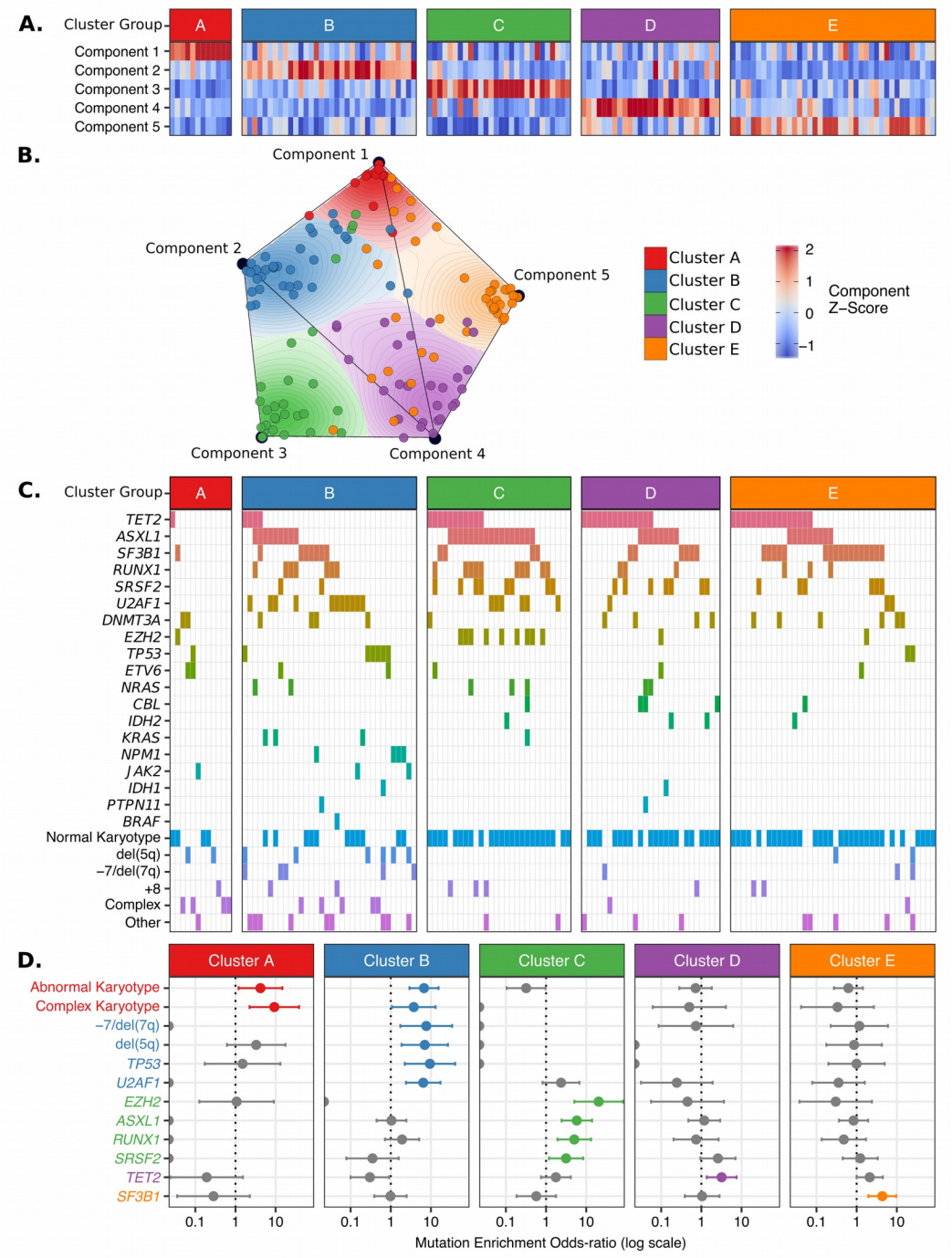


Figure 2.1 Onco-GPS computational clustering approach identified five NMF components which clustered into five patient groups with distinct DNA methylation states. A) NMF cmoponent amplitudes (rows) are plotted for each patient (columns), by methylation cluster membership. **B)** A 2-dimensional projection of each patient's NMF component amplitudes where each point is a single patient colored by cluster membership. C) Five methylation clusters categorize patients (columns) with distinct genetic and cytogenetic abnormalities (rows). D) Odds ratio of enrichment for patients with particular genetic lesions within each methylation cluster. Significantly enriched lesions (FDR <0.1) are highlighted in color.

Methylation Clusters are Enriched for Distinct Patterns of Genetic Lesions

Each of the five methylation clusters were significantly enriched for specific genetic and cytogenetic lesions (Figure 2.1C-D and Supplementary Table 2.2). Cytogenetic abnormalities were unequally distributed across clusters, with abnormal and complex karyotype patients enriched in Clusters A and B and depleted in Clusters C through E (Cluster A: complex OR 9.3, FDR<0.1; Cluster B: abnormal OR 7.6, FDR<0.01). Alterations of chromosome 7 (OR 9.4, FDR<0.1) and del(5q) (OR 7.0, FDR<0.1) were enriched in Cluster B. Total mutation burden was lowest in Cluster A, with a mean of 0.8 mutations per patient compared to a mean of 1.5-2 mutations per patient for Clusters B through E (Figure 2.1D). Mutations known to convey adverse risk occurred more frequently in Clusters B and C. Cluster B had significant enrichment of mutations in *TP53* (OR 9.4, FDR<0.05) and *U2AF1* (OR 6.4, FDR<0.01). Cluster C showed enrichment of mutations in *EZH2* (OR 20.4, FDR<0.01), *ASXL1* (OR 5.8, FDR<0.01), and *RUNX1* (OR 5.0, FDR<0.05). The splicing factors were mutually exclusive and found in different clusters, with *U2AF1* mutations enriched in Cluster B (OR 6.4, FDR<0.01), *SRSF2* mutations enriched to a less than significant level in Cluster C (OR 3.2, FDR>0.1) and *SF3B1* mutations enriched in Cluster E (OR 4.4, FDR<0.05). *TET2* mutated patients were present in several clusters but uniquely enriched in Cluster D (OR 3.2, FDR<0.1).

Methylation Components Correlate with Specific Genetic Lesions

The enrichment of specific mutations within each methylation cluster appeared to be driven by the association of these mutations with the single NMF component that primarily defined each cluster (Figure 2.2). For example, while samples with mutations in *ASXL1* and *RUNX1* were enriched in Cluster C (defined primarily by NMF component 3), samples assigned to other clusters with mutations in these same genes also possessed higher NMF component 3 scores (Figure 2.2B). *TET2* mutant samples associated strongly with NMF component 4 and were enriched in Cluster D, although *TET2* mutant samples assigned to Clusters A, B, C, and E also had higher component 4 scores than their wild-type counterparts. Using an information coefficient-based permutation test, we found unique associations between NMF components and genetic lesions (Figure 2.2). The most significant component-mutation associations were for Component 3-*ASXL1* (IC 0.417, $p<0.001$), Component 3-*RUNX1* (IC 0.351, $p<0.001$), Component 3-*EZH2* (IC 0.311, $p<0.001$), and Component 4-*TET2* (IC 0.394, $p<0.001$). The most prevalent splicing factor mutations were each associated with different NMF components (Component 3-*SRSF2*, IC 0.232, $p\leq0.01$; Component 4-*SRSF2*, IC 0.262, $p\leq0.01$; Component 2-*U2AF1*, IC 0.257, $p\leq0.01$; Component 5-*SF3B1*, IC 0.211, $p=0.02$).

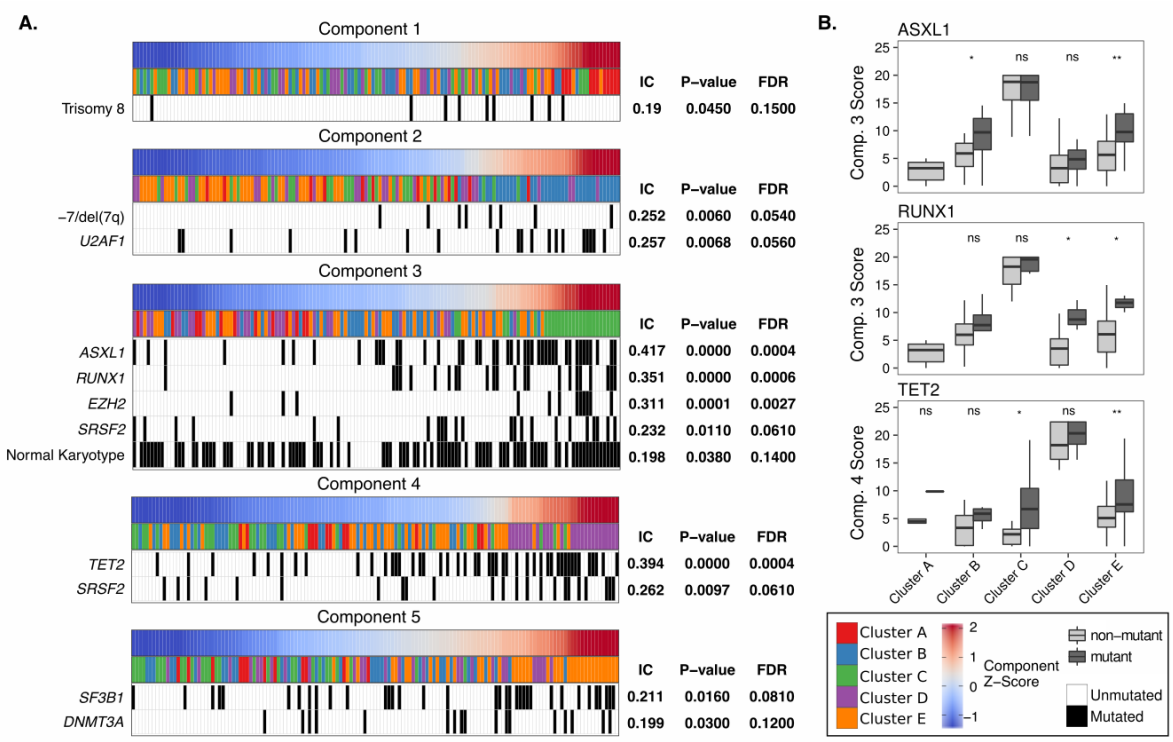


Figure 2.2 Individual NMF component amplitudes correlate with specific genetic lesions. Each NMF component is associeated with specific genetic lesions, and these genetic profiles are unique for each individual component (with the exception of *SRSF2* mutations which were associated with both components 3 and 4). **A)** patient samples (columns) are plotted in order of their component amplitudes (first row, blue=low, red=high) with cluster membership (second row, colored tiles) and presence of specific genetic mutations (black/white tiles). (IC = Information coefficient; *p*-values and False Discovery Rates are based on 100,000 permutations). **B)** Example of how Components 3 and 4 are associated with genetic lesions even in clusters which are more closely associated with different components. Boxplots for component scores are plotted by cluster for patients with or without the specific lesion for the 3 most highly associated lesions. (ns, not significant; *, p<0.05; **, p<0.01)

Clinical and Prognostic Significance of Epigenetic Clusters

Small differences in clinical features were observed between methylation clusters (Figure 2.3A) including platelet and neutrophil counts. In addition, the median age among Cluster A patients was significantly lower at 63 years compared to a median of 72 years for Clusters B through E (*p*=0.01). HMA-treated patients were distributed across all clusters with no enrichment by response status (Supplementary Figure 2.4). OS curves for patients in each cluster identified two

major patterns with those in Clusters B and C displaying inferior OS compared to patients in Clusters A, D, and E (Figure 2.3B). We combined clusters with similar median OS into "High" (Clusters B and C) and "Low" (Clusters A, D and E) cluster risk groups. In univariate analysis, High and Low cluster risk groups had statistically significant differences in OS (HR, 1.95; 95% CI,1.24–3.08; p<0.01) (Table 2.2, Supplementary Table 2.3). By a multivariate analysis that assessed known prognostic variables in the IPSS-R, DNAm cluster risk group was the most statistically significant single predictor of OS (HR, 2.02; 95% CI 1.25–3.27 ; p<0.01) (Table 2.2). In a second multivariate model which also considered somatic mutations, DNAm cluster risk group remained a significant predictor for OS (HR, 1.6; 95% CI 0.95–2.71 ; p=0.08) (Table 2.2). Notably, methylation cluster risk group was a stronger predictor of OS in IPSS-R lower vs. higher risk patients and retained prognostic value in HMA treated patients (Supplementary Figures 2.5). A subset of patients without known prognostic mutations could also be stratified by methylation cluster risk groups, which indicates that the prognostic value of methylation clusters is not fully driven by the differential enrichments of prognosis-associated somatic mutations across clusters (Supplementary Figure 2.6) (Bejar et al. 2011; Haferlach et al. 2014; Thol et al. 2012).

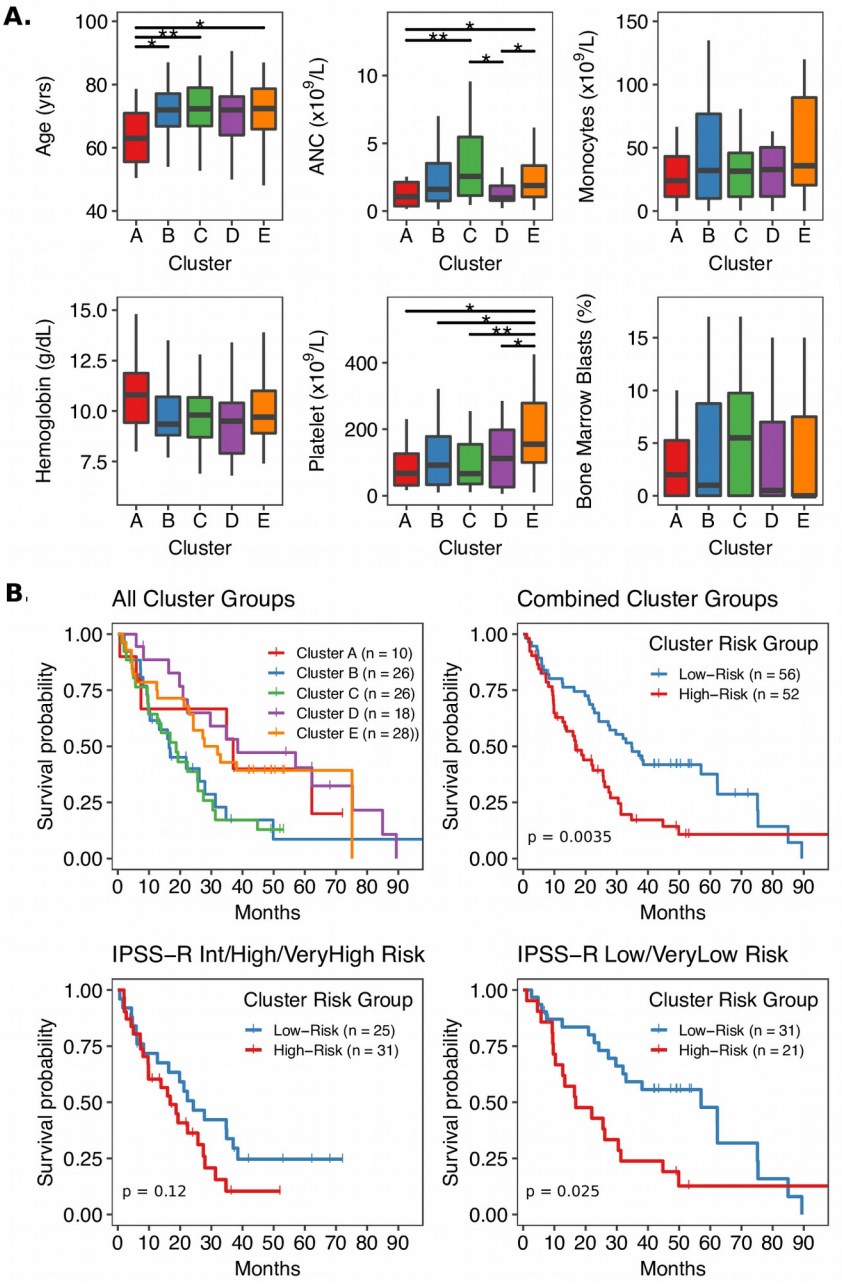


Figure 2.3 Methylation cluster display differences in clinical features and overall survival. **A)** Median with ranges (25th to 75th percentiles) are shown for patient age, absolute neutrophil count (ANC), absolute monocyte count, hemoglobin, platelet count and bone marrow blast percentage. Variables with significant pairwise comparisons between clusters by Wilcoxon rank-sum test are indicated. (*, p<0.05; **, p<0.01) **B)** Kaplan-Meier (KM) curves stratified by methylation cluster are shown for all patients with available survival data (top left). KM curves for combined cluster risk groups are shown, where the “High-Risk” patients are those in Clusters B and C, and the “Low-Risk” patients are those in Clusters A, D, and E (top right). KM curves for patients with Intermediate, High, or Very High IPSS-R risk are stratified by cluster risk groups (bottom left), and KM curves for patients with Low or Very low IPSS-R risk are stratified by cluster risk groups (bottom right).

Table 2.2. Methylation cluster risk membership retains prognostic significance by multivariate analysis

Model 1: Clinical Features and Cluster Risk Retained by Multivariable Analysis

		Univariable		Multivariable	
	N (%)	**HR [95% CI]**	**P-value**	**HR [95% CI]**	**P-value**
Cluster Risk Group					
High vs. Low	52 (48)	1.95 [1.24-3.08]	0.003	2.02 [1.25-3.27]	0.004
Clinical Features					
Karyotype			0.13		
Abnormal (Not Complex) vs. Normal	33 (31)	1.35 [0.82-2.24]	0.24	1.43 [0.85-2.39]	0.174
Complex vs. Normal	9 (8)	2.20 [0.93-5.20]	0.071	2.79 [1.15-6.75]	0.023
Bone Marrow Blasts (%)			0.014		
5 - 10 vs. < 5	23 (21)	2.14 [1.22-3.77]	0.008	2.18 [1.23-3.86]	0.007
11 - 30 vs. < 5	23 (21)	1.77 [1.01-3.09]	0.046	1.40 [0.79-2.49]	0.252

Model 2: Clinical Features, Somatic Mutations, and Cluster Risk Retained by Multivariable Analysis

		Univariable		Multivariable	
	N (%)	**HR [95% CI]**	**P-value**	**HR [95% CI]**	**P-value**
Cluster Risk Group					
High vs. Low	52 (48)	1.95 [1.24-3.08]	0.003	1.60 [0.95-2.71]	0.076
Clinical Features					
Karyotype			0.13		
Abnormal (Not Complex) vs. Normal	33 (31)	1.35 [0.82-2.24]	0.24	1.30 [0.75-2.28]	0.353
Complex vs. Normal	9 (8)	2.20 [0.93-5.20]	0.071	2.91 [1.16-7.28]	0.022
Bone Marrow Blasts (%)			0.014		
5 - 10 vs. < 5	23 (21)	2.14 [1.22-3.77]	0.008	2.08 [1.16-3.73]	0.014
11 - 30 vs. < 5	23 (21)	1.77 [1.01-3.09]	0.046	1.55 [0.85-2.83]	0.15
Somatic Mutations					
RUNX1 mutated vs. not mutated	19 (18)	2.12 [1.22-3.68]	0.008	1.97 [1.07-3.64]	0.031
EZH2 mutated vs. not mutated	9 (8)	2.91 [1.37-6.18]	0.005	2.19 [0.91-5.24]	0.079
TP53 mutated vs. not mutated	9 (8)	3.16 [1.41-7.08]	0.005	2.40 [0.96-6.01]	0.062

*P-values for individual categories within variables were calculated using the Wald test. P-values for full variables correspond to a log-rank test. Multivariable models were constructed by optimizing the AIC which is why some multivariable p-values are less than 0.05.

DMRs Between Methylation Clusters Are Enriched for Distinct HSPC Genetic Regulatory Features

While the five methylation clusters were defined using a subset of the most variably methylated regions, we identified genome-wide methylation patterns associated with each cluster. Cluster-specific differentially methylated regions (DMRs) were extracted by comparing members of each cluster with non-members at all CpG tiles with sufficient coverage (Figures 2.4, Supplementary Figure 2.7).

DMRs specific to Cluster B and Cluster D were almost entirely hypermethylated, where DMRs specific to Cluster C and Cluster E were primarily hypomethylated. Cluster A-specific DMRs were both hypermethylated and hypomethylated. We then examined DMRs associated with specific genetic subgroups of patients. In general, comparison of mutated vs. wild type patients *within a single cluster* yielded only a small amount of differential methylation. However, comparing patients with shared mutations *across different clusters* identified much larger differences (Supplementary Figure 2.7). For example, there were few DMRs when comparing all *TET2*-mutated patients to *TET2*-wild type patients (11 hypermethylated- and 10 hypo-methylated DMRs). This is illustrated in differential methylation volcano plots comparing all patients with specific gene mutations to wild-type patients (Supplementary Figure 2.8A). Differential methylation becomes more pronounced when comparing those with a sole gene mutation versus without the mutation, such as patients with only *TET2* mutations versus *TET2*-wild type patients (Supplementary Figure 2.8B). However, DMR differences between *TET2* mutants with different cluster membership were significantly larger than those between *TET2* mutants and non-mutants within the same cluster (Figure 2.7), suggesting that *TET2* non-mutants in the same cluster had converged to a comparable epigenetic state.

To determine if cluster-specific DMRs occurred more frequently at specific types of genomic regulatory regions, we measured overlap between DMRs and regulatory regions defined in the Roadmap Consortium 15-state chromatin state

model for human mobilized CD34^{+} cells and calculated enrichment of DMRs in each type of chromatin state (Figure 2.4) (Roadmap Epigenomics Consortium et al. 2015). Cluster A DMRs were specifically enriched in regions flanking the transcription start site of genes that are actively transcribed in CD34^{+} cells, with a trend towards hypermethylation. Clusters C and D DMRs were both highly enriched in states predicted to act as enhancers, with opposing trends in methylation. Clusters B and E DMRs were highly enriched for states predicted to be bivalent promoters and bivalent enhancers.

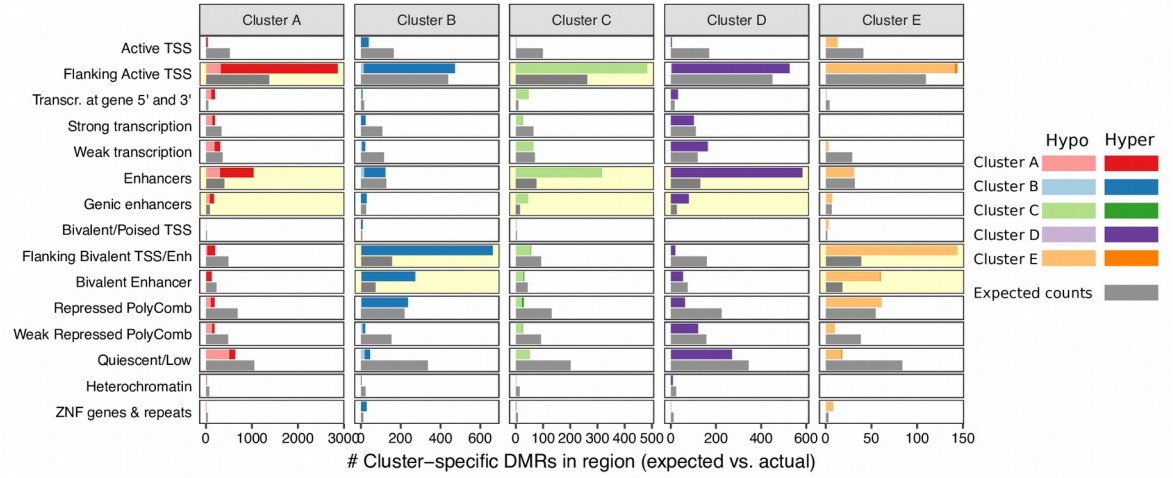


Figure 2.4 Cluster-specific differentially methylated regions are enriched for distinct regulatory chromatin states in reference CD34^{+} cell epigenome. Cluster-specific DMRs display distinct enrichment and patterns of differential methylation at epigenetic regulatory segments in a reference hematopoietic stem cell (HSC) as defined by the chromHMM 15-state genome segmentation model created from integrated epigenetic datasets (Roadmap Consortium). Clusters with significant positive enrichment of DMRs within a given segment are highlighted in light yellow. (TSS = transcription start site; Enh = Enhancer; Transcr. = Transcribed; Darker colored bars = hypermethylated DMRs; lighter colored bars = hypomethylated DMRs; gray bars = expected counts which is defined as (number of cluster-specific DMRs) x ((number of CpG-tiles within the given genome segment) / (all CpG tiles covered))

NMF Component Amplitudes for CpG's Correlate with Distinct Reference HSPC Genetic Regulatory Features

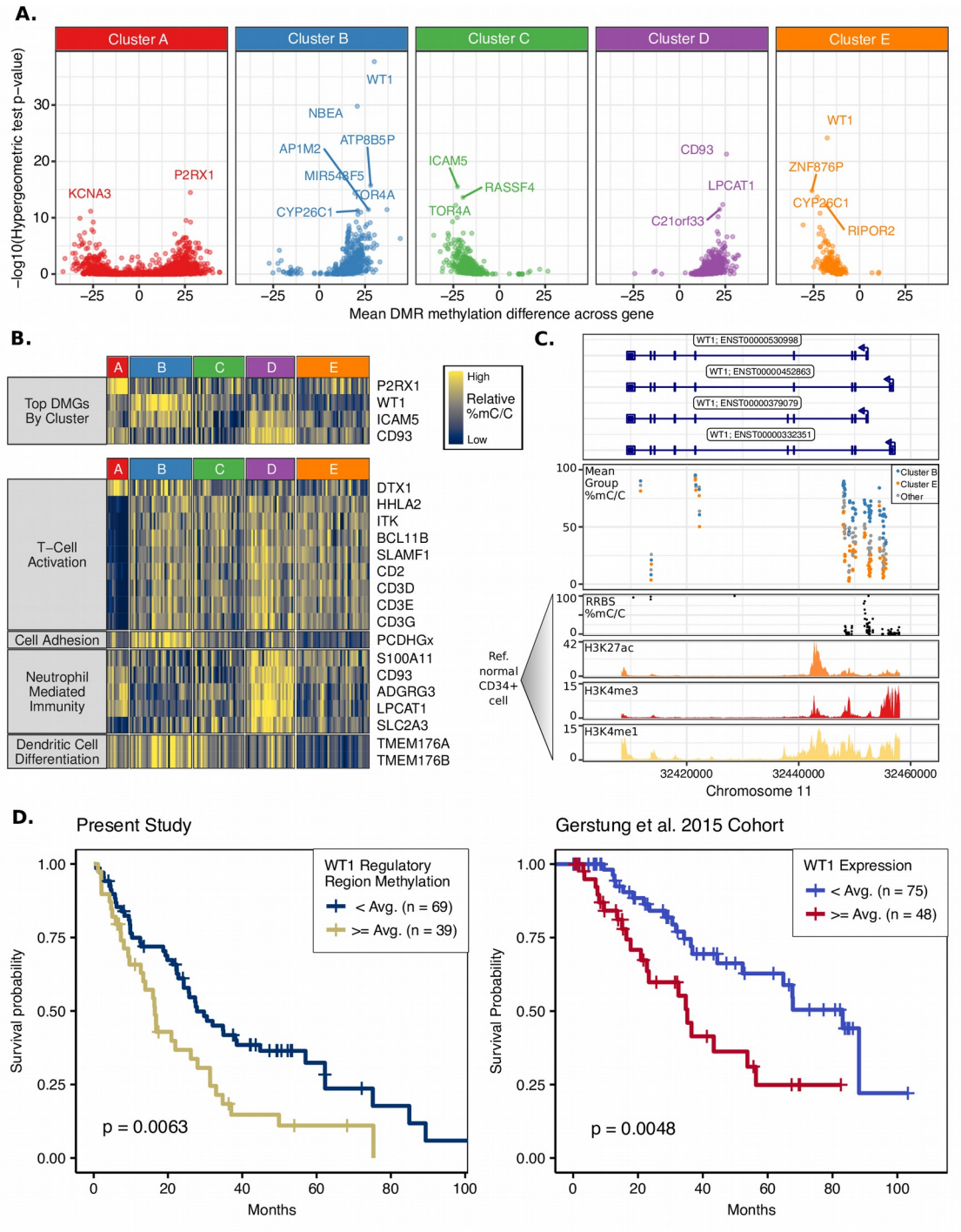
We also measured the association of each NMF component with chromatin states defined in reference $CD34^+$ cells. As each CpG tile is assigned an amplitude for each NMF component we can calculate the strength of association by comparing the component amplitudes of CpGs falling within a given chromatin state to those falling outside a given state. We calculated the information coefficient between each NMF component and each state as a measure of the association strength (where IC=1 is the strongest positive association and IC=-1 the strongest negative association). We observed that each component was associated with unique chromatin states, and the pattern and strength of association was in some cases different than that of the cluster-specific DMRs (Supplementary Figure 2.9). The top scoring CpG tiles for component 3, which is highly associated with Cluster C, were specifically enriched for polycomb repressed states. Notably, Cluster C was specifically enriched for patients carrying mutations in genes associated with polycomb regulation (*EZH2*, *ASXL1*, *SRSF2*) and component 3 was highly associated with mutation of these genes independent of cluster membership. Component 2 CpG tiles were highly and specifically associated with bivalent chromatin states. Component 3 CpG tiles were associated with promoter and gene flanking states. Component 4 CpG tiles were specifically associated with enhancer states coinciding with findings for Cluster 4 specific DMRs.

Differentially Methylated Genes are Enriched in Distinct Pathways

We next performed Gene Ontology (GO) term enrichment analysis of cluster-specific differentially methylated genes (DMGs) and identified enrichment for many ontologies related to T-cell activation and differentiation for Cluster A DMGs (Figure 2.5). Cluster B was solely enriched for biological processes related to cell-cell adhesion, largely driven by a region of differential methylation centered on the *protocadherin gamma* gene cluster. Cluster D DMGs were enriched for genes involved in neutrophil mediated immunity. The most differentially methylated gene between clusters was *WT1*, which was hypermethylated in Cluster B and hypomethylated in Cluster E (Figure 2.5A-B). The majority of DMRs within *WT1* were located in a downstream regulatory region that had high levels of H3K27ac, H3K4me1, and H3K4me3 signal in reference HSPC data, which are considered activating histone marks (Shlyueva, Stampfel, and Stark 2014). Patients in our study who had above average methylation in the 2500bp region downstream of the WT1 TSS had significantly shorter OS (Figure 2.5C). We then examined *WT1* gene expression and survival data from the Gerstung et al. study of MDS patients and found that patients with above average expression of *WT1* had significantly shorter OS (Figure 2.5D) (Gerstung et al. 2015). We examined the relationship between methylation and expression of *WT1* from an external cohort of 200 patients with AML published by The Cancer Genome Atlas (TCGA) Research Network (The Cancer Genome Atlas Research Network 2013) and found a positive correlation between hypermethylation and increased gene expression at CpG loci that overlapped with our study suggesting an association between

methylation at these sites and gene expression (Supplementary Figure 2.10). The next most differentially methylated gene was *CD93*, which was hypermethylated at a regulatory region surrounding the TSS in our prognostically lower risk Clusters A, D, and E. We observed a trend for improved survival in patients with above average methylation of this regulatory region while gene expression data from Gerstung et al. identified a novel and significant association between survival and *CD93* expression (Supplementary Figure 2.11). We observed a strong negative correlation between CD93 expression and DNAm in the TCGA cohort (Supplementary Figure 2.12), implying that adverse risk clusters B and C with lower average methylation may have had higher expression that correlates with the survival results observed in the Gerstung et al. data.

Figure 2.5 Cluster-specific differentially methylated genes are enriched in distinct pathways and may contribute to prognostic differences between clusters. A) Gene enrichment for DMRs (y-axis = -log10(*p*-value) from hypergeometric test of enrichment for DMRs present within a gene +/- 1kb, relative to all CpG-tiles within gene; x-axis = mean methylation difference of DMRs within gene +/- 1kb) **B)** Heatmap of average methylation for the most highly differentially methylated genes for each set of cluster-specific DMRs (Top; samples in columns, genes in rows). *WT1* was the most highly differentially methylated gene for both Clusters B and E. Heatmap of the average methylation of genes in significantly enriched biological process gene ontologies from gene ontology enrichment analysis of cluster-specific DMGs (Bottom; samples in columns, genes in rows; biological process ontology indicated by left panel of heatmap). **C)** methylation of cluster-specific differentially methylated CpG tiles within the *WT1* gene (2nd panel). Grey points represent the average methylation of patients in clusters A, C, and D which all had similar levels of methylation at these loci. Reduced Representation Bisulfite Sequencing (RRBS) methylation as well as H3K27ac, H3K4me3, and H3K4me1 Chip-seq signal tracks correspond to a reference dataset of GM-CSF mobilized CD34^{+} cells from a healthy male donor (3rd through 6th panels; Roadmap Consortium epigenome E051). Gene model track (top panel) correspond to isoforms of *WT1* found in RefSeq with their corresponding Ensembl ID's. **D)** Kaplan-Meier curves for the Gerstung et al. 2015 cohort stratified by average *WT1* expression (left) and for the present study stratified by average *WT1* regulatory region methylation (region +/-2500bp from TSS) (right).


A.
Cluster A
Cluster B
Cluster C
Cluster D
Cluster E
−log10(Hypergeometric test p−value)
Mean DMR methylation difference across gene
KCNA3
P2RX1
WT1
NBEA
ATP8B5P
AP1M2
MIR548F5
TOR4A
CYP26C1
ICAM5
RASSF4
TOR4A
CD93
LPCAT1
C21orf33
WT1
ZNF876P
CYP26C1
RIPOR2
B.
Top DMGs By Cluster
P2RX1
WT1
ICAM5
CD93
High
Relative %mC/C
Low
T−Cell Activation
DTX1
HHLA2
ITK
BCL11B
SLAMF1
CD2
CD3D
CD3E
CD3G
Cell Adhesion
PCDHGx
Neutrophil Mediated Immunity
S100A11
CD93
ADGRG3
LPCAT1
SLC2A3
Dendritic Cell Differentiation
TMEM176A
TMEM176B
C.
WT1; ENST00000530998
WT1; ENST00000452863
WT1; ENST00000379079
WT1; ENST00000332351
Mean Group %mC/C
Cluster B
Cluster E
Other
RRBS %mC/C
Ref. normal CD34+ cell
H3K27ac
H3K4me3
H3K4me1
32420000
32440000
32460000
Chromosome 11
D.
Present Study
WT1 Regulatory Region Methylation
< Avg. (n = 69)
>= Avg. (n = 39)
Survival probability
p = 0.0063
Months
Gerstung et al. 2015 Cohort
WT1 Expression
< Avg. (n = 75)
>= Avg. (n = 48)
Survival Probability
p = 0.0048
Months

2.4 Discussion and Conclusions

In this study, we applied a computational clustering approach based solely on the DNAm profiles of 141 bone marrow DNA samples from patients with MDS to identify subgroups with different DNAm states defined at key regulatory regions of the genome. Using this agnostic approach, five methylation subtypes were described that possess unique patterns of enrichment for genetic lesions, differences in regulatory element methylation, and associations with survival that are independent of prognostic clinical variables and somatic mutations. Our results demonstrate that somatic genetics are insufficient to predict the methylation state for a given patient with MDS, and that DNAm profiles may supplement prognostic information in the context of known molecular and clinical variables.

A key conclusion of our study is that in MDS, various mutational profiles can share a common DNA methylation state even if these states are enriched for specific mutated genes. Somatic mutations alone are not sole determinants of DNAm which may be integrating additional cues from elements like the microenvironment, differentiation state of cells, or age-related changes. This could reflect preferential selection for mutations in a given pre-existing epigenetic state or shared mechanisms engaged by distinct lesions that converge on common epigenetic profiles. For example, we observed that mutually exclusive mutations in *EZH2* and *SRSF2* were both enriched in Cluster C, consistent with studies suggesting these mutations share pathogenic mechanisms (Meggendorfer et al. 2012; Papaemmanuil et al. 2013). We also observed that patients with *TET2*

mutations were distributed to Clusters B, C, and E but enriched only in Cluster D. In differential methylation analysis comparing *TET2* mutants and non-mutants *within* a single cluster, we found few differences with no strong trends in methylation. In contrast, when we compared *TET2* mutants within a single cluster to *TET2* mutants in other clusters, we observed differences in methylation that were consistent with the global differences between clusters, suggesting that *TET2* mutant patients can diverge to different epigenetic states, potentially driven by patterns of co-incident mutations or other factors, thus suggesting that DNAm differences cannot be attributed simply to particular mutations or cytogenetic abnormalities. Mutations in *U2AF1*, *SRSF2*, and *SF3B1* are known to occur in a mutually exclusive manner possibly because they are generally not tolerated as co-incident lesions or because they may share a common pathogenic mechanism (S. C.-W. Lee et al. 2016, 2018). In our study, mutations of each of these genes were enriched in separate epigenetic clusters and associated with distinct NMF components, indicating that these mutations drive distinct oncogenic states and likely engage unique pathogenic mechanisms.

Another important finding in our study was the enrichment of cluster-specific differentially methylated regions located at distal regulatory regions of the HSC epigenome. Much of the previous work on DNAm in MDS has focused on CpG-rich promoter regions or LINE elements (Römermann et al. 2008; Zhao et al. 2014; Garcia-Manero et al. 2011; Qin et al. 2011). While we identified DNAm differences at promoters, there were an equal or greater number of differences at non-promoter

regions of the genome. Notably, the *TET2* mutated samples enriched in Cluster D had the greatest representation of DMRs at enhancer regions as did the loci that defined Component 4. These observations are consistent with the putative role of *TET2* in maintaining the methylation state at enhancer regions (Hon et al. 2014; Kasper Dindler Rasmussen and Helin 2016; Yamazaki et al. 2015; Kasper D Rasmussen et al. 2015) and may help explain why a number of studies in AML and MDS have struggled to find clear relationships between *TET2* mutations and DNAm (Yamazaki et al. 2012; Pérez et al. 2012; Figueroa, Abdel-Wahab, et al. 2010; Yamazaki et al. 2015; Meldi et al. 2015; Ko et al. 2010). Consideration of the underlying oncogenic state may resolve the relative impact of mutations like those in *TET2* whose effects may vary in different contexts. Our data suggest that the lack of a relationship between DNAm and *TET2* mutations may be partially explained by the presence of a subset of *TET2* wild-type patients displaying an enhancer hypermethylation phenotype and *TET2* mutant patients often harbor co-mutations that may also strongly impact their global methylation patterns (e.g., mutations in *ASXL1*, *RUNX1*, or *EZH2*). Overall, our differential methylation analysis demonstrates that the differences in methylation between genetic subgroups were vastly overshadowed by differences driven by methylation cluster membership.

We did not observe significant differences in cytopenias or bone marrow blast percentages across clusters. Yet, when examining clinical outcomes, we discovered that DNAm clusters had two distinct patterns of prognostic risk. While

these OS differences are modest and require independent validation, cluster groups retain their prognostic significance even when accounting for somatic mutations and clinical factors included in the IPSS-R. This suggests that additional prognostic information may be integrated by convergent epigenetic states influenced by, but not exclusive to, somatic mutations. In support of this idea, we showed that even among the subset of patients lacking mutations in genes with known prognostic impact, DNAm cluster risk groups could stratify patients by OS. Additionally, we identified several genes with significant differential methylation between clusters, including *WT1,* which may contribute to the prognostic associations we identified. This is consistent with prior work that has associated *WT1* expression with poor prognosis in MDS and AML (Cilloni et al. 2003; Niavarani et al. 2016; Kobayashi et al. 2016; Yanan Jiang et al. 2018). More novel was the association between shorter OS and *CD93* hypomethylation in our cohort and overexpression in an external MDS cohort, as *CD93* is a transmembrane receptor implicated in the pathogenesis of several malignancies (Olsen et al. 2015; Lugano et al. 2018; Langenkamp et al. 2015). Its role in myeloid malignancies is relatively unknown, however, it was shown to be essential to the oncogenic potential of non-quiescent leukemia stem cells in MLL rearranged AML, therefore this gene may warrant further investigation in MDS (Iwasaki et al. 2015).

Finally, our findings are consistent with those of Shiozawa et al, which identified two major subtypes of MDS with differences in genetic lesions and time to transformation to AML using an analogous unsupervised classification approach

on gene-expression data from primary MDS $CD34^+$ cells (Shiozawa et al. 2017). Although a different genomic analysis was utilized on a more purified cell population, their study similarly concluded that somatic genetics were insufficient to define clinically relevant disease characteristics in MDS. In the two groups defined in their study, mutations in *TET2* and *SF3B1* were enriched in, but not unique to, the subgroup with longer time to AML transformation, while abnormalities in chromosome 7 and mutations in *RUNX1* and *TP53*, among others, were more frequent in the subgroup with shorter time to AML transformation. These findings are consistent with the genetic abnormalities enriched in the two cluster risk groups defined by our study, where mutations in *TET2* and *SF3B1* were solely enriched in our low-risk clusters D and E, while chromosome 7 abnormalities and *TP53* and *RUNX1* mutations were enriched only in our high-risk clusters B and C.

A potential limitation of our study is that while samples were collected from treatment-naïve MDS patients, only a subset of patients were known to have later received HMA therapy. Although HMA-treated patients were distributed across all methylation clusters, there was no signal associated with eventual response. Larger methylation profiling studies of MDS patients with known treatment outcomes are ongoing and may increase our ability to incorporate predictive epigenetic and genetic biomarkers into clinical practice.

In conclusion, we demonstrate that using novel computational methods to agnostically classify primary MDS patients by their DNAm patterns can identify

subtypes of MDS characterized by distinct patterns of genetic lesions, regulatory region methylation, and prognostic risk. Our results highlight the importance of DNAm in MDS disease biology as a convergent oncogenic phenotype and support future studies investigating DNAm profiles as clinically relevant biomarkers.

2.5 Materials and Methods

Patients and Samples

A total of 154 patients with treatment-naïve MDS were considered for this study. Samples and data from 140 patients were obtained and processed as previously described (Bejar et al. 2011; Bejar, Levine, and Ebert 2011; Steensma et al. 2009; Bejar, Lord, et al. 2014b). Samples and data from an additional 14 patients were included from the University of California San Diego (UCSD) Moores Cancer Center. All samples were collected with patient consent under protocols approved by Institutional Review Boards and in accordance with the Declaration of Helsinki. Clinical data to determine the Revised International Prognostic Scoring System (IPSS-R) score was available at the time of sample collection (Greenberg et al. 2012).

DNA Mutation Analysis

Genomic DNA samples were isolated from bone marrow aspirate mononuclear cells and most were genetically characterized in previous studies. Samples collected at U.C. San Diego were genetically characterized by targeted capture of DNA and sequenced on the Illumina platform. Using the Burrows Wheeler Aligner (BWA v0.7.12) MEM modules for paired end reads, sequenced

reads were aligned against the reference genome (hg19). Duplicate reads were flagged and removed using Picard tools (V1.91). GATK v3.2 was used for base recalibration prior to variant calling, and also for local realignments for insertion/deletions using the reference variant databases. Somatic variants were called using LoFreq v2.1.1 for all variants at >=1% variant frequency. Additional variant filtering after variant calling are used with the following parameters: VF (variant frequency) <0.05, read depth at variant site <20, GQ and/or QUAL scores <30, IndelRepeatFilter >8. Filtered called variants were first annotated using ANNOVAR (February 2016). Variant located outside corresponding protein coding region or splice site, synonymous variants that were not predicted to alter splicing were filtered out. To remove common polymorphisms, variants with population frequencies of at least 1% in either 1000 genomes (October 2014) or ExAC (v.3.1) were also filtered out. Of the remaining variants, potential functional impact was assessed using SIFT (v.5.1.0), Polyphen2 (v2.2.2), and mutation assessor. Variants predicted to alter splicing were assessed as described in Jian et al., (2014). Variants were assessed if they have been previously identified in various databases including looked up in COSMIC (V77) and cBioPortal, as well as ClinVar. All candidate somatic variants were manually reviewed in the Integrated Genome Viewer for accuracy.

Generation of BSPP methylomes

BSPP libraries were generated and sequenced using a standard 150bp paired-end read protocol on Illumina HiSeq 2500. Sequencing reads were

trimmed, processed and aligned as previously described (Diep et al. 2012; Deng et al. 2009). Briefly, adapter content, biased read positions and poor quality bases were removed from bisulfite-converted sequencing reads (trim-galore version 0.4.0) and reads were aligned to the hg19 reference genome. For alignment of bisulfite converted sequencing reads, all cytosines were converted to thymines, and for the reverse-complement strand, all guanines to adenines, so that reads could be mapped to a three-letter genome. Reads were mapped to both the Watson and Crick strands of the bisulfite-converted hg19 genome (Burroughs-Wheeler algorithm mem version 0.7.12 with options '-B2 -c1000'). Alignments with mapping-quality scores <5 were discarded and only the alignment with the highest mapping quality score was kept for each read. Overlapping paired-end reads were clipped (BamUtil clipOverlap function) and encoded read sequences were replaced by the original read sequences in the final BAM alignment files which were used for quantifying the methylation levels of CpGs. CpGs with ≥10x and <500x read coverage were included and summarized into 25bp tiles by the coverage-weighted average of CpGs within the 25bp region. Filtering and quality control analysis yielded 141 samples covering 246,088 25bp tiles for our final cohort.

DNA Methylation Quality Control

Sample methylation files were filtered to exclude CpG tiles located on the X and Y chromosomes, as well as mitochondrial DNA. Methylation files for all samples were combined into a single matrix for QC and subsequent analyses.

Methylation percentage for each CpG cytosine was quantified as the number of reads in which that cytosine was methylated divided by the total number of reads covering that position. Samples with missing data for >15% of CpG tiles that were covered by >80% of all samples were removed, and CpG tiles that were missing in >5% of the remaining samples were excluded from downstream analyses (sample n=144 after filtering). We employed an iterative process for missing value imputation and outlier identification by principal component analysis (adapted from (Hannum et al. 2013)), where missing data were imputed using K-nearest neighbors (R package 'impute') then the first two principal components were computed for the completed matrix (R package 'pcaMethods'). The first principal component score for each sample was converted into a Z-score by subtracting the population mean and dividing by the population standard deviation. This Z-score was converted to a False Discovery Rate using the gaussian cumulative distribution function (R 'pnorm' function) and the Benjamini-Hochberg procedure (Benjamini and Hochberg Yosef 1995) and samples with an FDR <=0.2 were removed. This procedure was repeated (starting at imputation) until there were no more outlier samples, yielding the final cohort of n=141 samples covering 246,088 25bp tiles.

Identification of Epigenetic Subtypes by Onco-GPS

Using the Onco-GPS computation clustering framework (https://github.com/UCSD-CCAL/onco-gps-paper-analysis), NMF was performed on the 3% most variably methylated CpG tiles (n=7382) to decompose these methylation data into

a smaller matrix of NMF components that served as inputs for consensus clustering (Brunet et al. 2004; Kim et al. 2017). NMF is a feature extraction algorithm that combines similar attributes (in this case methylation values at collections of CpG loci) into summary components which represent cohesive properties of the data (in this case localized methylation differences among patients), thus reducing redundancy while maintaining interpretability. The number of NMF components and consensus clusters was selected based on the cophenetic correlation coefficient peak and visual inspection of the clustering result after evaluating different numbers of clusters (Supplementary Figure 2.2). We observed a peak in the CCC for Kc=5 NMF components after repeating the NMF decomposition 100 times for each K in the range of 2 to 8 (Supplementary Figure 2.2). Using the H-matrix from the Kc=5 NMF decomposition, we performed consensus hierarchical clustering repeated 1000 times by randomly sampling with replacement 80% of the samples for each round of clustering, using one minus the information coefficient as distance metric and Ward's linkage. A 2-dimensional representation of the 5-dimensional NMF H-matrix values for each sample was generated, with samples regionally designated by their cluster membership using the Onco-GPS functions 'GPSmap' and 'set_sample_phenotypes'. Since the CCC was higher with Kc=2 components, we also evaluated this solution but found that the while the clustering was stable, it did not give us enough resolution to identify genetically homogeneous groups, therefore we carried forward our 5 component

solution. Unsupervised and consensus hierarchical clustering were also performed on the 3% most variably methylated CpG tiles for comparison.

Differential Methylation Analysis and Genomic Annotation

Differential methylation analyses were performed on 25bp CpG tiles with ≥10x and <500x coverage in >60% of samples in each comparison group, with a minimum of 3 samples per group (R package 'MethylSig' version 0.4.4). Differentially methylated CpG tiles were defined as those with FDR corrected *p*-value <0.05 and a mean methylation difference >20% between comparison groups unless otherwise specified.

Analysis of Gerstung et al. Gene Expression Microarray Data

Gene expression datasets (Affymetrix GeneChip Human Genome U133 Plus 2.0 arrays) and metadata were downloaded from NCBI GEO with accession number GSE58831. Probe intensity values were normalized using the 'gcrma' R package from Bioconductor and normalized probe intensities were averaged for genes with more than one probe. Comparison groups for survival analysis in the Gerstung et al. data were defined based upon the average expression of the given genes (Gerstung et al. 2015). (46)

Analysis of TCGA LAML RNA-seq and Illumina 450k methylation Data

Pre-processed level 3 RNA-seq and Illumina 450k methylation array data were downloaded from the The Cancer Genome Atlas data portal (https://portal.gdc.cancer.gov). See figure legends for supplemental figures S14 and S16 for specific analyses performed.

Statistics

Differences in Kaplan Meier survival curves were assessed using the log-rank test. Univariate and multivariate models of OS were constructed using Cox proportional hazards regression. Variables with univariate *p*-values <0.2 were evaluated for inclusion in the final multivariable model and the final model was constructed using a forward and backward stepwise procedure optimized by the Aikake Information Criterion (AIC). Variables with multiple categories (such as IPSS-R categories) were assessed for inclusion in the final multivariable model based on the full variable. An FDR corrected Fisher's Exact test was used for all tests of enrichment including genetic lesions in specific clusters, DMRs within RefSeq genes and chromatin state segments, and gene ontology (GO) terms. GO term enrichment analysis was performed with R package “clusterProfiler” version 3.6.0. Associations between NMF component amplitudes and specific genetic lesions were tested via the information coefficient and an empirical permutation test (n=100,000 permutations per comparison) to determine statistical significance of the association (Kim et al. 2016).

Data Sharing Statement

The processed methylation data files have been deposited to the NCBI Gene Expression Omnibus under the accession GSE129828.

2.6 Acknowledgments

Chapter 2, in part, has been submitted for publication of the material as it may appear in *Blood Advances*, 2019, Reilly, Brian; Tanaka, Tiffany N.; Diep, Dinh;

Yeerna, Huwate; Tamayo, Pablo; Zhang, Kun; Bejar, Rafael. “DNA Methylation Identifies Genetically and Prognostically Distinct Subtypes of Myelodysplastic Syndromes”. The dissertation author was one of two primary investigators and authors of this paper. We would like to thank Donna Neuberg for her review and suggestions of the prognostic modeling carried out in this chapter. This work was supported by NIH grants (U01-CA217885. P.T. and H.Y., R01-HG009285, P.T., P30-CA023100, P.T. and H.Y.) and Evans Foundation grants (R.B.). This work was also supported by the American Foundation for Pharmaceutical Education (B.R.).

2.7 Supplementary Figures and Tables

Supplementary Table 2.1 Patient characteristics stratified by treatment group.

	Untreated	Treated	Combined	*P* value
Number of cases, N (%)	93 (66)	48 (34)	141	
Sex (male/female), N	72/21	31/17	103/73	0.11
Age (y), median (range)	72.3 (48-96.6)	71.0 (47-87)	72 (47-96.6)	0.32
FAB, N (%)				
RA	40 (43)	7 (15)	47 (33)	<0.01
RARS	18 (22)	9 (19)	27 (19)	1
RAEB/RAEBT	35 (38)	18 (38)	53 (38)	1
RCUD/MD	0 (0)	4 (8)	4 (3)	0.01
CMML	0 (0)	7 (15)	7 (5)	<0.01
Other	0 (0)	3 (6)	3 (2)	0.04
Cytogenetics*, N (%)				
Normal	66 (71)	23 (48)	89 (63)	0.01
Abnormal, not complex	20 (22)	16 (33)	41 (29)	0.19
Complex	7 (7)	4 (8)	11 (8)	1
Blood Counts, median (range)				
WBC ($x10^9$/L)	4.1(0.9–63)	3.1 (0.9–95.2)	3.7 (0.9–95.2)	0.94
ANC ($x10^9$/L)	2.1 (0.2–28)	1.0 (0.1–21.9)	1.6 (0.1–28)	0.4
Hgb (g/dL)	9.9 (6.9–15.2)	9.2 (5.8–13.4)	9.7 (5.8–15.2)	0.05
Platelet ($x10^9$/L)	108 (11–987)	99 (6–902)	102 (6–987)	0.14
BM Blast (%)†	0 (0–28)	6 (1–37)	1 (0–37)	0.01
IPSS-R‡, N (%)				
Very Low	22 (24)	1 (2)	23 (16)	<0.01
Low	27 (29)	4 (8)	31 (22)	<0.01
Intermediate	13 (14)	7 (15)	20 (14)	1
High	16 (17)	8 (17)	24 (17)	1
Very High	8 (9)	6 (13)	14 (10)	0.55
Treatment, N (%)				
Azacitidine	–	20 (42)	–	–
Decitabine§	–	28 (58)	–	–

*5 patients in the treatment group had incomplete cytogenetic information. † 20 patients were missing data for bone marrow blast % (all in treatment group) and were excluded from survival modeling. ‡ 29 patients were missing data (7 from the untreated group, 22 from the treated group) for one or more variables in the IPSS-R and could not be classified.§ One patient stopped treatment after 1 cycle of decitabine

Supplementary Table 2.2 Enrichment odds ratios for specific genetic abnormalities within epigenetically defined clusters.

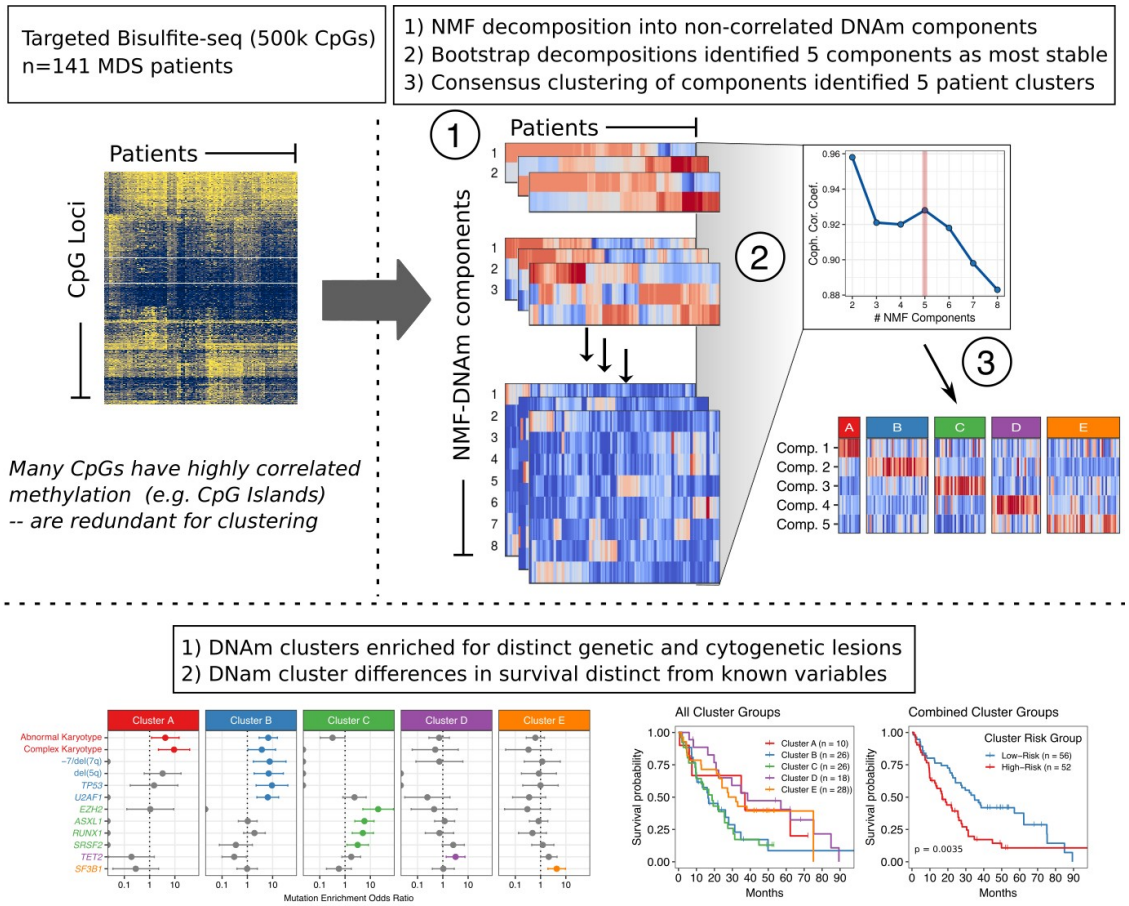
	Cluster A	Cluster B	Cluster C	Cluster D	Cluster E
TET2	0.2 (0.0-1.5)	0.3 (0.1-0.9)	1.8 (0.8-4.1)	3.2 (1.4-7.5) *	2.1 (0.99-4.5)
ASXL1	0 (0-0)	1.0 (0.4-2.4)	5.8 (2.4-14) ***	1.2 (0.5-2.9)	0.8 (0.4-1.9)
DNMT3A	1.7 (0.3-8.3)	1.3 (0.4-4.2)	0.29 (0.0-2.3)	1.7 (0.5-5.5)	2.0 (0.7-5.8)
ETV6	5.4 (0.9-31.2)	1.7 (0.3-9.0)	0.8 (0.1-7.2)	0.9 (0.1-7.4)	0.6 (0.1-4.9)
EZH2	1.1 (0.1-9.0)	0 (0-0)	20.4 (5.1-82) ***	0.5 (0.1-3.7)	0.3 (0.0-2.4)
NRAS	0 (0-0)	1.7 (0.3-9.0)	4.0 (0.9-19.1)	2.1 (0.4-11.7)	0 (0-0)
RUNX1	0 (0-0)	1.9 (0.7-5.1)	5.0 (1.9-13.2) **	0.7 (0.2-2.7)	0.5 (0.1-1.7)
SF3B1	0.3 (0.0-2.3)	1.0 (0.4-2.5)	0.6 (0.2-1.8)	1.0 (0.4-2.8)	4.4 (2.0-9.7) **
SRSF2	0 (0-0)	0.4 (0.1-1.6)	3.2 (1.2-8.4)	2.6 (1.0-7.1)	1.2 (0.5-3.4)
TP53	1.5 (0.2-13.1)	9.4 (2.2-40) **	0 (0-0)	0 (0-0)	1.0 (0.2-5.0)
U2AF1	0 (0-0)	6.4 (2.4-17.3) ***	2.4 (0.8-6.8)	0.2 (0.0-1.9)	0.4 (0.1-1.6)
Abnormal Karyotype	4.2 (1.2-14.7)	6.7 (2.9-15.5) ***	0.3 (0.1-0.97)	0.7 (0.3-1.8)	0.6 (0.3-1.4)
Complex Karyotype	9.3 (2.2-38.5) *	3.7 (1.1-13.1)	0 (0-0)	0.5 (0.1-4.1)	0.3 (0.0-2.7)
-7/del(7q)	0 (0-0)	7.6 (1.7-33.8) *	0 (0-0)	0.7 (0.1-6.2)	1.2 (0.2-6.0)
del(5q)	3.3 (0.6-17.5)	7.0 (1.9-26.5) *	0 (0-0)	0 (0-0)	0.9 (0.2-4.2)
Trisomy 8	1.5 (0.2-13.1)	1.2 (0.2-5.9)	2.6 (0.6-11.3)	0.6 (0.1-5.3)	1.0 (0.2-5.0)

Fisher's exact test FDR corrected p-value for enrichment or depletion in each cluster:
* FDR < 0.1; ** FDR < 0.05, *** FDR<0.01

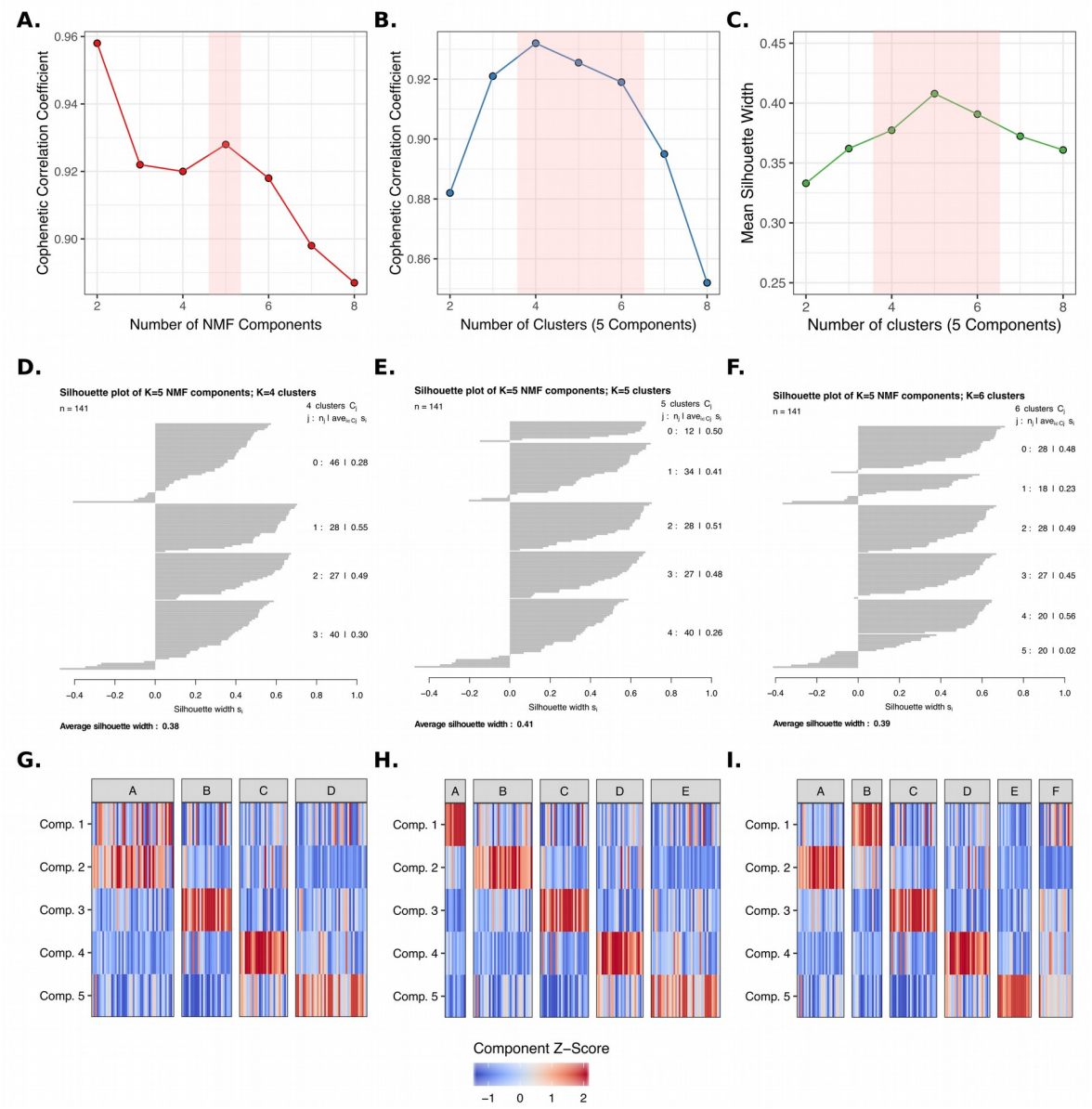
Supplementary Table 2.3 Univariate Cox proportional hazards regression on clinical and genetic features

	N (%)	HR [95% CI]	p-value
Cluster Risk Group			
High vs. Low	52 (48)	1.95 [1.24-3.08]	0.004
Clinical Features			
Age, >= 70 vs. Age < 70	68 (63)	1.29 [0.81-2.04]	0.285
Sex, M vs. F	79 (73)	1.21 [0.72-2.04]	0.463
IPSS-R			0.062
Very-High vs. Very-Low	14 (13)	3.15 [1.40-7.08]	0.005
High vs. Very-Low	24 (22)	1.83 [0.89-3.79]	0.102
Intermediate vs. Very-Low	18 (17)	1.70 [0.81-3.59]	0.163
Low vs. Very-Low	31 (29)	1.36 [0.69-2.67]	0.371
Karyotype			0.13
Abnormal (Not Complex) vs. Normal	33 (31)	1.35 [0.82-2.24]	0.24
Complex vs. Normal	9 (8)	2.20 [0.93-5.20]	0.071
Bone Marrow Blasts (%)			0.014
11 - 30 vs. < 5	23 (21)	1.77 [1.01-3.09]	0.046
5 - 10 vs. < 5	23 (21)	2.14 [1.22-3.77]	0.008
Hemoglobin (g/dL)			0.125
< 8.0 vs. >= 12	15 (14)	1.03 [0.39-2.71]	0.958
>= 8.0 - 9.9 vs. >= 12	46 (43)	2.03 [0.93-4.41]	0.075
>= 10.0 - 11.9 vs. >= 12	34 (31)	1.42 [0.64-3.13]	0.387
Platelets (x10^9/L)			0.03
< 50 vs. >= 450	30 (28)	3.03 [1.15-8.00]	0.025
>= 50 - 99 vs. >= 450	22 (20)	1.65 [0.60-4.56]	0.335
>= 100 - 149 vs. >= 450	14 (13)	1.44 [0.48-4.31]	0.514
>= 150 - 449 vs. >= 450	33 (31)	1.38 [0.52-3.64]	0.52
Absolute Neutrophil Count (x10^9/L)			0.737
< 0.5 vs. >= 10	10 (9)	0.65 [0.19-2.22]	0.49
0.5 - 1.79 vs. >= 10	46 (43)	0.56 [0.20-1.59]	0.278
1.8 - 9.99 vs. >= 10	47 (44)	0.58 [0.21-1.65]	0.31
Somatic Mutations (Mutated vs. Non-mutated)			
TET2	34 (31)	0.71 [0.43-1.16]	0.171
ASXL1	30 (28)	1.75 [1.06-2.88]	0.028
SF3B1	27 (25)	0.95 [0.58-1.57]	0.855
RUNX1	19 (18)	2.12 [1.22-3.68]	0.008
SRSF2	18 (17)	1.08 [0.61-1.90]	0.792
U2AF1	15 (14)	1.17 [0.63-2.19]	0.618
DNMT3A	13 (12)	1.00 [0.50-2.00]	0.991
EZH2	9 (8)	2.91 [1.37-6.18]	0.005
TP53	9 (8)	3.16 [1.41-7.08]	0.005
ETV6	5 (5)	2.29 [0.83-6.36]	0.111
NRAS	5 (5)	1.12 [0.41-3.07]	0.827
CBL	3 (3)	1.74 [0.42-7.13]	0.442
IDH2	2 (2)	1.46 [0.36-5.97]	0.601
KRAS	2 (2)	2.04 [0.50-8.38]	0.321
NPM1	4 (4)	0.95 [0.30-3.03]	0.933

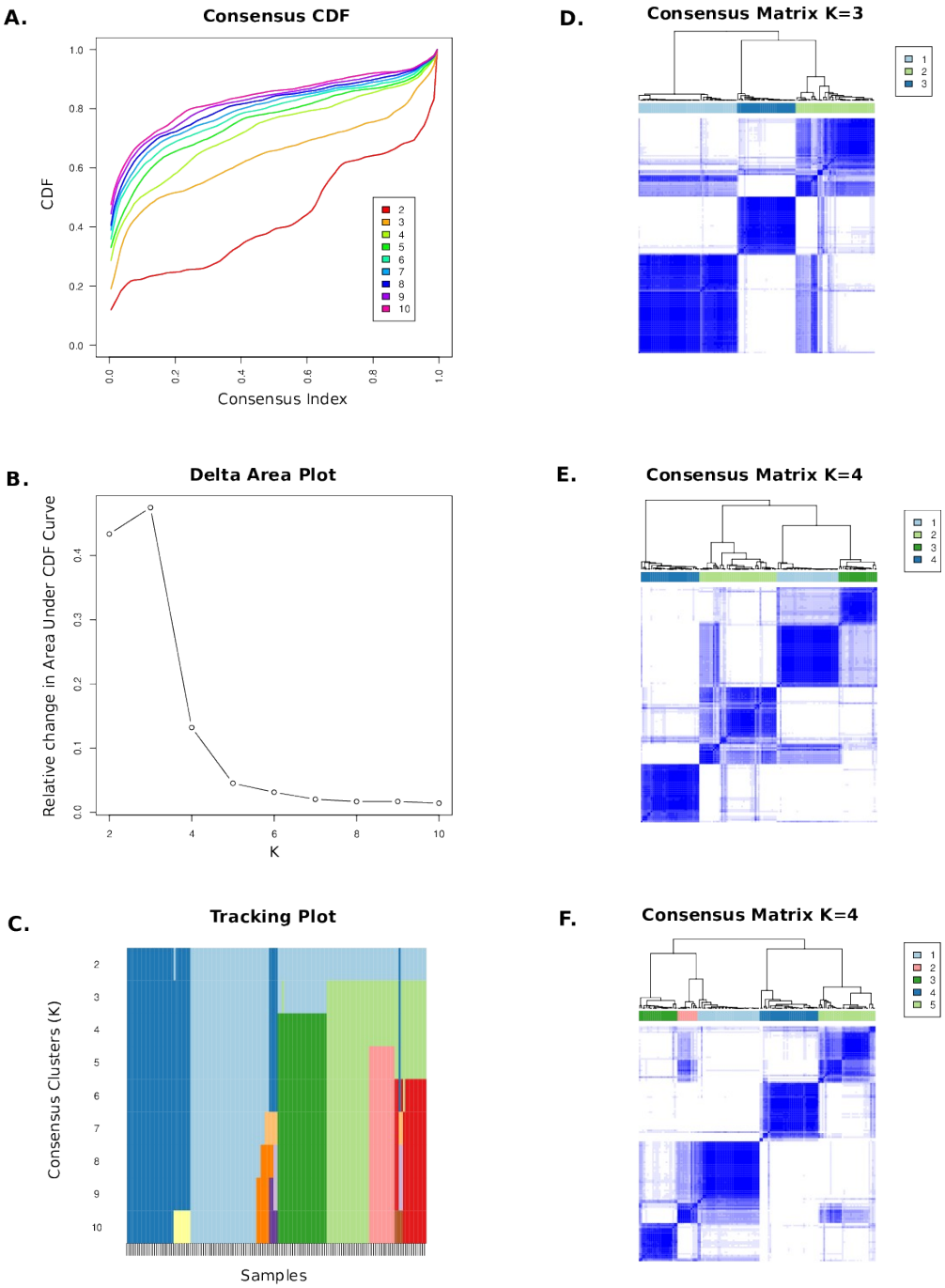
*P-values for individual categories within variables were calculated using the Wald test. P-values for full variables correspond to a log-rank test.



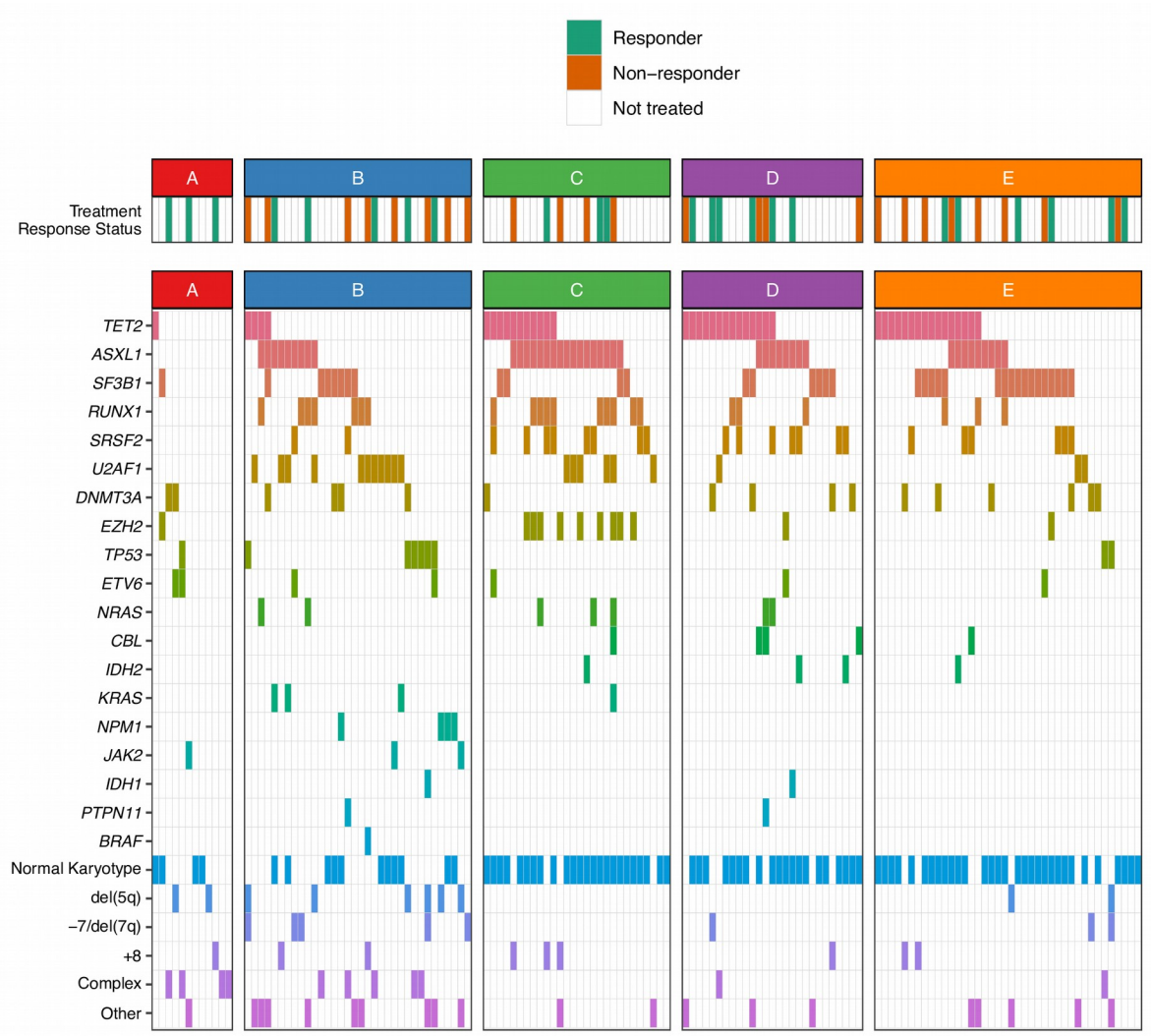
Supplementary Figure 2.1 Study design and findings overview.



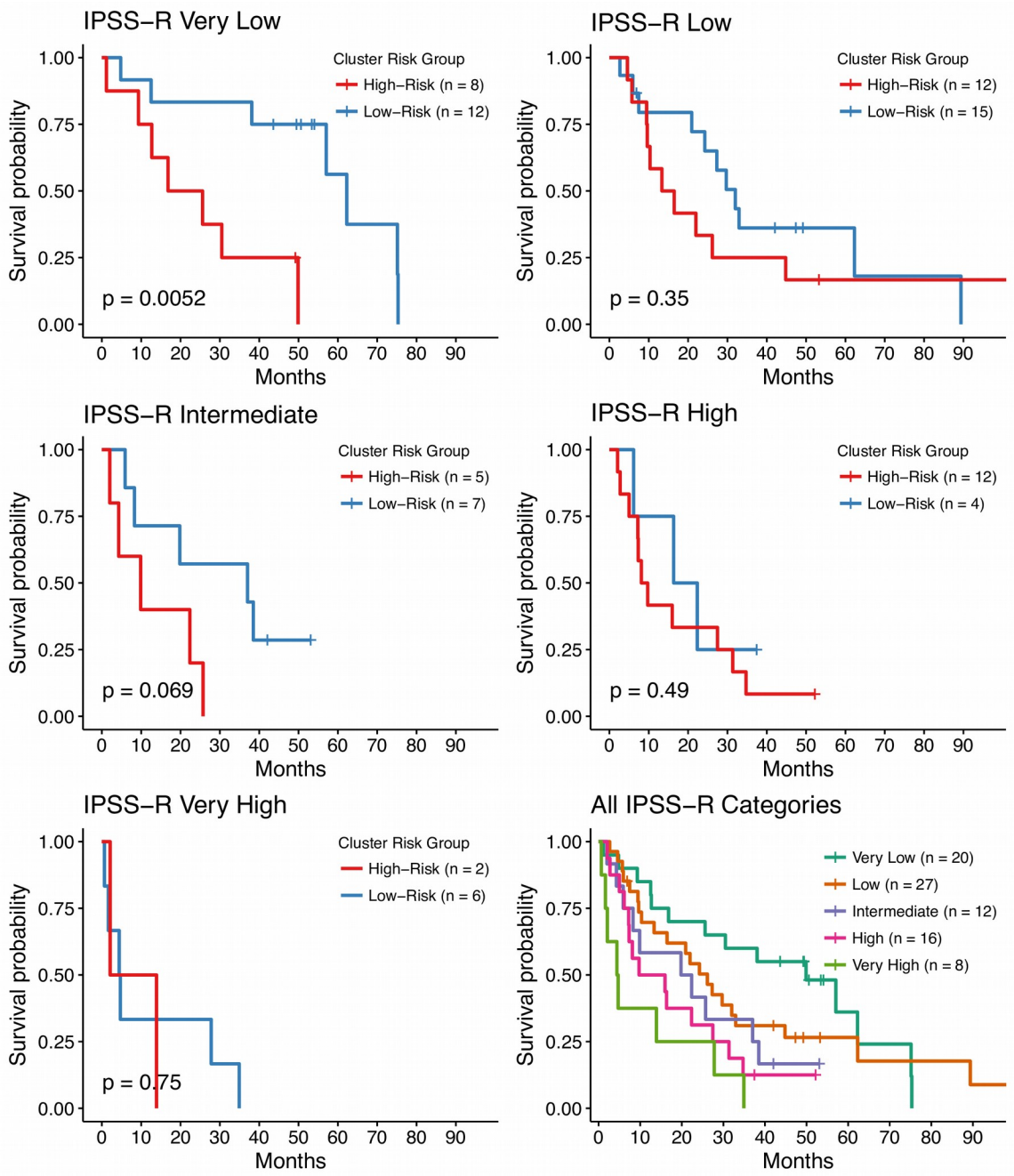
Supplementary Figure 2.2 Data-driven selection of the number of NMF components and methylation clusters. A) cophenetic correlation coefficient (measure of numerical stability of clustering result) for different numbers of NMF components. **B)** Cophenetic correlation coefficient for number of clusters after consensus clustering. **C)** Average silhouette width (measure of cluster cohesion and separation from neighboring clusters) for different numbers of consensus clusters. **D-F)** Silhouette width values for individual samples in clustering solutions with 4-6 consensus clusters. **G-I)** clustering heatmaps for the K=5 NMF component solution with 4-6 consensus clusters.



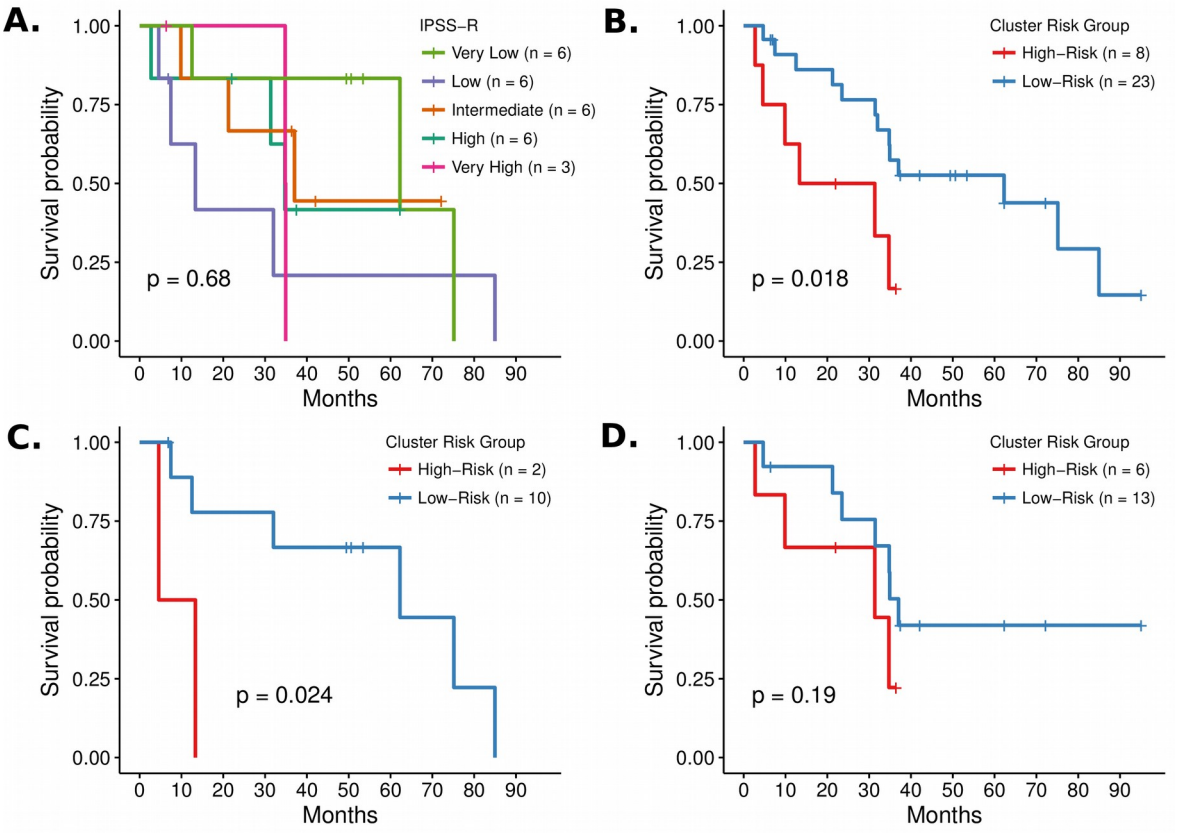
Supplementary Figure 2.3 Consensus clustering of raw methylation data is in agreement with a five cluster solution. The top 3% most variable CpG tiles were clustered using hierarchical clustering and Ward's agglomeration method, repeated 1,000 times with subsampling of 80% of CpG tiles to provide a measure of clustering stability. **A)** The Consensus CDF and **B)** Delta Area plots indicated a clustering solution close to 5 clusters was most stable, with greater than five clusters providing very little increase in the area under the CDF curve. **D-F)** This can also be seen in the consensus matrix heatmaps.



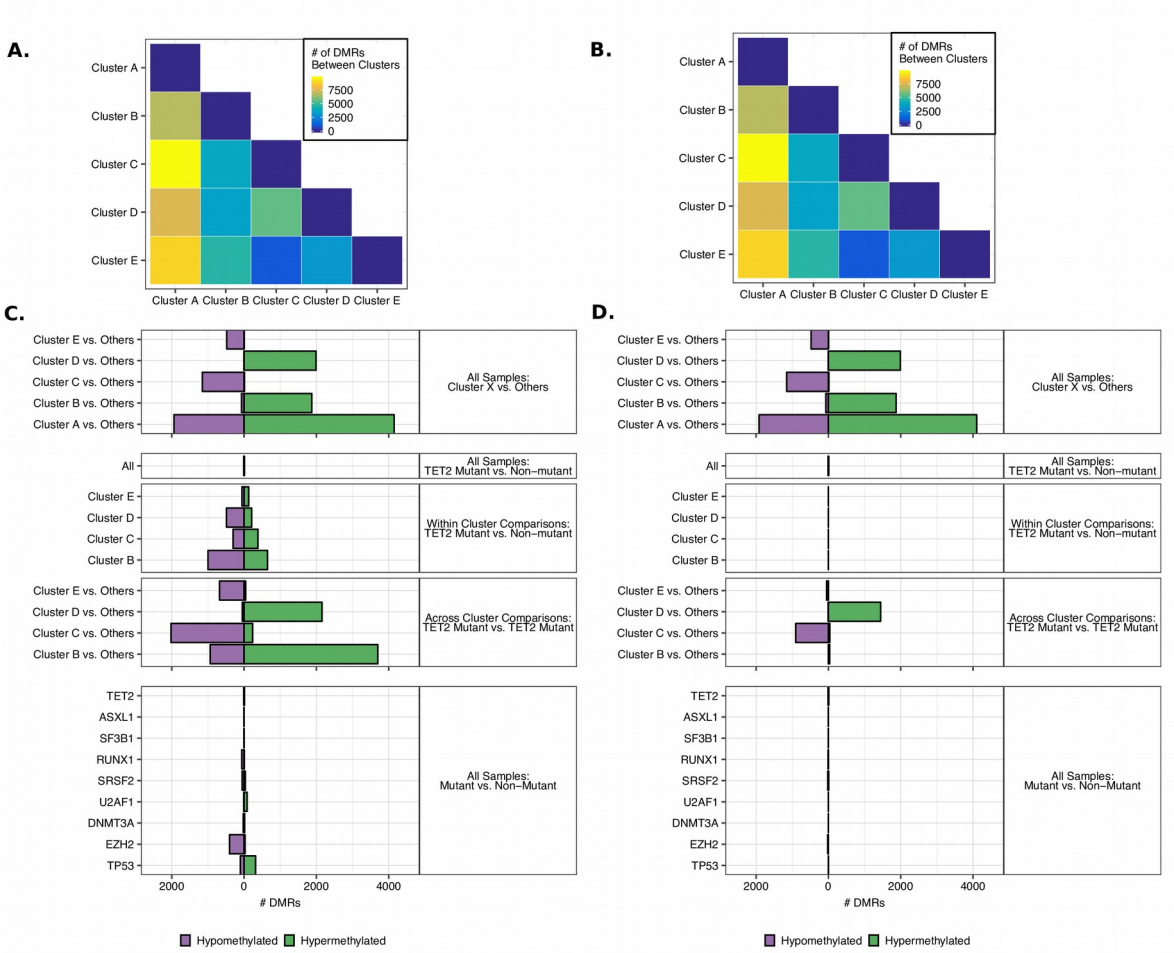
Supplementary Figure 2.4 Hypomethylating agent treatment response status and genetic abnormalities of the five DNA methylation clusters identified. Clusters are the same as shown in Figure 2.1C with the addition of the bar at the top indicating treatment response status. 48 patients received at least one cycle of HMA therapy, one patient stopped after one cycle, leaving 47 with treatment response outcomes.



Supplementary Figure 2.5 Survival of patients by methylation cluster risk group for each IPSS-R category. Including the HMA-treated patients, we stratified patients by their cluster risk group (high risk vs. low risk) for each IPSS-R category, and a significant difference in survival was seen between cluster risk groups among patients classified as very low risk by the IPSS-R. A similar although non-significant trend was seen in all other IPSS-R risk categories.

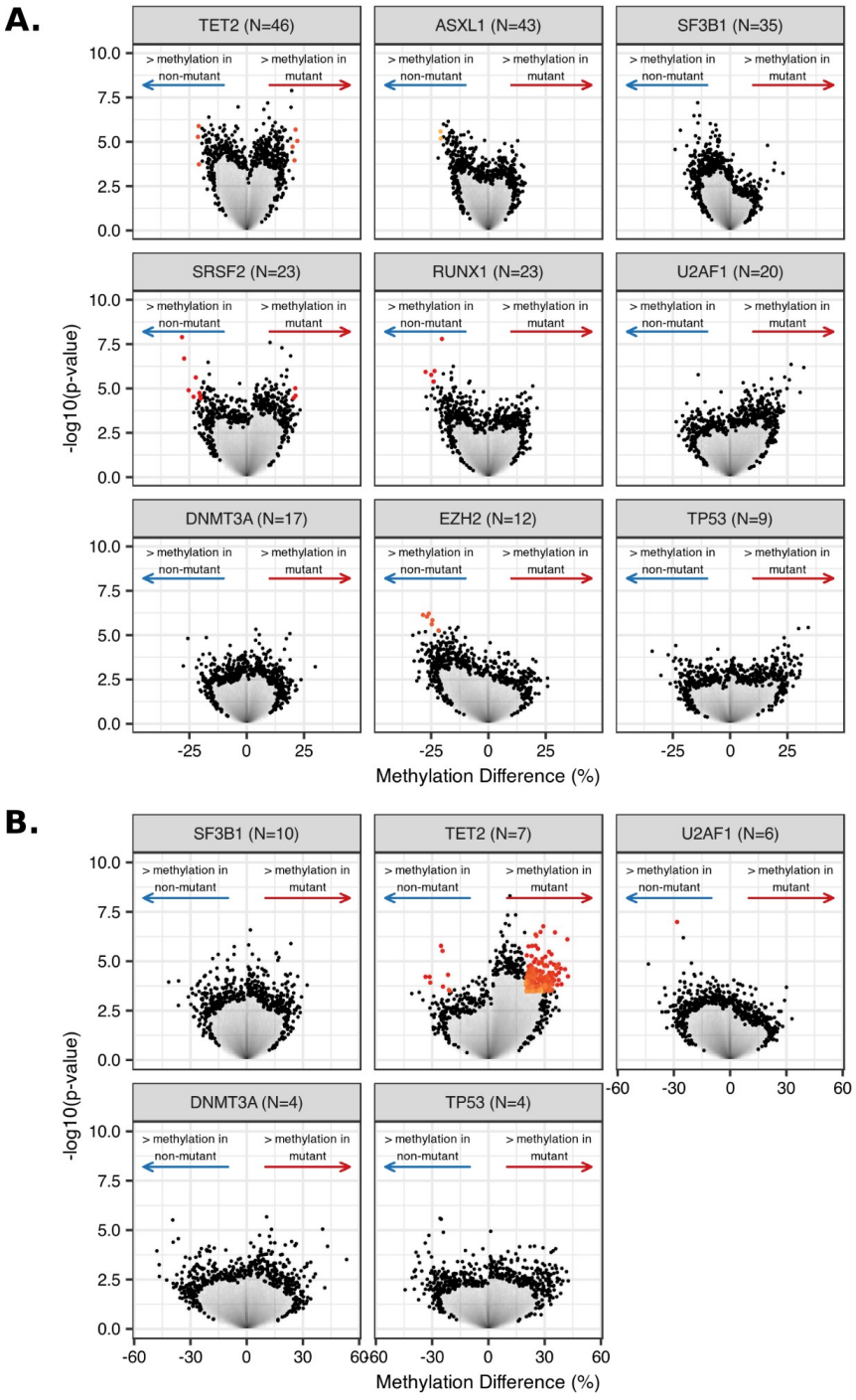


Supplementary Figure 2.6 Methylation cluster risk group informs prognostic risk even in patients lacking any known prognosis associated somatic mutations. Survival of patients lacking somatic mutations in genes associated with prognosis in MDS (*TP53*, *ASXL1*, *RUNX1*, *U2AF1*, *CBL*, *EZH2*, *SF3B1*, *SRSF2* and *IDH2*), stratified by **A)** IPSS-R category, **B)** Epigenetic cluster risk group. **C)** IPSS-R Low and very low risk patients and **D)** IPSS-R intermediate, high, and very high patients stratified by methylation cluster risk group.

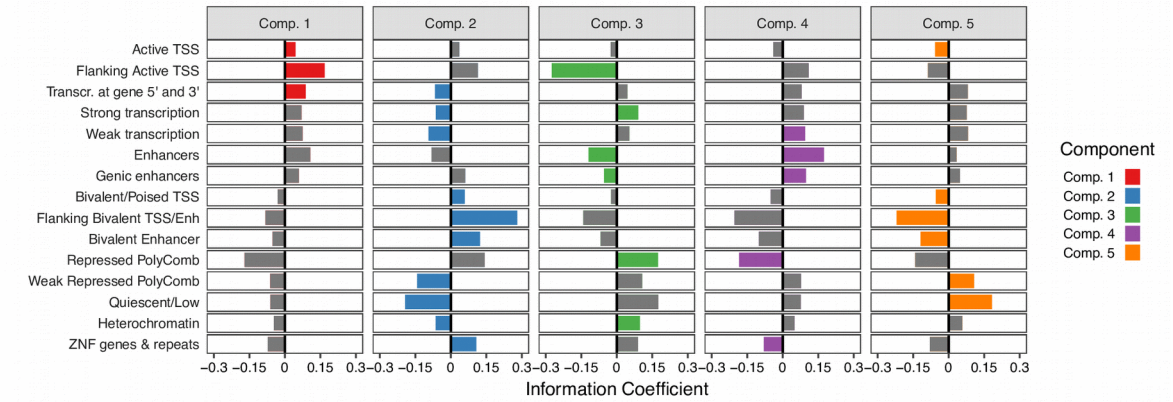


Supplementary Figure 2.7 Global differences in differential methylation by cluster and for select genes.

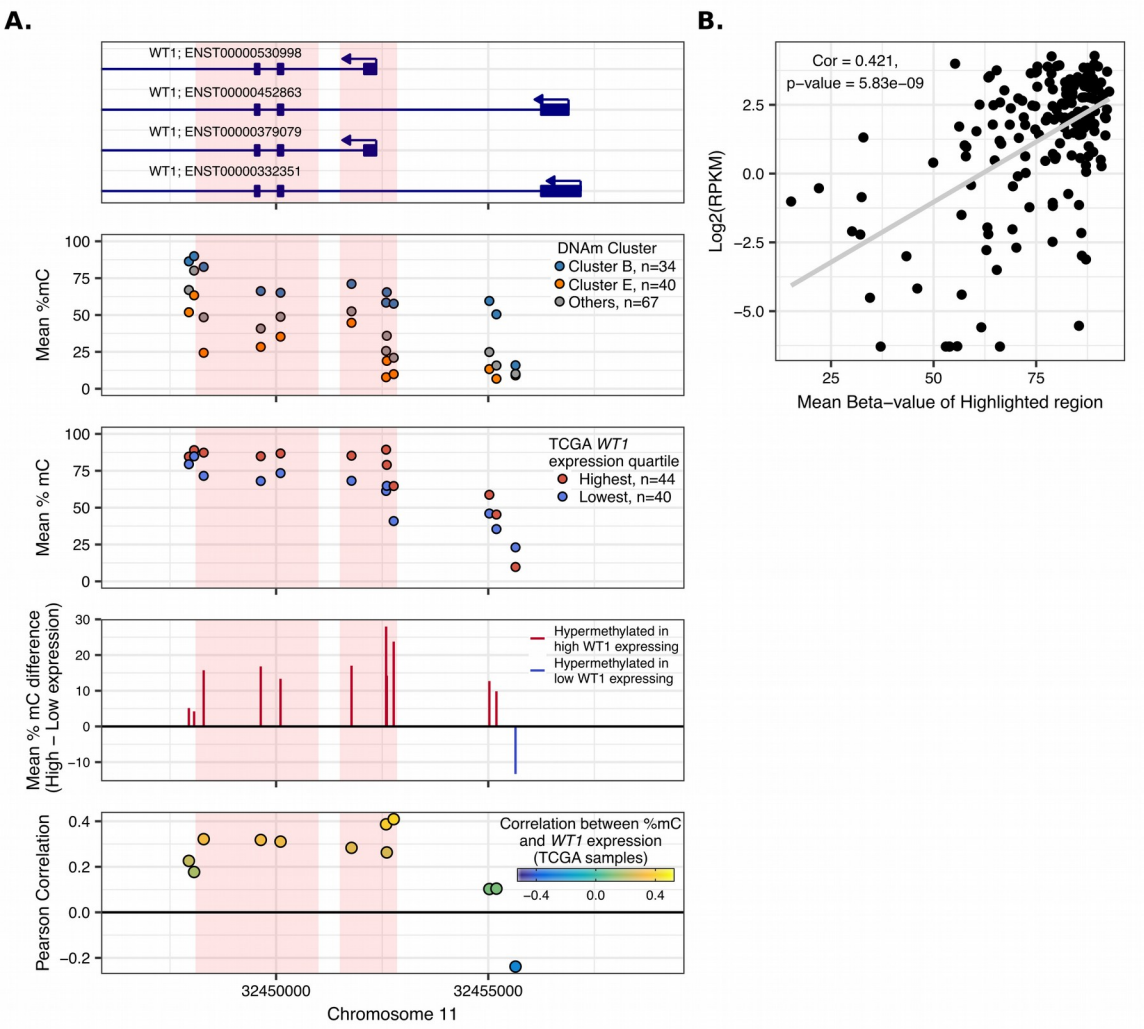
(A/B) Pairwise differential methylation heatmap between clusters reveals that Cluster A is most dissimilar compared with other clusters. **(C/D)** Global differential methylation patterns for Cluster-specific DMRs (top), comparisons across clusters within *TET2* mutant/non-mutant groups where indicated (middle, 3 panels), comparisons based on presence/absence of indicated somatic mutations. DMRs in (A) and (C) defined with a P-value cutoff of 0.05 and methylation difference of >20%; DMRs in (B) and (D) defined with a Q-value cutoff (FDR-corrected, more stringent) of 0.05 and methylation difference >20%. Notice that the number and direction of cluster-specific differentially methylated regions is fairly similar when limiting comparisons to *TET2* mutant patients belonging to different clusters, indicating that cluster-membership is the major driver of methylation differences as opposed to mutations in *TET2*.



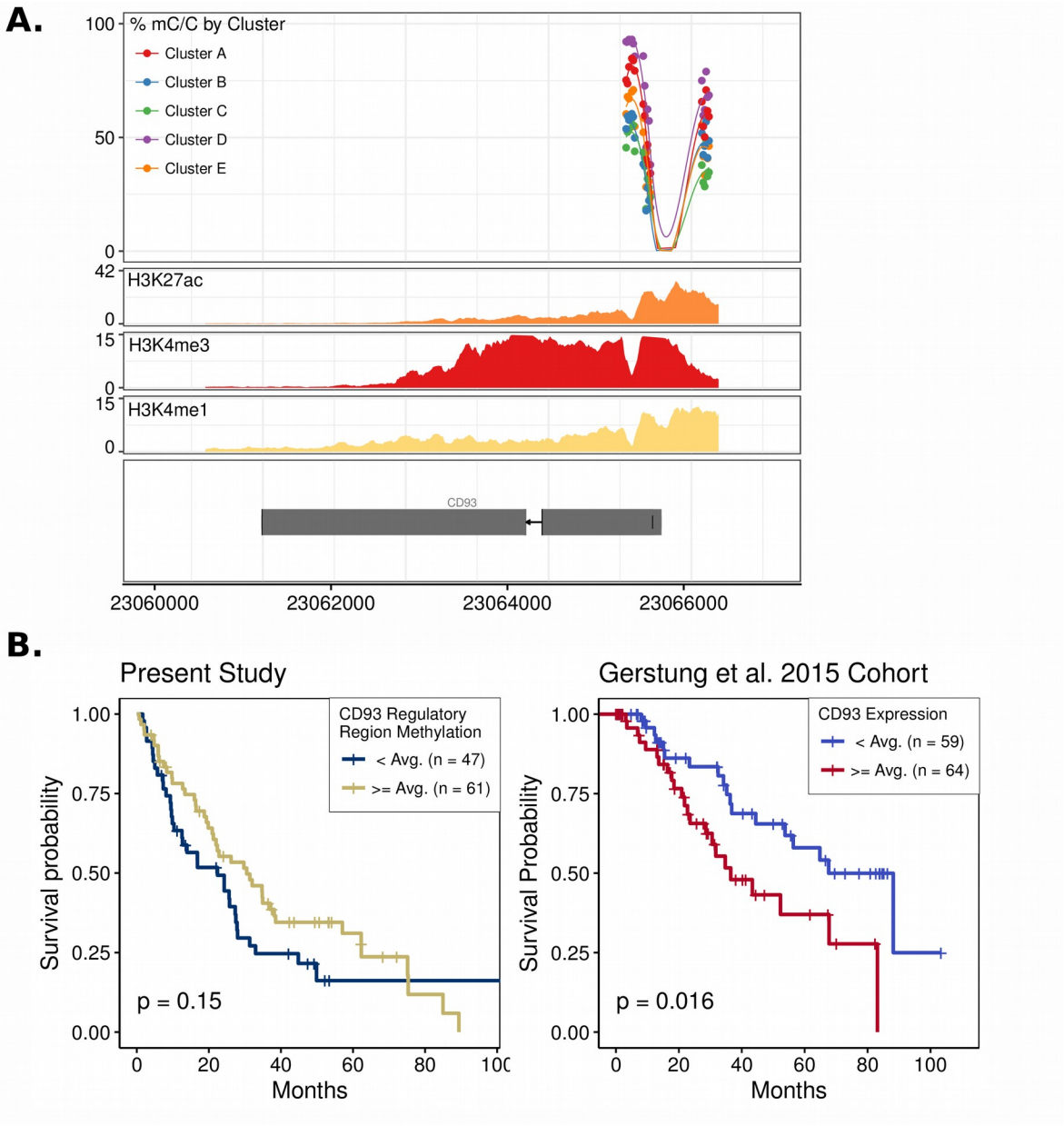
Supplementary Figure 2.8 Differentially methylated CpG-tiles between groups of patients based on presence or absence of somatic mutations. A) Volcano plots for comparisons between ALL patients harboring a somatic mutation in the indicated gene vs. patients that are wild-type for that gene. **B)** Volcano plots for comparisons between patients whose ONLY detected mutation is in the indicated gene vs. patients who are wild-type for that gene. Note: only genes with >3 patients in each group were selected for comparison.



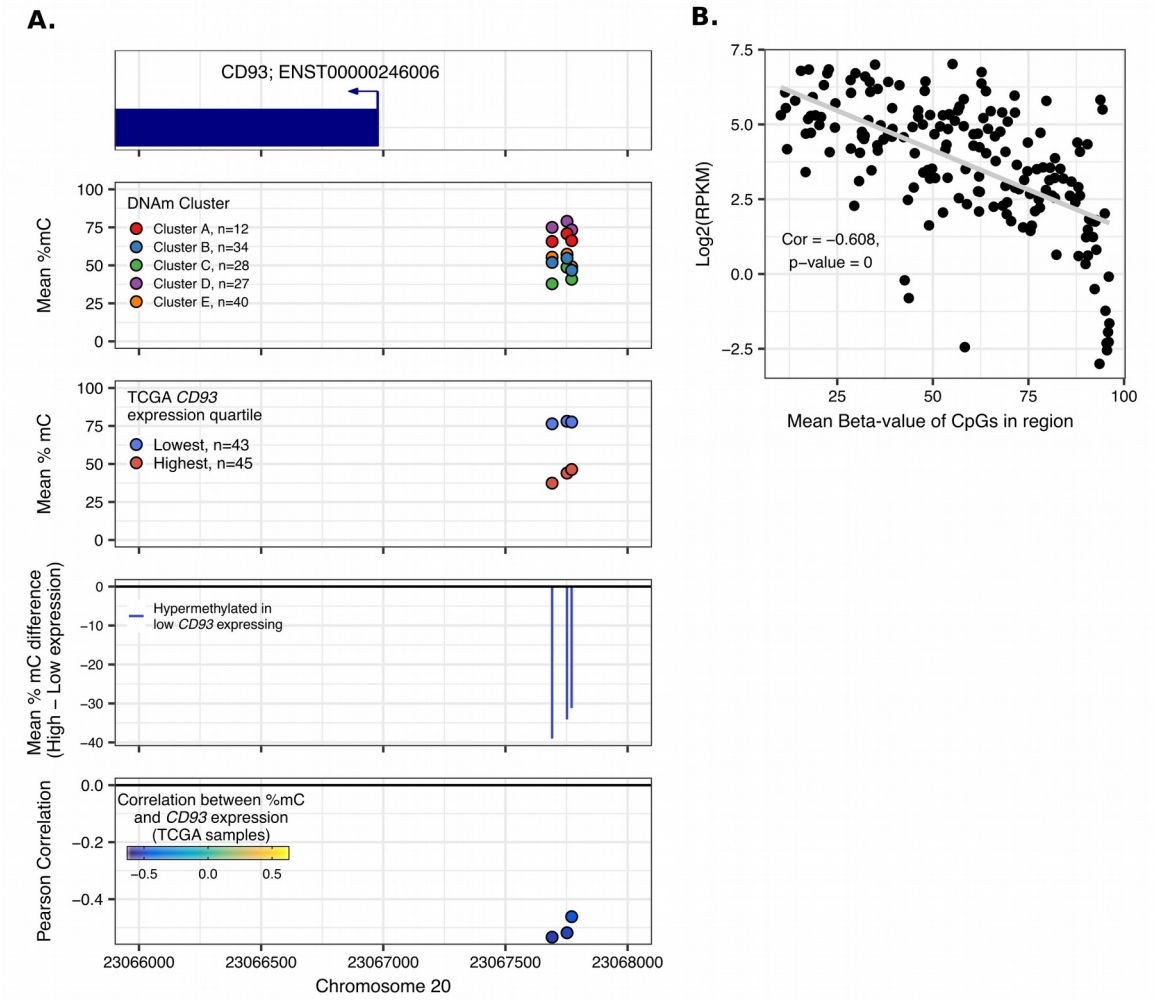
Supplementary Figure 2.9 Individual NMF component amplitudes are associated with distinct regulatory chromatin states in reference CD34+ cell epigenome. NMF component amplitudes show distinct patterns of association with epigenetic regulatory segments in a reference hematopoietic stem cell (HSC) as defined by the chromHMM 15-state genome segmentation model created from integrated epigenetic datasets (Roadmap Consortium). NMF component associations with given genome segments as represented by the information coefficient value on the x axis (the most strongly positively and negatively associated components are colored for each regulatory segment; i.e. two colored bars per segment).(TSS = transcription start site; Enh = Enhancer; Transcr. = Transcribed)



Supplementary Figure 2.10 Correlating *WT1* methylation and expression using TCGA LAML cohort data. Note: only CpGs with coverage in both TCGA and our data are shown. **A)** (top) transcript model plot of the WT1 gene zoomed in to focus on CpGs with large differences in our cohort; (middle-top) mean % methylation of patients in DNAm cluster groups for CpGs overlapping the region; (middle-middle) mean % methylation of patients in TCGA LAML cohort for each CpG loci overlapping region; (middle-bottom) difference in % methylation between highest expressing and lowest expressing quartiles of TCGA patients for WT1; (bottom) correlation between WT1 expression and % methylation across all TCGA patients with paired RNAseq expression and Illumina 450k methylation datasets for each CpG overlapping region. **B)** Mean % methylation of all CpGs located in red-highlighted region in (A) vs. log2 WT1 expression in RPKM for each TCGA patient (each point represents one patient). Spearman correlation of the % methylation vs. expression and significance of correlation indicated in plot.



Supplementary Figure 2.11 Survival by *CD93* methylation (this study) and expression (Gerstung et al. cohort). A) methylation of cluster-specific differentially methylated CpG tiles within the *CD93* gene with loess smoothened linear model (top panel). H3K27ac, H3K4me3, and H3K4me1 Chip-seq signal tracks correspond to a reference dataset of GM-CSF mobilized CD34+ cells from a healthy male donor (2nd through 4th panels; Roadmap consortium, epigenome E051). Gene model track correspond to isoforms of *CD93* found in UCSC knownGene table for hg19 (bottom panel). **B)** Kaplan-Meier curves for the present study stratified by average *CD93* regulatory region (+/-2500bp from TSS) DNA methylation (left) and for the Gerstung et al. 2015 cohort stratified by average *CD93* expression (right).



Supplementary Figure 2.12 Correlating *CD93* methylation and expression using TCGA LAML cohort data. Note: only CpGs with coverage in both TCGA and our data are shown. **A)** (top) transcript model plot of the *CD93* gene zoomed in to focus on CpGs that are covered in both our cohort and the TCGA LAML cohort; (middle-top) mean % methylation of patients in DNAm cluster groups for CpGs overlapping the region; (middle-middle) mean % methylation of patients in TCGA LAML cohort for each CpG loci overlapping region; (middle-bottom) difference in % methylation between highest expressing and lowest expressing quartiles of TCGA patients for *CD93*; (bottom) correlation between *CD93* expression and % methylation across all TCGA patients with paired RNAseq expression and Illumina 450k methylation datasets for each CpG overlapping region. **B)** Mean % methylation of all (3) CpGs located in region in (A) vs. log2 *CD93* expression in RPKM for each TCGA patient (each point represents one patient). Spearman correlation of the % methylation vs. expression and significance of correlation indicated in plot.

Chapter 3: *TET2* Dependent Effects of 5-Azacitidine on Gene Expression and DNA methylation Dynamics

3.1 Abstract

Myelodysplastic syndromes (MDS) are a group of myeloid malignancies characterized by clonally impaired hematopoiesis and an increased propensity for the development of acute myeloid leukemia (AML). DNA methyltransferase inhibitors (DNMTIs) like 5-Azacitidine (5-Aza) are the only drugs approved for the treatment of higher-risk MDS, however less than 50% of patients respond, and there no reliable predictors of response. Somatic mutations in the DNA methylation regulating gene *tet-methylcytosine dioxygenase 2* (*TET2*) have been shown to be associated with response to DNMTIs in some patients, however the mechanisms responsible for this association remain unknown. Using bisulfite padlock probes, mRNA sequencing, and hydroxymethylcytosine pull-down sequencing, we show that *TET2* loss particularly influences DNA methylation (5mC) and hydroxymethylation (5hmC) patterns at erythroid gene enhancers, and is associated with down-regulation of erythroid gene expression in the human erythroleukemia cell line TF-1. We go on to show that 5-Aza disproportionately induces expression of these down regulated genes in TET2KO cells and we show that this effect may be related to 5mC changes at erythroid gene enhancers after 5-Aza exposure. This work highlights the role of 5mC and 5hmC changes at distal regulatory elements in altering the expression of differentiation-associated gene signatures, and sheds light on how 5-Aza may be more effective in patients harboring *TET2* mutations.

3.2 Introduction

DNA methyltransferase inhibitors (DNMTIs) such as azacitidine (5-Aza) are the only class of drugs approved for the treatment of higher-risk myelodysplastic syndromes (MDS), and are increasingly gaining use in other cancers as monotherapy (J.-P. J. Issa et al. 2005; Kantarjian et al. 2012; Stahl et al. 2018), as well as in combination with other epigenetic drugs (Azad et al. 2017; Connolly et al. 2017; Juergens et al. 2011). DNMTIs are cytosine analogs that become incorporated into DNA where they serve as substrates for and covalently trap DNA methyltransferase enzymes (DNMTs), leading to their proteasomal degradation and causing global hypomethylation of the genome as cells divide (J. P. J. Issa and Kantarjian 2009). Despite wider usage of these drugs, overall response rates are poor, with less than 50% of MDS patients experiencing a clinical benefit, while the mechanisms which sensitize patients to these drugs remain elusive (Garcia-Manero 2008; Fenaux and Ades 2009). Previous work in our laboratory and others identified TET2 mutation as a biomarker of response in MDS patients (Bejar, Lord, et al. 2014b; Itzykson et al. 2011), where mutant patients were almost twice as likely to respond as their non-mutant counterparts, however the mechanisms behind this relationship remain unknown.

TET2 is a member of the tet methylcytosine dioxygenase enzyme family which catalyze the conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) and is one of the most frequently mutated genes in MDS(Tahiliani et al. 2009b; Bejar et al. 2011; Haferlach et al. 2014; Ito et al.

2010b). Tet proteins are the only known pathway for active demethylation of 5mC through further oxidation of 5hmC into 5-carboxycytosine which is recognizable by thymine-DNA glycosylase which remove and replace it with unmethylated cytosine(He et al. 2011b; Maiti and Drohat 2011b). Tet proteins serve diverse roles but are especially important in maintaining low methylation levels at transcriptional enhancers in hematopoietic cells (Kasper D Rasmussen et al. 2015), a finding corroborated by TET2-mutated patients exhibiting a hypermethylation phenotype at enhancers (Yamazaki et al. 2015). Methylation at enhancer elements is extremely dynamic during cellular differentiation (Schultz et al. 2015b; Rönnerblad et al. 2014), and it appears that tet proteins are essential for certain types of differentiation to occur properly (Sheaffer et al. 2014b; Hon et al. 2014).

There is some debate regarding the precise mechanisms leading to patient responses to DNMTIs, but at least in the context of TET2 mutant cases, the response does not appear to be driven by changes in the clonal burden of tumor, but rather via epigenetic induction of differentiation (Merlevede et al. 2016; Unnikrishnan et al. 2017b). The dynamics of DNA methylation during DNMTI treatment have been a subject of intense study as groups look for DNAm-based biomarkers of response to DNMTIs. Notably, both responding and non-responding patients exhibit similar levels of demethylation immediately following DNMTI treatment (Welch et al. 2016b), however only patients who respond exhibit long-term demethylation after several cycles (months) of therapy (Shen et al. 2010b;

Merlevede et al. 2016), suggesting that the rate of remethylation in these patients may be a significant factor determining response.

We hypothesized that the baseline as well as the kinetics of DNA methylation changes during 5-Aza treatment may be relevant for the increased response rates in *TET2*-mutant MDS patients, therefore we set out to determine the specific DNAm changes and the accompanying biological pathways affected during 5-Aza treatment. To accomplish this we employed CRISPR/Cas9-mediated inactivation of *TET2* and next-gen sequencing approaches to study 5mC, 5hmC and gene expression dynamics during 5-Aza exposure and recovery in a TF-1 erythroleukemia cell line model. We identified an altered sensitivity and length of recovery after 5-Aza exposure as well as differences in demethylation and remethylation rate in TET2KO cells. We go on to show that erythroid differentiation related signatures are down-regulated in TET2KO, how these processes are disproportionately effected in KO cells upon 5-Aza treatment, and how it relates to 5hmC and 5mC changes particularly at enhancers. Our findings highlight the subtleties of the relationship between DNA methylation deregulation and differentiation defects in TET2KO cells, and shed light on why TET2KO patients may benefit more from DNMTI therapy.

3.3 Results

Targeting the TET2 catalytic domain with CRISPR/Cas9 effectively abolishes TET2 expression and function in TF-1 cells.

As a model for studying the role of TET2 in response to 5-Aza we selected the TF-1 erythroleukemia cell line. TF-1 makes a suitable model due to its growth factor dependence and sensitivity to differentiating agents, as well as its lack of mutations in genes recurrently mutated in MDS (Barretina et al. 2012). We used paired CRISPR Cas9 guide-RNAs (gRNA) to target the key catalytic residues of the TET2 dioxygenase domain in TF-1 cells with the goal of deleting a ~1.2kb region of exon 11. To control for any potential off-target DNA damage we used separate gRNA sequences for each replicate as shown in Figure 3.1. After isolating single-cell colonies and PCR screening potential TET2KO clones, we identified one clone from each pair of CRISPR gRNAs that had bi-allelic inactivation of the TET2 catalytic domain. Notably, after PCR screening ~100 clones per gRNA pair, we never observed a clone with a bi-allelic 1.2kb deletion, likely due to the statistical improbability of this occurring in a single cell as mentioned in (Canver et al. 2014). Nonetheless, we validated at the protein level that there was no full-length TET2 protein present in either of our TET2KO TF-1 clones (Figure 3.1B). We observed some clone-specific differences in normal growth patterns as might be expected from the initial clonal variability of the bulk cell population (Figure 3.1C). To confirm that the catalytic function of TET2 had been impacted we performed 5hmC DNA immunoblot, however we did not have the sensitivity to detect significant differences in 5hmC content at baseline (Supplementary Figure 3.1). In order to enhance the 5hmC signal in our clones we took advantage of the recent findings that TET protein function can be enhanced

by increased L-ascorbic acid (L-AA) concentration (Agathocleous et al. 2017; Cimmino et al. 2017). L-AA treatment increased 5hmC content >2 fold in TET2 WT clones, while TET2KO clones had a more modest ~1.2 fold increase. Finally, to more sensitively detect changes in 5hmC and to determine which genomic regions were effected, we performed 5-hydroxymethylation pull-down followed by next-generation sequencing (HMCP-seq). The majority of 5hmC enriched regions (~51,000) were shared between TET2KO and WT cells. Of the 2,381 differentially hydroxymethylated regions (DHMR) shown in Figure 3.1D (unadjusted $p<0.05$) 1,861 (78%) had increased 5hmC in TET2WT, while 520 were increased in TET2KO. When examining only the most significantly different DHMRs (FDR < 0.1) 319/328 (97%) had increased 5hmC signal in TET2WT. Notably, the DHMR peaks were enriched for regions of active transcription and active enhancers and promoters in a reference hematopoietic stem cell epigenome (Figure 3.1F). Combined these results indicate that at baseline TET2 may not be the primary regulator of 5hmC in TF-1 cells, but its loss of function leads to reduced 5hmC signal at enhancer regions and impairs 5hmC deposition upon L-AA stimulation, which is in agreement with the known roles of TET2 (Kasper Dindler Rasmussen and Helin 2016; Kasper D Rasmussen et al. 2015; Hon et al. 2014; Yamazaki et al. 2015).

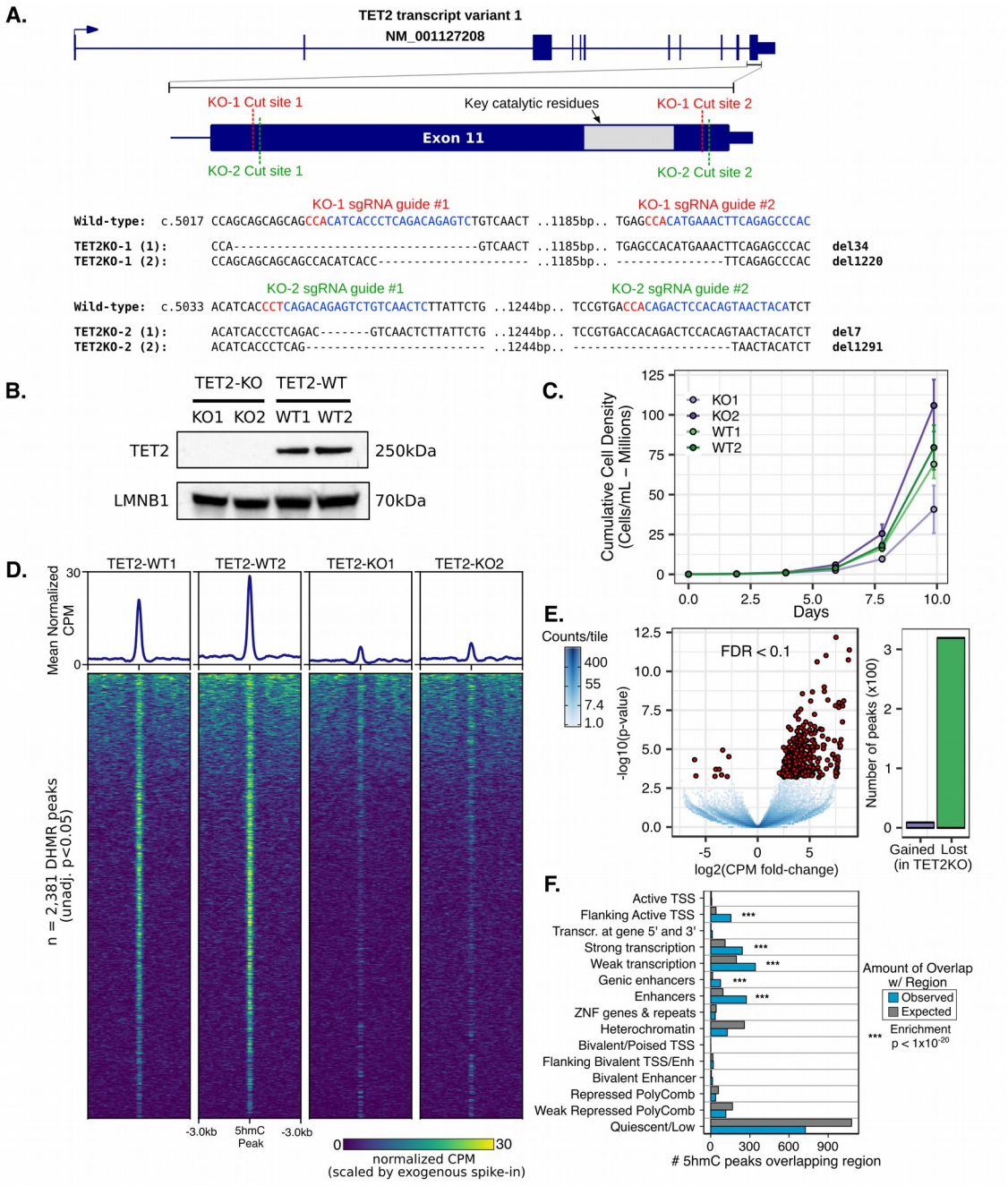


Figure 3.1 CRISPR/Cas9 mediated loss of TET2 alters growth and hydroxymethylation patterns in TF1. A) Paired CRISPR sgRNA design targeted key catalytic residues of TET2. Two independently derived TF1 clones with bi-allelic inactivating deletions were characterized. **B)** Western blot confirms loss of full-length TET2 protein. **C)** Growth curves for TF1 clones display differences in normal growth patterns for TET2 knockout. **D)** 5-hmC pull down and sequencing (HMCP) peak heatmaps and mean normalized counts profile for differentialy hydroxymethylated regions (DHMR). **E)** Differential 5hmC peak enrichment volcano plot TET2KO vs. WT. **F)** Number of DHMR (TET2KO v. WT) observed vs. expected overlapping regulatory chromatin states in reference CD34+ HSC epigenome (from reference epigenome E051 of Roadmap Epigenomics Consortium)

TET2KO and WT TF-1 exhibit differential sensitivity and remethylation kinetics after 5-Aza exposure.

To investigate the effects of 5-Aza in our cell lines we first had to determine the minimal dose that inhibited DNA methyltransferases (DNMTs) while maintaining high cellular viability to avoid convolution of DNAm differences due to clonal shifts versus true DNAm changes induced by 5-Aza. We observed that multiple lower intensity treatments of 5-Aza lead to prolonged down-regulation of DNMT1, while maintaining high cellular viability (Supplementary Figure 3.2), and settled on an optimal dosing regimen of 200nM 5-Aza every 24hrs for 3 days. With a 3 day dosage regimen we observed modest but significant increase in sensitivity to 5-Aza by TET2KO TF-1 cells (Figure 3.1A). Acute DNAm changes induced by 5-Aza are known to correlate with the population doubling time (PDT) in cancer cell lines (Yang et al. 2014), therefore we monitored the PDT after 5-Aza exposure as a surrogate measure of DNAm recovery. Notably, TET2KO clones had a significant increase in PDT compared with WT clones, however the length of time to recovery of normal PDT was similar between the two (~12 days) (Figure 3.2B). Interestingly, the time to recovery for the leukemia TF-1 cell line is much shorter than that of colon cancer cell lines reported in (Yang et al. 2014). After determining the optimal dosage regimen and time points for DNAm changes after 5-Aza, we used Bisulfite Padlock Probe sequencing (BSPP) to measure DNA methylation and RNAseq to measure gene expression dynamics during treatment and recovery from 5-Aza treatment (Figure 3.2C). Surprisingly, at baseline (pre-

treatment) there were almost an equal number of hyper and hypo-methylated CpGs when comparing TET2WT and KO (12,310 hyper- and 10,246 hypo-methylated) (Figure 3.2D). As expected, after 5-Aza exposure there was a global trend for hypomethylation in both genotypes. We next defined a set of CpGs termed “dynamic CpGs” for each genotype which had the largest differences in methylation between treatment time points (posterior probability >0.90 that difference was >20%) to focus on the most relevant %mC changes during treatment. Of the 621,000 CpGs covered in all samples in our dataset, only 9,787 met the dynamic CpG criteria, and the mean change in methylation for these CpGs was 31% in WT and 27% in KO (Figure 3.2E). Only 1,098 of the dynamic CpGs were shared between genotypes, while the majority were unique (WT 4,998; KO 3,691) (Figure 3.2F). Finally, when examining the global remethylation kinetics, TET2KO dynamic CpGs were both demethylated to a lesser extent and also remethylated at a lower speed than WT dynamic CpGs (Figure 3.2G).

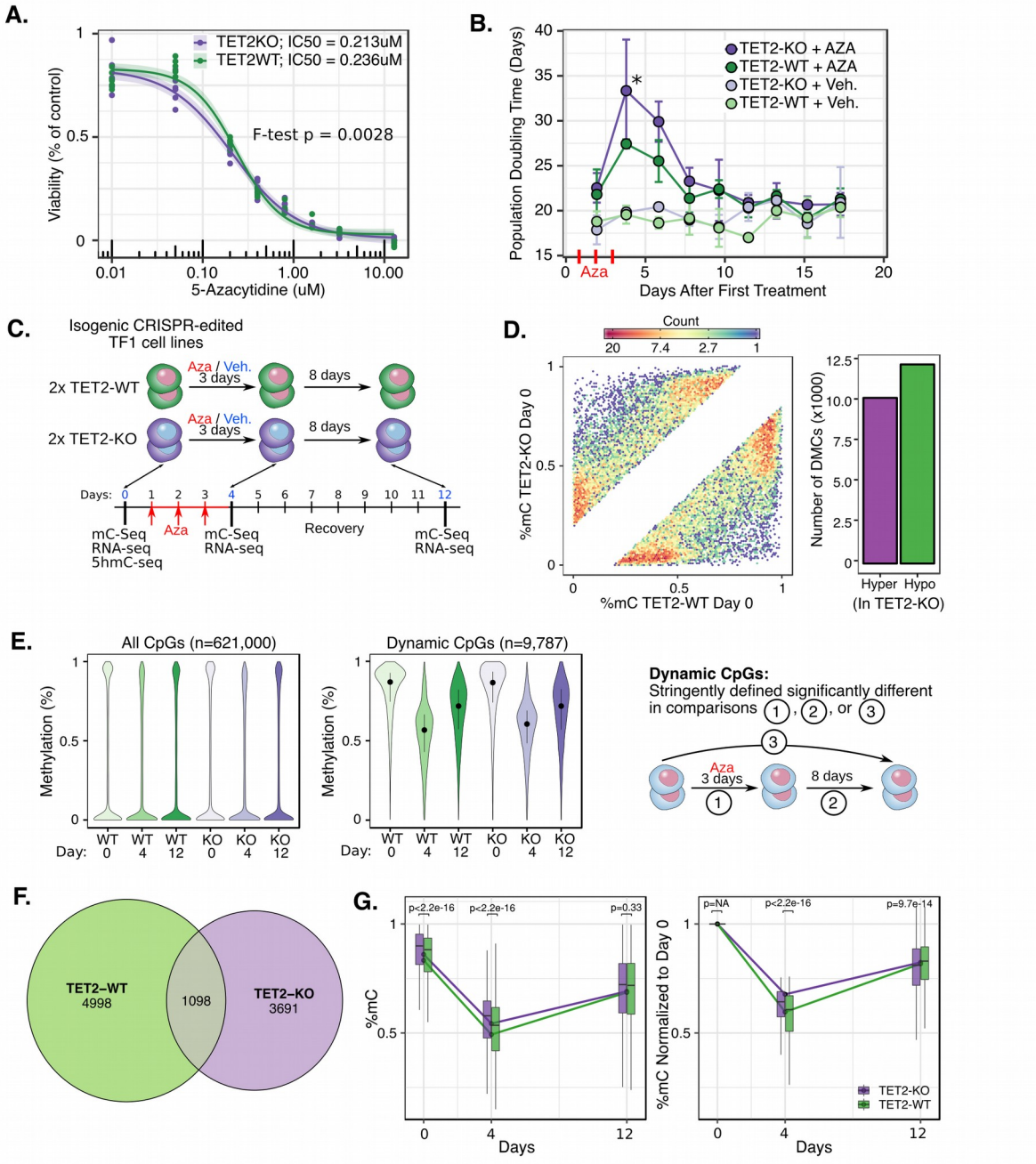


Figure 3.2 Differential response to 5-Aza exposure in TET2KO and WT TF-1. A) IC50 curve for 3x once daily dose of 5-Aza. **B)** Population doubling time changes after 5-Aza or vehicle exposure. **C)** Baseline DNA methylation differences between TET2KO vs. WT TF-1 prior to 5-Aza exposure. **D)** Schematic of 5-Aza exposure experiment and data generated at each time point. **E)** Global patterns of DNAm for (left) all CpGs covered and (middle) "Dynamic" CpGs that were significantly changed during treatment (posterior probability of >20% difference at FDR < 0.1 for comparisons shown in schematic at right). **F)** Venn diagram showing amount of overlap for Dynamic CpGs in TET2 KO vs. WT. **G)** Time profiles for dynamic CpGs reveal differences in remethylation rate for TET2KO vs. WT dynamic CpGs.

Dynamic CpG loci exhibit different patterns and proportions of remethylation kinetics

To investigate different patterns of remethylation kinetics we sought to categorize CpGs based on the speed and extent with which they recovered their initial methylation value. Rather than try to define a host of arbitrary thresholds to determine categories, we instead employed an unsupervised hierarchical clustering of methylation values z-scored across time for each CpG locus in each genotype to define groups of loci with similar time-profiles in an unbiased manner. To decide at what level to cut the clustering tree, we examined the range of K=2 to K=9 potential clusters and calculated the mean silhouette width for each potential clustering solution, whereby the higher the mean silhouette width, the better the solution is able to separate loci into well-defined clusters (Rousseeuw 1987). We observed a peak in the mean silhouette width for K=4 clusters of CpG time-profiles, and upon visual inspection of the different results we decided that 4 clusters captured the majority of the variability in time profiles without over-fitting the results (Supplementary Figure 3.3). The four remethylation patterns identified included three groups of CpGs that were significantly demethylated, with varying speeds/levels of methylation recovery (fast, intermediate, and slow recovery), as well as one group of CpGs that actually gained methylation after 5-Aza exposure and fully recovered by day 12 (Figure 3.3A-B). Interestingly, when we explored how well these four clusters reflected differences in the amount of recovery of methylation patterns by the ‘recovery ratio’ (Figure 3.3D), we observed that the

clusters were perfectly separated based on their recovery ratio. We examined the quantity and relative proportion of dynamic CpGs belonging to each cluster and observed that TET2KO dynamic CpGs had a smaller proportion of CpGs assigned to the rapid-recovery cluster (KO, n=1,467 (30.6%); WT, n=2,546 (41.7%) and a much larger proportion of CpGs assigned to the slow-recovery cluster (KO, n=1,111 (23.2.6%); WT, n=804 (13.2%)), in agreement with our previous finding that TET2KO dynamic CpGs displayed slower global remethylation kinetics. Finally, to correlate these categories of dynamic CpGs, we determined the amount of overlap of each clusters CpGs with regulatory chromatin states defined in a reference CD34+ hematopoietic stem cell epigenome (Figure 3.3D). In both genotypes we observed striking enrichment for dynamic CpGs located in regions of active transcription as well as in active enhancers. Notably, there was a trend for actively transcribed regions as well as gene-body enhancers to be more rapidly remethylated (enrichment: rapid > intermediate > slow recovery), while the trend was in reverse for non-gene-body enhancers and transcribed regions at gene 3' or 5' (slow > intermediate > rapid recovery). The one region where the trends were not identical between genotypes was for enhancers not located in gene bodies, where the highest enrichment was in the slow-recovery cluster for TET2KO, while for WT the highest enrichment was for intermediate-recovery CpGs, but the differences in enrichment weren't significant ($p = 0.23$).

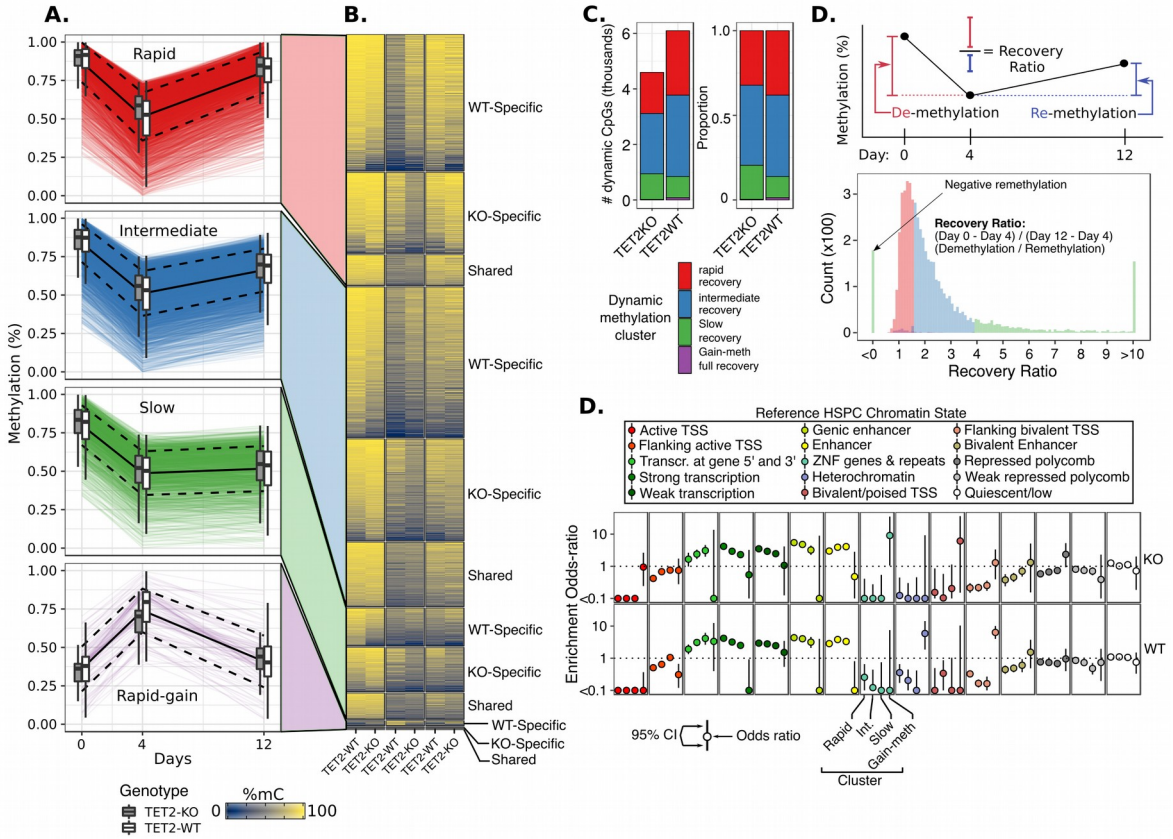


Figure 3.3: Hierarchical clustering of time-scaled CpG methylation values resulted in four major clusters characterized by differences in remethylation rates. A) Time profile of dynamic CpG loci for each cluster. Each colored line represents a single CpG locus; black solid line represents mean of all loci in cluster; dashed black lines represent mean +/- 1 standard deviation; boxplots show summary statistics for WT and TET2KO. **B)** Heatmap of dynamic CpG mean methylation values for each genotype stratified by whether the locus was a shared dynamic CpG or unique to either genotype. **C)** Proportion of dynamic CpGs belonging to each cluster. **D)** (top) schematic of the 'recovery ratio' which is plotted in histogram below, colored by cluster. Clusters were perfectly separated based on the recovery ratio statistic. **E)** Odds ratio of enrichment of dynamic CpGs for regulatory chromatin states defined in a reference normal CD34+ HSPC epigenome (Roadmap Epigenomics, E051).

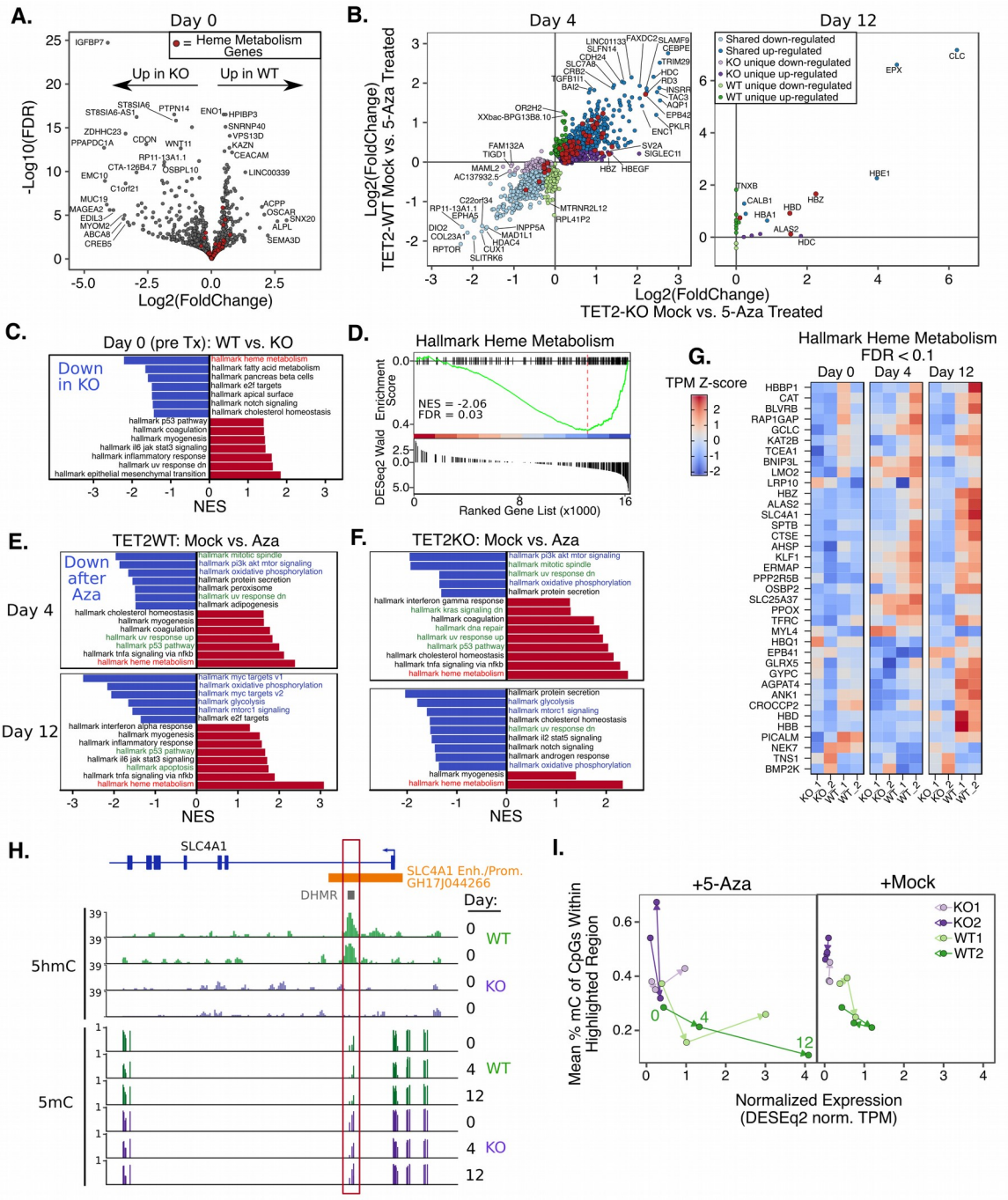
Dynamic gene expression changes before and after 5-Aza exposure reveal alterations in differentiation-associated gene signatures

To investigate differential gene regulation between genotypes, as well as the pathways effected by 5-Aza exposure, we performed RNAseq on the same days as methylation profiling as indicated in Figure 3.2C. We first determined differentially expressed genes at baseline (Day 0, pre-treatment) and identified

414 down- and 404 up-regulated genes between TET2KO and WT. We performed gene-set enrichment analysis (GSEA) (Subramanian et al. 2005) to determine differentially regulated pathways between TET2KO and WT and observed highly significant enrichment of ‘hallmark heme metabolism’ (down-regulated in KO relative to WT) and ‘hallmark epithelial mesenchymal transition (EMT)’ (up-regulated in KO relative to WT) gene sets (Figure 3.5A,C), along with others. Next we determined differentially expressed genes between 5-Aza and vehicle (DMSO) treated conditions on days 4 (24hrs after final 5-Aza) and 12 (8 days after final dose). On day 4 we observed widespread changes in transcription, with 2,912 genes differentially expressed in TET2KO and 2,518 in WT, with the majority of genes exhibiting the same behavior in both conditions (shared: 1,011 down, 854 up; KO unique: 432 down, 634 up; WT unique: 370 down, 270 up) (Figure 3.4B). By Day 12 the widespread transcriptional changes induced by 5-Aza were almost completely normalized, with only 25 genes differentially expressed in either genotype compared with vehicle (shared: 8 up; KO unique: 5 up; WT unique: 9 up, 3 down). As might be expected of a DNA damaging agent like 5-Aza, gene sets involved in DNA damage response, cell cycle, and growth/metabolism were all down-regulated, while inflammation and apoptosis gene sets were up-regulated on Day 4 for both WT and KO. By day 12 many of the same gene sets were still enriched, indicating that although many genes were no longer ‘significantly’ differentially expressed, there were likely still large numbers of small-magnitude changes that had not fully recovered. Notably, the heme metabolism erythroid

differentiation gene set was the most significantly up-regulated gene set in both genotypes on day 4, and in the case of WT cells, its enrichment grew more significant by day 12. When we examined the expression patterns of genes in the heme metabolism gene set over time, it was clear that the TET2WT cells gained progressively higher expression from day 0 through day 12, while the TET2KO cells peaked on day 4 before trending downwards by day 12. To investigate this trend in more detail we examined our integrated genomic datasets at the key erythroid differentiation marker gene, erythrocyte anion exchanger, band 3 (SLC4A1). In a genome browser track of the gene (Figure 3.4H) we observed that there was a differentially hydroxymethylated region within the enhancer/promoter region of the gene, and directly overlapping this region were dynamic CpGs which had opposing patterns of remethylation kinetics in TET2KO and WT, where WT lost methylation and did not recover, while the KO rapidly recovered its original methylation. When we correlated the methylation of this locus with expression patterns it was clear that 5-Aza treatment was able to induce expression of this gene in WT but to a much lesser extent in KO.

Figure 3.4: Gene expression changes after 5-Aza treatment reveal differences in erythroid gene signatures. A) Differential expression between TET2KO and WT at baseline. **B)** Differential expression after 5-Aza exposure in TET2KO and WT. **C)** Gene Set Enrichment Analaysis (GSEA) for baseline gene expression differences (KO vs. WT). **D)** GSEA plot for the most significantly enriched gene set in (C): Heme metabolism genes. **E-F)** GSEA results for comparisons of mock vs. 5-Aza treated TF1 on Day 4 and Day 12 after treatment. **G)** Heatmap displaying genes in the Hallmark Heme Metabolism gene set which were differentially expressed between TET2 KO and WT TF1 at any time point. **H)** Genome browser track highlighting 5mC and 5hmC differences centered in the SLC4A1 promoter/enhancer region. Enhancer annotation derived from GeneHancer Regulatory Interactions (double elite) table on UCSC genome browser. DHMR = Differentially Hydroxymethylated Region (as determined by differential peak enrichment analysis). **I)** Scatterplot displaying the relationship between % 5mC of the highlighted region in (I) and gene expression over time in TET2KO and WT TF1. Line arrows in plot indicate directionality of time (Day 0→ Day 4→ Day 12).


A.
Day 0
= Heme Metabolism Genes
Up in KO
Up in WT
-Log10(FDR)
Log2(FoldChange)
B.
Day 4
Day 12
Shared down-regulated
Shared up-regulated
KO unique down-regulated
KO unique up-regulated
WT unique down-regulated
WT unique up-regulated
Log2(FoldChange) TET2-WT Mock vs. 5-Aza Treated
Log2(FoldChange) TET2-KO Mock vs. 5-Aza Treated
C.
Day 0 (pre Tx): WT vs. KO
Down in KO
hallmark heme metabolism
hallmark fatty acid metabolism
hallmark pancreas beta cells
hallmark e2f targets
hallmark apical surface
hallmark notch signaling
hallmark cholesterol homeostasis
hallmark p53 pathway
hallmark coagulation
hallmark myogenesis
hallmark il6 jak stat3 signaling
hallmark inflammatory response
hallmark uv response dn
hallmark epithelial mesenchymal transition
NES
D.
Hallmark Heme Metabolism
Enrichment Score
NES = -2.06
FDR = 0.03
DESeq2 Wald
Ranked Gene List (x1000)
G.
Hallmark Heme Metabolism
FDR < 0.1
TPM Z-score
Day 0
Day 4
Day 12
E.
TET2WT: Mock vs. Aza
Down after Aza
Day 4
Day 12
hallmark mitotic spindle
hallmark pi3k akt mtor signaling
hallmark oxidative phosphorylation
hallmark protein secretion
hallmark peroxisome
hallmark uv response dn
hallmark adipogenesis
hallmark cholesterol homeostasis
hallmark myogenesis
hallmark coagulation
hallmark uv response up
hallmark p53 pathway
hallmark tnfa signaling via nfkb
hallmark heme metabolism
hallmark myc targets v1
hallmark myc targets v2
hallmark glycolysis
hallmark mtorc1 signaling
hallmark e2f targets
hallmark interferon alpha response
hallmark inflammatory response
hallmark il6 jak stat3 signaling
hallmark apoptosis
NES
F.
TET2KO: Mock vs. Aza
hallmark pi3k akt mtor signaling
hallmark mitotic spindle
hallmark uv response dn
hallmark oxidative phosphorylation
hallmark protein secretion
hallmark interferon gamma response
hallmark kras signaling dn
hallmark coagulation
hallmark dna repair
hallmark uv response up
hallmark p53 pathway
hallmark cholesterol homeostasis
hallmark tnfa signaling via nfkb
hallmark heme metabolism
hallmark protein secretion
hallmark glycolysis
hallmark mtorc1 signaling
hallmark cholesterol homeostasis
hallmark uv response dn
hallmark il2 stat5 signaling
hallmark notch signaling
hallmark androgen response
hallmark oxidative phosphorylation
hallmark myogenesis
hallmark heme metabolism
NES
H.
SLC4A1
SLC4A1 Enh./Prom. GH17J044266
DHMR
Day:
5hmC
5mC
WT
KO
I.
+5-Aza
+Mock
KO1
KO2
WT1
WT2
Mean % mC of CpGs Within Highlighted Region
Normalized Expression (DESEq2 norm. TPM)

TET2KO cells are not blocked from differentiation in the presence of EPO, but 5-Aza enhances differentiation capacity

To determine whether the differences in expression of erythroid differentiation genes were at levels relevant for impacting the differentiation capacity of these cells we measured Glycophorin-A (GPA) cell surface expression by flow-cytometry as a surrogate for differentiation along the erythroid lineage during Aza exposure and erythropoietin (EPO) induced differentiation (Figure 3.5). As expected, TET2WT cells had higher GPA expression at baseline, and 5-Aza exposure increased expression to a similar degree in both genotypes (~3-4fold). We also observed a trend for increased GPA expression over time, albeit to a lesser degree, in the WT vehicle treated condition, potentially due to spontaneous differentiation or due to vehicle exposure, however in TET2KO cells we did not see any significant increases in GPA expression in the same conditions. In agreement with our previous RNAseq results, the KO cells displayed peak GPA expression on day 4, while the TET2WT cells steadily increased over time, even through day 20. When we calculated the fold-change in GPA expression for 5-Aza vs. vehicle treated conditions, it was apparent that TET2KO condition had significantly greater relative GPA induction (Figure 3.5B). Finally, we took advantage of the ability of TF-1 cells to differentiate along the erythroid lineage when exposed to erythropoietin to assess how TET2 loss and 5-Aza exposure would impact differentiation (Figure 3.5C). There were wide variations in the relative amounts of differentiation among clones from both genotypes, but in general, all clones

retained the ability to differentiate. 5-Aza exposure in WT and to a lesser extent KO clones increased the clones capacity for differentiation as can be seen through relative increases in GPA expression as well as reddening of the cell pellets which is indicative of higher hemoglobin expression.

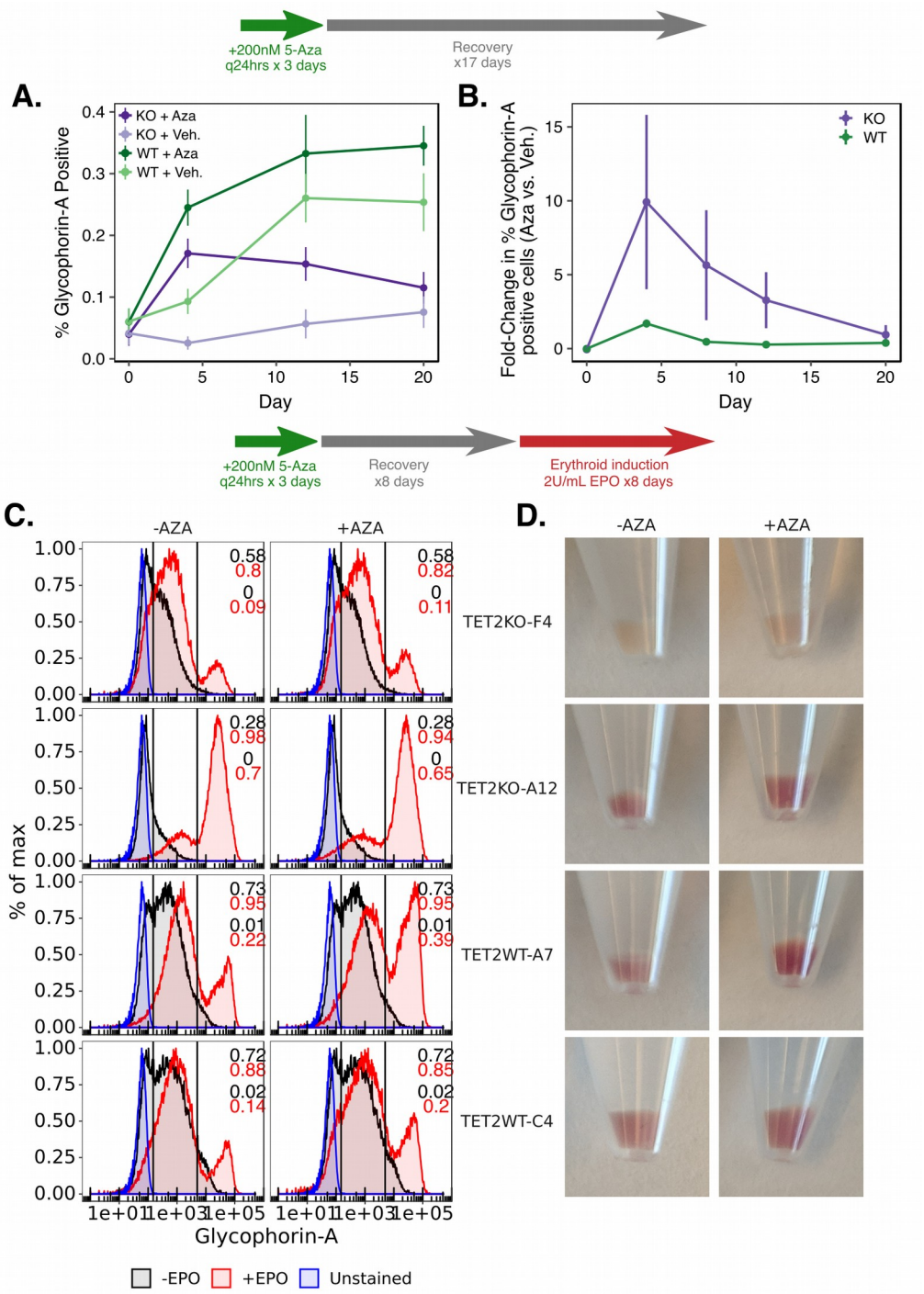


Figure 3.5 Expression of the erythroid cell surface protein Glycophorin-A (GPA) is differentially induced in KO v. WT after 5-Aza but not during erythroid differentiation. A) Percentage of GPA-positive cells over time after 5-Aza or vehicle exposure as measured by flow cytometry. Drug/vehicle was added on Days 1-3 as in previous experiments. **B)** Fold-change in cell-surface GPA expression over time after 5-Aza vs. vehicle exposure. **C)** Erythroid differentiation induction by recombinant erythropoietin as measured by GPA flow cytometry (left). Red pigmentation of cell pellets is indicative of the amount of hemoglobin expression (right).

3.4 Discussion and Conclusions

DNA methyltransferase inhibitors such as 5-azacitidine are increasingly used alone and in combination with other therapies in treatment regimens for a variety of cancers, yet the mechanisms predicting therapeutic response to these agents remains elusive. We have previously shown in MDS patients, that the presence of a *TET2* mutation in patient tumors was predictive of patient responses to DNMTIs, and in the present study we attempt to better understand the mechanisms behind that association. We applied a number of pharmacological and high throughput sequencing approaches to establish the effects of *TET2* loss in a TF-1 leukemia cell line models, and we studied how this loss may impact responses to the DNA methyltransferase inhibitor azacitidine. We identified a number of processes which were differentially effected in TET2KO cells, including altered sensitivity and recovery after 5-Aza exposure, as well as down-regulation of typical erythroid differentiation processes, and we showed how these relate to DNAm and 5hmC alterations in these cell lines. Our results highlight the impact of TET2 loss on the DNA methylation kinetics of differentiation-associated genes and their enhancers, and how 5-Azacitidine can target these regions to improve baseline differentiation capacity.

We observed key differences in the pharmacological and epigenetic responses to 5-Aza between TET2KO vs WT TF-1. Notably, both genotypes had similar dose-response curves but KO cells had significantly increased population

doubling time after 5-Aza. We also observed that TET2KO cells had globally reduced demethylation and recovery extent. The reduced DNAm demethylation in KO is likely related to the lengthening of population doubling time, as passive demethylation by 5-Aza is dependent on DNA replication and cell division, and therefore because there are fewer cell divisions in TET2KO after 5-Aza, there is less demethylation as well. Interestingly we did observe up-regulation of gene sets involved in DNA damage response (hallmark p53 pathway, hallmark UV response down) in TET2KO at baseline. Several recent studies have shown that TET2 and 5hmC play essential roles in genome stability, and deficiencies of TET enzymes and 5hmC lead to defects in chromosome segregation and DNA damage responses (Y. W. Zhang et al. 2017; L.-L. Chen et al. 2018; Kafer et al. 2016). Together, these data may suggest that TET2KO cells are deficient in DNA damage responses and are likely less capable of coping with the added DNA damage induced by 5-Aza exposure, thus more profound stalling of cell cycle and less passive DNA demethylation occurs. This mechanism also helps explain findings from our previous study showing that 5-Aza treatment significantly decreased peripheral blood chimerism after competitive bone marrow transplantation in Tet2-null but not WT cells (Bejar, Lord, et al. 2014b).

A notable finding of our study was our characterization of the time-scale of dynamic changes in DNAm and gene expression after Aza exposure. Notably, while we observed relatively rapid recovery of transcriptional changes induced by 5-Aza with less than 1% of day 4 DEGs present on day 12, many of the DNA

methylation changes observed on day 4 were not fully recovered by day 12. This fact is relevant in terms of the role of Aza as a differentiating agent because recent work has shown that DNAm changes, especially at bivalent promoters, even in the absence of gene expression changes can impact the later differentiation ability of stem cells (Verma et al. 2018). We determined there were four major patterns of remethylation and observed differences in the proportion of loci falling in each pattern, where TET2KO exhibited a significantly higher proportion of slow-remethylating and significantly lower proportion of rapid-remethylating CpGs. We observed enrichment for regions of active gene-transcription and active enhancers when calculating the overlap of CpGs from each dynamic cluster in relation to reference HSPC chromatin states. Interestingly, actively transcribed regions as well as gene-body enhancers showed a positive trend in enrichment, with the highest enrichment for rapid-recovery CpGs and lowest for slow-recovery CpGs. DNAm in gene-bodies is known to positively correlate with gene expression, and our findings confirm these observations, while also demonstrating that gene-body enhancers exhibit similar behavior (Neri et al. 2017; Yang et al. 2014). Interestingly, active enhancers located outside of gene bodies displayed the opposite trend, where they were more strongly enriched in the intermediate- and slow-recovery than the rapid-recovery clusters. This finding is interesting in that it suggests that the long-term DNAm changes induced by 5-Aza occur somewhat selectively at enhancers while minimally impacting other, more rapidly recovered cellular processes.

A key finding of our study was the influence of *TET2* loss on differentiation-associated gene expression signatures in TF-1 cells, and how these signatures are disproportionately influenced by 5-Aza exposure in KO compared to WT cells. Erythroleukemic TF-1 cells retain the ability to differentiate along the erythroid lineage in response to a variety of stimuli, and even in the absence of differentiation inducing agents, some population of cells (~5%) constantly undergoes spontaneous differentiation (Kitamura et al. 1989; Losman et al. 2013; Wang et al. 2013). At baseline, the most differentially enriched gene sets between TET2KO and WT TF-1 were those associated with heme metabolism (down-regulated), and epithelial mesenchymal transition (EMT) (up-regulated). The combination and direction of these two gene sets in particular are interesting, as both are involved in opposing directions of differentiation, with heme metabolism involving the terminal differentiation of erythroblasts, and EMT which has been described as a reversal of epithelial differentiation (T. Chen et al. 2017). Notably, both KO and WT 5-Aza treated cells had significant up-regulation of heme metabolism genes relative to vehicle on days 4 and 12 after Aza exposure, however WT had much higher expression of these genes than KO on all days (Figure 4G). At least some of this discrepancy in erythroid gene expression between genotypes is likely due to loss of hydroxymethylation at differentiation-associated enhancers in KO condition leading to differences in remethylation kinetics as we showed for the erythrocyte anion exchanger, Band 3 (SLC4A1). When we examined Glycophorin-A (GPA) protein expression as a measure of

erythroid differentiation during 5-Aza exposure, we observed that a substantial amount of vehicle-treated WT cells became GPA-positive over time, while virtually no KO cells did. This may have been due to vehicle inducing differentiation or due to spontaneous differentiation over time, however in either case it shows how KO cells are less sensitive to weak differentiation signals. When examining the fold-change in GPA expression compared to vehicle-treated cells, 5-Aza treated KO cells had significantly higher induction of GPA relative to WT. Notably, both TET2KO and WT TF-1 retained erythroid differentiation capacity when exposed to a strong differentiation stimulus, such as erythropoietin. All of these results suggest that DNA methylation and hydroxymethylation changes induced by *TET2* loss shift TF-1 cells towards a more undifferentiated state (high EMT, low heme metabolism genes), making them *relatively* less sensitive to weak differentiation signals (such as 5-Aza or vehicle exposure) when compared to wild-type. However these DNAm and hmC changes make TET2KO cells more targetable by 5-Aza as shown by the fold induction in differentiation genes being much greater in KO vs WT. Our results are in accord with studies demonstrating how TET2 loss can impair differentiation of stem cells (Madzo et al. 2014; Hon et al. 2014; Verma et al. 2018) as well as studies showing that reduction of 2-hydroxyglutarate (which inhibits TET2) can improve erythroid differentiation (Losman et al. 2013; Wang et al. 2013).

A potential limitation of our study is that we base our conclusions upon a relatively small number of single-cell derived TF-1 clones which are likely to have

high clonal variation. However, given the relatively large effect sizes we observed and consistency across time for each genotype in these highly variable clones, it may also strengthen the credibility of our results that we were able to establish these findings in the presence of such variability. Another unexpected and potential limitation is that our drug vehicle, DMSO, appeared able to induce differentiation-associated markers in wild-type cells over time, even at the exceedingly low concentrations (~2 ppm) used in these experiments. While this was not desirable for our experiment, it actually helped us better understand the differences in differentiation capacity between WT and KO cells.

In conclusion, we demonstrate that CRISPR/Cas-mediated *TET2* inactivation in erythroleukemia cell lines pushed cells towards a more undifferentiated state which lengthened their recovery times after 5-Aza exposure, as well as disproportionately induced erythroid differentiation signatures after treatment. We also showed that these changes likely have to do with 5hmC and 5mC dynamics at erythroid enhancers, which correlated with gene expression. Overall our study helps shed light on the relationship between TET2 mutation and increased sensitivity to 5-Aza in MDS patients, and provides several insights on potential mechanisms which should be further studied.

3.5 Materials and Methods

Cell culture

The human GM-CSF cytokine dependent erythroleukemia cell line TF-1 (ATCC CRL-2003) were maintained in RPMI-1640 medium (Life Technologies) supplemented with 10% fetal bovine serum (Omega Scientific) and 1% penicillin/streptomycin/L-glutamine (Thermo Fisher cat #10378016). Cells were maintained at a density between $5x10^4$ and $5x10^5$ viable cells/mL and passaged every 2-3 days. Cell viability was assessed by Trypan Blue exclusion.

TET2 gene editing with CRISPR/Cas9

CRISPR single-guide RNAs (sgRNA) were designed with CHOPCHOP (chopchop.cbu.uib.no) and then cloned into px458 vector (addgene, 48138) following the protocol in Ran et al (Ran et al. 2013; Labun et al. 2016). We followed the approach of Bauer et al. using 2 CRISPR sgRNAs to delete the dioxygenase active site residues of TET2 following the design in Figure 3.1A (Bauer, Canver, and Orkin 2015) 2ug of px458 vectors with incorporated sgRNAs or unmodified px458 (for wild-type controls) were nucleofected into parental TF-1 cells with the Lonza Nucleofector Kit V following the manufacturer protocol. Nucleofected TF-1 were allowed to grow undisturbed for 24hrs before sorting live GFP-positive cells using a BioRad S3 Cell Sorter. Bulk GFP-positive sorted cell populations were allowed to grow undisturbed for 3 days before isolating single-cell clones by limiting dilution into 96-well plates. Single-cell clones were screened for inactivating out-of-frame indels/deletions by PCR followed by Sanger sequencing, and later by immunoblotting for TET2 protein and 5-hydroxymethylcytosine. CRISPR sgRNA sequences are as shown in Figure 3.1A.

Western blotting

Cell density was calculated before 10 million cells were lysed directly in 500uL 2x Laemmli sample buffer (Bio-rad) freshly prepared with 2.5% B-mercaptoethanol. Lysates were then sonicated at 75% amplitude for 15 seconds to shear DNA and then boiled at 95C for 5 minutes. 20UL of whole cell lysates were separated by 7.5% Tris-Glycine (Bio-rad miniprotean TGX) SDS-PAGE and proteins were transferred onto PVDF membranes. Immunodetection of proteins was carried out in TBST (20mM Tris-Hcl, 500mM NaCl, 0.05% Tween-20, pH 7.5) supplemented with 5% non-fat milk and anti-TET2 (CST-18950), anti-DNMT1 (Abcam ab92314), anti-Lamin-B1 (proteintech 12987-1-AP), or anti-β-actin (Biolegend poly6221), followed by incubation with HRP-conjugated secondary antibodies.

L-ascorbic acid (L-AA) treatment and 5hmC DNA immunoblotting

First LAA exposure dose was titrated to a level that minimized cell death 24hrs after exposure. Cells were treated with a minimally cytotoxic dose of 100uM LAA and genomic DNA was extracted 24hrs after exposure for dot blot analaysis. 5-hydroxymethylcytosine DNA immunoblotting was carried out as described previously.(M. Lee et al. 2017) Briefly, DNA was denatured and spotted in 2-fold serial dilutions on a nitrocellulose membrane in a Bio-Dot apparatus. A synthetic oligonucleotide containing a known quantity of 5hmC was used as standard. Immunodetection of 5hmC was carried out using anti-5hmC antibody at 1:5000 dilution (Active Motif) followed by incubation with 1:10,000 dilution of HRP-

conjugated secondary antibody and visualization with West-Q Picu Dura ECL Solution (GenDEPOT). 5hmC dot blot signal was quantified using imageJ. A standard curve was constructed as in Ko et al. and the linear portion of the curve was used to calculate the moles of 5hmC per sample shown in Supplementary Figure 3.1 (Ko et al. 2010)

Drug treatment and viability assays

Viability of cultures that were treated with varying concentrations and dosing schedules of 5-Aza were assessed via CellTiter Blue viability assay following manufacturers protocol. Briefly, cells were seeded at 2,000 cells per well in 96-well culture plates with final volume of 100uL culture media 24 hours prior to addition of first drug dose. Viability was assessed 24hrs after each final treatment (48,72, or 96hrs for the 3 treatment schedules tested). IC50 curves were modeled using the four parameter log-logistic regression model using the R package ‘drc’. Differences in IC50 curves between TET2-WT and TET2-KO cell lines were tested via one-way ANOVA (R ‘anova’ function). After dose-optimization, the final experiment from which genomic analyses were obtained was performed as follows: TF1 cell lines were seeded at a density of 100,000 cells/mL on Day 0 and allowed to equilibrate overnight. Freshly prepared 5-Aza or vehicle (DMSO) was added to cultures every 24hrs for a total of three doses of 200nM starting on Day 1. Media was replaced on Day 2 and Day 4, and every 2-3 days following until Day 12.

Induction of erythroid differentiation of TF-1 cells

Erythroid differentiation induction was carried out as described (Losman et al. 2013). Briefly, TF1 cells growing in GM-CSF containing medium were spun down and washed three times in fresh media lacking GM-CSF before being seeded at a density of 200,000 cells/mL in 20mL media lacking GM-CSF. After 24hrs of culture in these cytokine poor conditions, recombinant human erythropoietin (Epogen, Epoeitin alfa, Amgen) was added directly to cultures to a final concentration of 2U/mL. Fresh media with 2U/mL erythropoietin was added four days later and flow cytometry assessment of differentiation was performed 8 days after the first erythropoietin addition.

Flow cytometry

FACS sorting of GFP-positive px458 transfected cells was performed using the Bio-Rad S3 Cell Sorter. During 5-Aza treatment and erythroid differentiation experiments, cells were stained with various combinations of monoclonal antibodies to erythroid differentiation markers including PE-Cy5 conjugated anti-CD235AB (Biolegend cat. # 306605), FITC conjugated anti-CD71 (Biolegend cat. # 334104), and FITC conjugated anti-CD233 (Miltenyi Biotec cat. # 130-119-818) to measure erythroid differentiation by flow cytometry on an Attune NxT Acoustic Focusing Flow Cytometer (ThermoFisher). Data were analyzed in R 3.4 and the Bioconductor ‘FlowCore’ package.

Bisulfite Padlock Probe library construction and sequencing analysis

BSPP libraries were constructed as previously described, and sequenced using a standard 150bp paired-end sequencing protocol on the Illumina NovaSeq

(Diep et al. 2012). Sequencing reads were trimmed for low quality bases and adapter content using TrimGalore v0.4.0 and cutadapt v.1.18 before alignment to a 3-letter (bisulfite converted) version of the hg19 reference genome as previously described (Guo et al. 2017). Briefly, for alignment, trimmed reads were encoded such that all cytosines converted to thymines and for the reverse complement strand, all guanines to adenines, so that the reads could align to the three-letter genome. Reads were then mapped separately to both the Watson and Crick strands of the bisulfite-converted hg19 genome using BWA mem v.0.7.12 (options: '-B2 -c1000'). Alignment with mapping quality scores below 5 were discarded and for reads which mapped to more than one position, and only the highest quality mapping was kept for each read. Overlapping paired-end reads were clipped with BamUtil clipOverlap function. Encoded read sequences were replaced by the original read sequences in the final BAM alignment files which were then used for extracting methylation levels for CpGs.

DNA methylation

Statistical analyses of BSPP methylation data

Differential methylation between different time points and conditions were performed using the R package 'DSS' version 2.26.0. Differentially methylated CpGs were determined using the 'callDML' function of DSS with a posterior probability threshold of 0.90 or greater that the absolute methylation difference was greater than 20% (e.g. 90% chance the difference is 20% or greater). DNAm time-profile clustering was performed as follows: First 'dynamic CpGs' were

identified by determining CpGs with differential methylation between any of the three time points (Day 0 v. 4; Day 0 v. 12; Day 4 v. 12). This yielded a set of dynamic CpGs for each genotype (KO and WT) which had some amount of overlap but were largely unique to each genotype. Next, the mean methylation value for each dynamic CpG locus was converted to a Z-score across the three time points (Days 0, 4, 12) for each genotype. Next, we combined the Z-scaled methylation values for both KO and WT and performed hierarchical clustering with euclidean distance and Ward's agglomeration method. To determine the optimal number of clusters for downstream analysis we examined the methylation time profiles for a range of K's, and settled on the K=4 cluster solution because it captured the major patterns of remethylation and statistically it produced the most compact clustering solution as indicated by the mean cluster silhouette width statistic.

RNAseq library preparation and analysis

Total RNA was isolated from cell lines using the Machery-Nagel Nucleospin RNA kit according to manufacturer instructions. RNA sequencing libraries were constructed using the Illumina TruSeq stranded mRNA kit and sequenced on an Illumina NovaSeq by NovoGene Inc. Fastq files were trimmed to remove polyA signals and adapter content using cutadapt and quality checked using FastQC. Reads were aligned to the hg19 genome using STAR v2.6.1d and GENCODE v19. Reads mapping to repetitive elements were removed. Transcript counts were summarized using featureCounts v1.6.2 and differential gene expression was

calculated using the R package ‘DESeq2’. The threshold for determining differential expression was an adjusted p-value (FDR corrected) < 0.10. Normalized TPM values were calculated from DESeq2 normalized counts and transcript lengths from GENCODE v19. Gene set enrichment analysis (GSEA) were performed using the GSEA pre-ranked algorithm using the DESeq2 differential expression Wald statistic as the ranking metric. GSEA was carried out in R using the ‘clusterProfiler’ package **(Yu et al. 2012)**

5-hydroxymethylcytosine pull-down (HMCP) sequencing with reference exogenous genome spike-in normalization

We employed the HMCP with reference exogenous genome (ChIP-Rx) method of Orlando et al. for quantitative normalization of HMCP signal according to an exogenous spike-in genome (Orlando et al. 2014). As exogenous control we used the DNA isolated from phage *T4gt*, which carries mutations in both *-agt* and *-bgt* glycosyltransferases and thus the vast majority of cytosines are only hydroxymethylated without glycosylation (Weigele and Raleigh 2016). HMCP was carried out using the 5-hydroxymethylcytosine pull-down assay (Cambridge Epigenetix) according to manufacturer instructions with slight modifications. Briefly, exogenous phage *T4gt* DNA was spiked into the TF1 cell DNA at a 1:50,000 ratio before shearing with a Covaris E220 to an average length of 150 bp. Sheared DNA samples were end-repaired and adapter ligated with barcoded Illumina TruSeq adapters. Libraries were denatured and a new copy strand was synthesized. DNA fragments containing 5hmC were pulled down using a two step

process where 5hmC were first glycosylated with T4BGT (Thermofisher) using UDP-6-azide-glucose, followed by labeling with PEG (polyethylene glycol)-biotin via copper-free CLICK chemistry. The biotin-containing DNA was then pulled down and purified using streptavidin magnetic beads. Finally, purified libraries were PCR amplified with Kapa real-time library amplification kit, and sequenced using single-end 50bp protocol on the Illumina NovaSeq.

Sequencing reads from HMCP were aligned to a combined genome of hg19 and the T4 phage genome (Genbank: AF158101.6) using a 2-step mapping procedure where raw reads were first mapped with BWA mem v0.7.15 (options: -M -t 8) and then any unaligned reads were trimmed for low quality base calls and adapter content with Trim-galore v0.4.3 (options: –paired –length 35 –stringency 3 –three_prime_clip_R1 1 –three_prime_clip_R2 1) before realignment using BWA-mem with more stringent mapping parameters (options: -M -t 8 -B 6). Reads aligning to ENCODE Blacklisted regions or random chromosomes, as well as any reads which mapped to both the T4 and hg19 genomes were removed. Normalization was performed based on a scaling factor for each sample which was calculated as in Orlando et al. with the modification that T4 phage genomic reads were calculated instead of Drosophila genomic reads. 5hmC peak calling was performed using MACS2 v.2.1.1 (options: callpeak -t IPbam -c INPUTbam -f BAMPE -n Name -g hs -q 0.01 --keep-dup all --nomodel --broad) and differential peak enrichment was calculated using MEDIPS v.1.34.0 and edgeR v3.24.3 (MEDIPS.meth function with options: p.adj=TRUE, diff.method='edgeR',

CNV=FALSE, MeDIP=FALSE, minRowSum=1, quantile=FALSE, TMM=FALSE) on the normalized read counts within peaks.

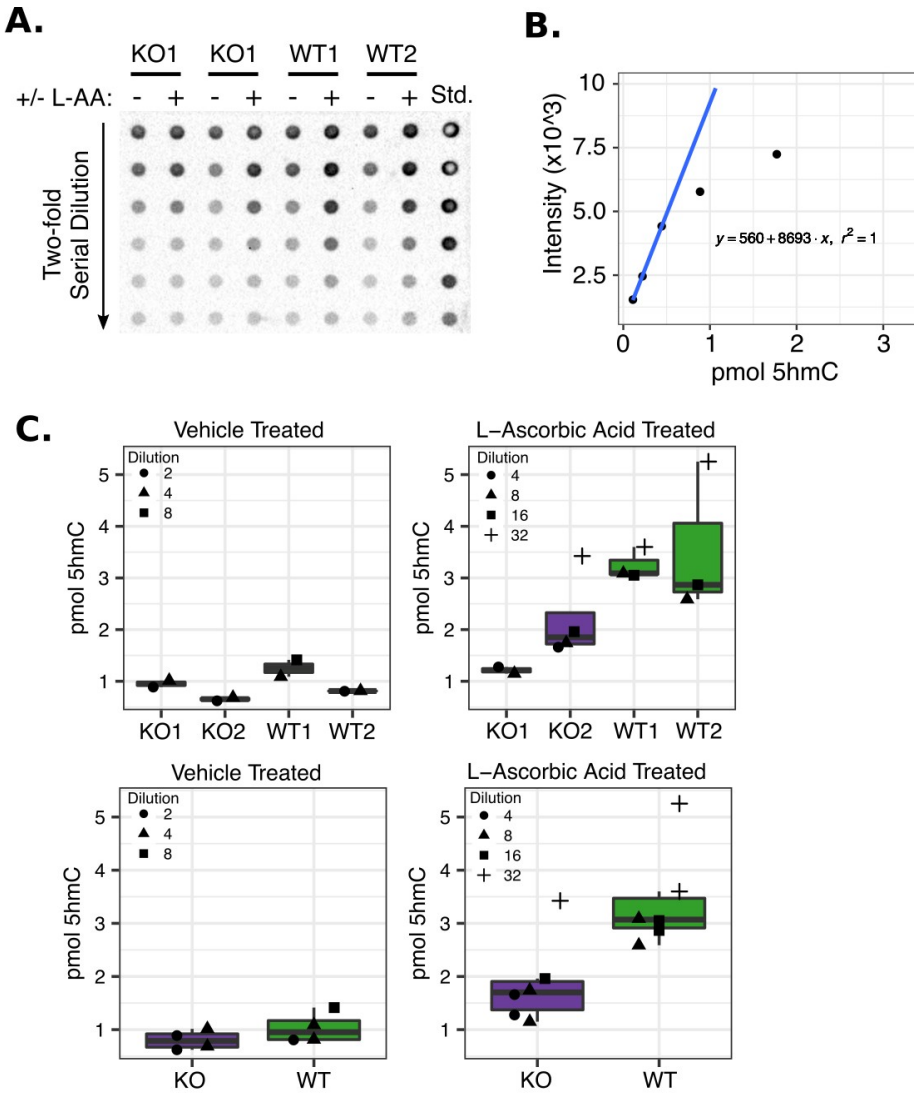
Statistical Analyses

All statistical analyses were performed in R v.3.4.4.

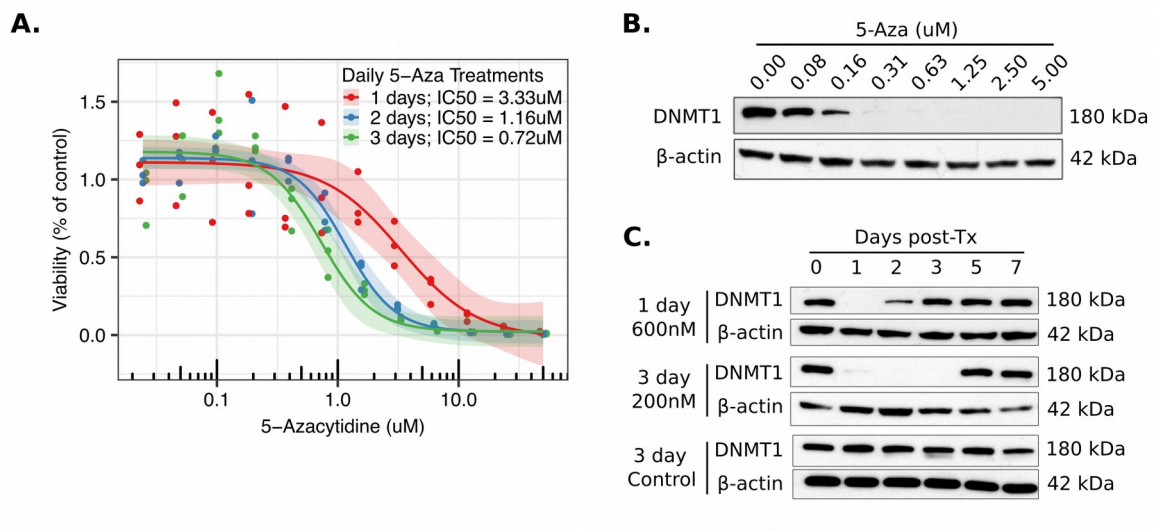
3.6 Acknowledgments

Chapter 3, in part is currently being prepared for submission for publication of the material. Reilly, Brian; Luger, Timothy; Park, Soo; Lio, Chan-Wang; González-Avalos, Edahí; Wheeler, Emily C.; Lee, Minjung; Diep, Dinh; Zhang, Kun, Huang, Yun; Rao, Anjana; Bejar, Rafael. "*TET2* Dependent Effects of 5-Azacitidine on DNA Methylation and Gene Expression Dynamics". The dissertation author was the primary investigator and author of this material. We would like to thank Michael Love and Pablo Tamayo for their advice on pathway enrichment analysis. We would also like to thank Frank Furnari and Renate Pilz for the fruitful discussions regarding the conception and design of several experiments in this study. We kindly thank Peter Weigele (NEB) for providing the original phage *T4gt* which was used in HMCP experiments. This work was supported by the Evans Foundation (B.R., R.B.) and a pre-doctoral fellowship of the American Foundation for Pharmaceutical Education (B.R.).

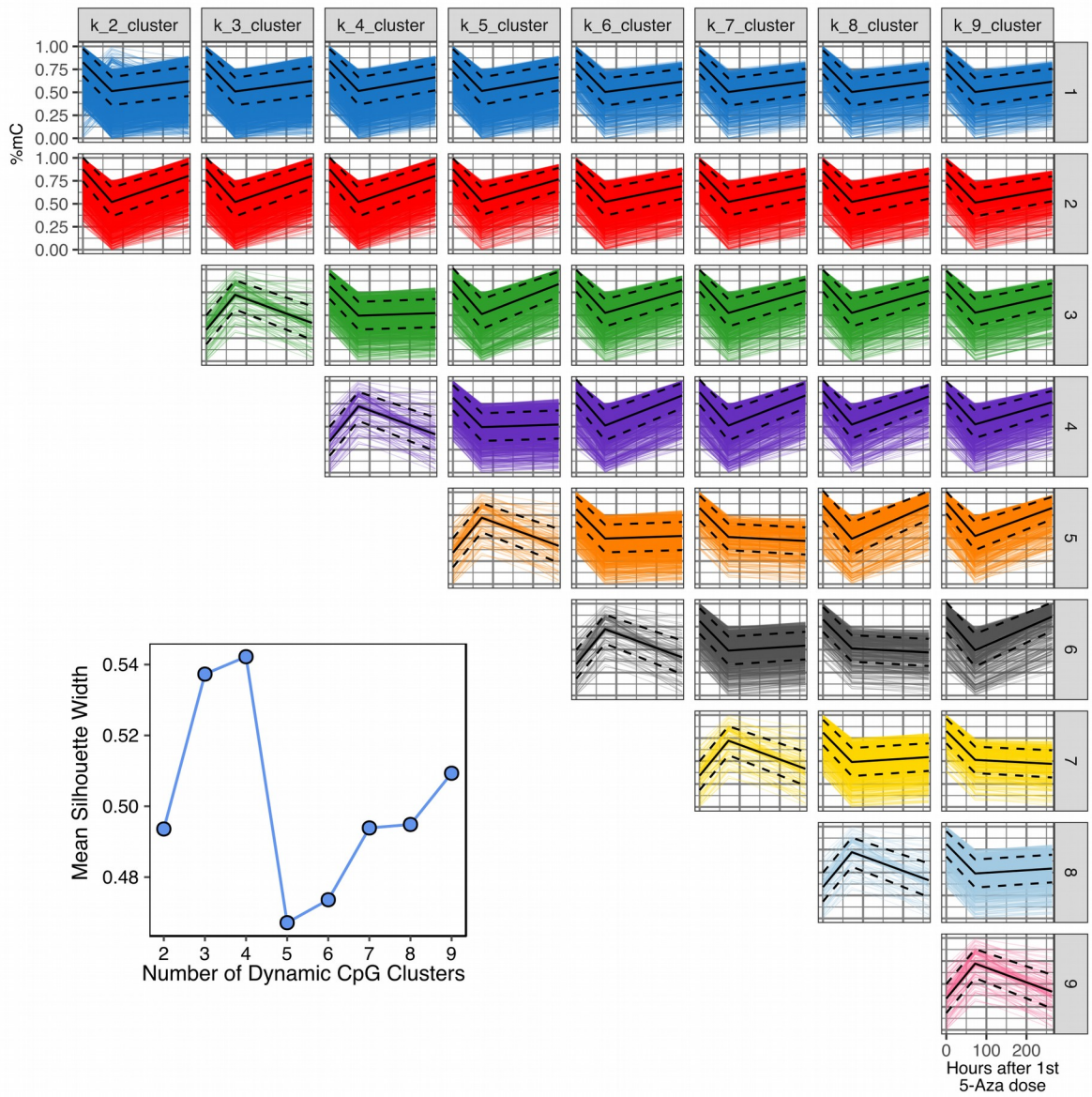
3.7 Supplementary Figures



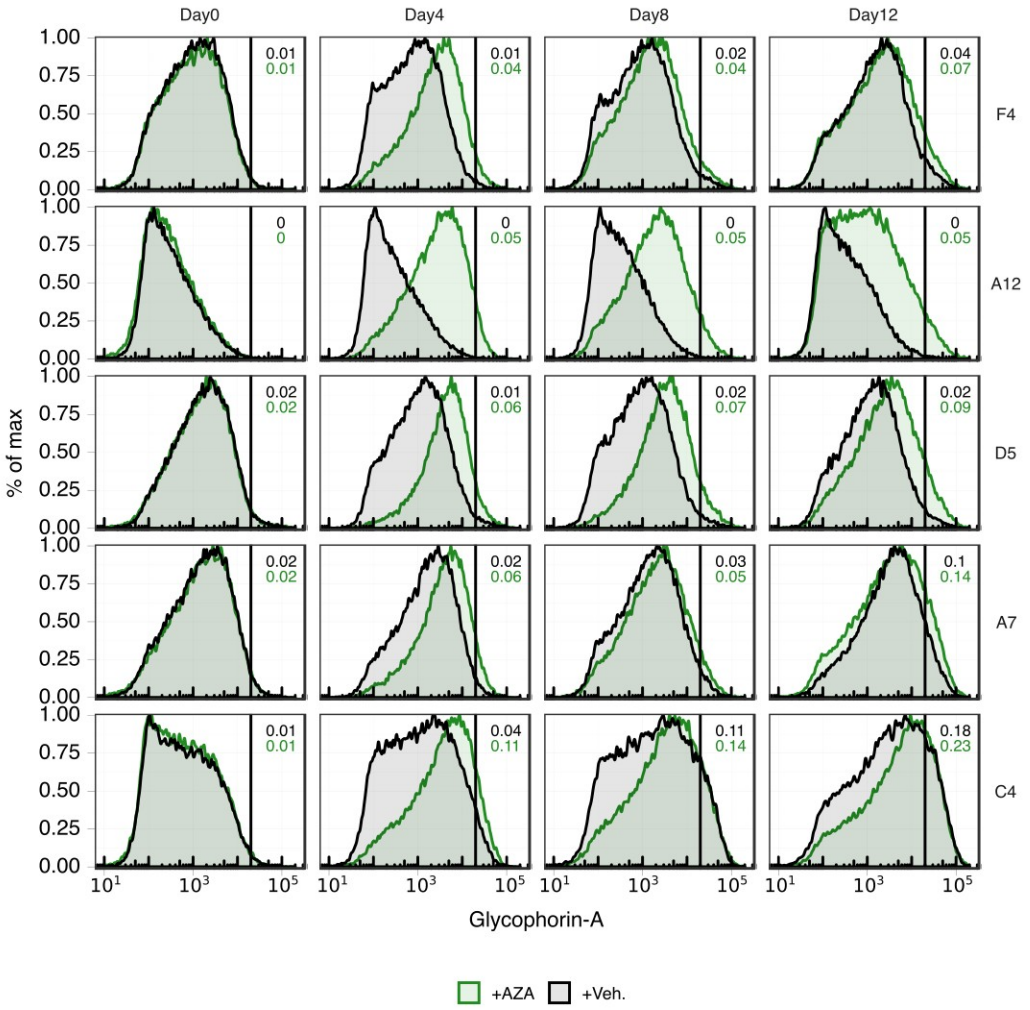
Supplementary Figure 3.1 L-ascorbic acid (L-AA) exposure enhances 5hmC accummulation to a greater extent in TET2-WT TF1 cells. Cells were treated with 100uM L-AA or vehicle 24hrs before DNA isolation. **A)** 5hmC dot blot for L-AA treated or untreated Cells. **B)** Standard curve of 5hmC densitometric signal vs. pmol (calculated from 5hmC standard with known 5hmC amount). **C)** Quantification of dot blot signal for TET2-WT and KO cells +/- L-AA.



Supplementary Figure 3.2 Optimizing 5-Aza dosage regimen in wild-type TF1. A) Dose-response curve of wild-type TF1 cell viability for three different lengths of 5-Aza treatments (1 dose per day for a total length of treatment between 1 and 3 days). Points represent replicates per condition, shaded region corresponds to 95% confidence band estimated by 4 parameter logistic regression model (EC50 and slope as parameter and lower bounded at zero; modeled with R package "drc"). **B)** Dose-response of DNMT1 down-regulation 24 hours after the indicated 5-Aza dose. **C)** Timecourse of DNMT1 down-regulation following the indicated dosage schedules



Supplemental Figure 3.3: Potential clustering solutions for defining groups of dynamic methylation clusters. Line plots in grid: each colored line represents a single CpG locus over time after 5-Aza treatment. Black line indicates mean value for that cluster group. Dotted black line indicates mean +/- 1 standard deviation. From left to right: clustering solutions with K=2 to K=9 clusters (derived from cutting the hierarchical cluster tree at different levels); top to bottom: the patterns of methylation in each cluster for that cluster solution. Mean silhouette width plot shows the mean silhouette width of all clusters for each individual clustering solution. K=4 clusters produced the most well-defined clusters as indicated by the peak in silhouette width.



Supplementary Figure 3.4 Flow cytometry staining of Glycophorin-A in cells exposed to 5-Aza or Vehicle over time. Threshold for positivity determined from erythropoietin induced erythroid differentiation samples shown in Figure 5C. Numbers correspond to percent of cells in the population which were above the positivity threshold for vehicle and 5-Aza treated conditions.

Chapter 4: Discussion and Future Directions

4.1 Summary

A better understanding of the role of DNA methylation and the mechanisms of its dysregulation in cancer is critical for the development of new therapies, predicting responses to current treatments, and improving outcome prediction in heterogeneous diseases like MDS. In this dissertation I focused on two major topics involving DNA methylation: 1) its role in MDS pathobiology and outcome prediction, and 2) the role of inactivation of the DNAm regulator *TET2* in treatment response to DNA methyltransferase inhibitor therapy.

In Chapter 2, I demonstrated how DNAm can be used to resolve some of the clinical and genetic heterogeneity in MDS to improve outcome prediction and enhance our understanding of MDS pathophysiology. I showed that an unbiased unsupervised analysis of DNAm patterns was able to identify groups of MDS patients that displayed distinct differences in their DNA methylomes, patterns and enrichment of genetic lesions, and had significant differences in their overall survival. I showed that the differences in outcome among clusters was not explained by any known prognosis-associated variable, indicating that DNAm plays an important and independent role in disease progression. Furthermore, by looking for differences between DNAm-based patient groups I identified that differences were enriched at genes known to be associated with prognosis, such as WT1, and also identified novel prognosis-associated genes, such as CD93, which were validated in external cohorts of MDS patients to be associated with outcome. Our findings confirm the important role of DNAm changes in these

disorders and highlight the heterogeneity in DNAm patterns among patients with shared somatic mutations.

In Chapter 3, I used a variety of high-throughput genomic approaches to investigate the impact of *TET2* loss in isogenic cell line models of myeloid malignancy, as well as to study how *TET2* loss may impact molecular responses to DNMTI treatment and recovery in these models. I first established and validated two independent isogenic *TET2*-inactivated TF-1 erythroleukemic cell lines (TET2KO). I used these novel cell lines to confirm previous findings that *TET2* loss induced focal loss of 5hmC and that this loss was highly enriched for enhancer elements in reference hematopoietic stem cell data sets. I demonstrated that at baseline, TET2KO cells exhibited a global reduction in differentiation-associated gene signatures and up-regulation of signatures related to stemness. When comparing responses to azacitidine (5-Aza), I showed that TET2KO TF-1 cells had altered sensitivity and lengthened recovery times after azacitidine (5-Aza) exposure, and that KO cells had a disproportionate induction of differentiation-associated genes compared with WT. I also showed how differences in gene expression of differentiation-associated genes such as erythrocyte anion exchanger, band 3 (*SLC4A1*), could be tied to differences in 5hmC and 5mC dynamics after 5-Aza exposure at a key *SLC4A1* transcriptional enhancer. In aggregate our study indicates that the alterations in 5mC and 5hmC at differentiation-associated gene enhancers in TET2KO cells pushes them towards

a less differentiated state, and that 5-Aza treatment can specifically target these alterations to normalize differentiation-associated gene signatures.

4.2 Implications and Future Perspectives

Molecular Sub-types of MDS

The five DNAm-based subgroups of MDS patients we identified in Chapter 2 appear to correlate well with recent work from other groups studying the molecular basis of MDS. While we are the first to report a large-scale unsupervised classification of MDS epigenomes, similar approaches have been applied to gene expression in large MDS cohorts. Shiozawa et al performed an unsupervised classification of MDS based on gene-expression signatures from CD34+ HSPCs from 100 patients and identified two major groups that were distinct in terms of their expression and genetic profiles as well as in terms of their time to progression to AML (Shiozawa et al. 2017). The two groups identified in that study displayed similar patterns of enrichment for mutations that were seen in the DNAm cluster risk groups in our study. For instance Shiozawa et al. identified that mutations in *EZH2*, *RUNX1*, *U2AF1*, *TP53* were enriched in the poor-prognostic risk group, while mutations in *SF3B1* and *TET2* were enriched in the favorable risk subgroup. This corresponds well to our findings, where only *TET2* and *SF3B1* mutations were enriched in the “low-risk” DNAm clusters while the “high-risk” clusters had enrichment for the same mutations mentioned in the high-risk group of Shiozawa et al. Similarly, in an as yet unpublished study utilizing

Illumina 450k methylation arrays to study DNAm in 320 patients with MDS (Morimoto et al. 2018), Morimoto et al identified five DNAm clusters whose patterns of somatic mutations were remarkably similar to our own, with four out of the five clusters identified having identical enrichment of somatic mutations to those seen in our study clusters B, C, D, and E. Although Morimoto et al did not correlate their clusters with survival, they did report IPSS-R risk categories which showed the same trends in disease risk as with our DNAm cluster-risk groups. Both of the studies mentioned above, as well as our own, identified subgroups based on different inputs (gene expression, promoter-focused CpG methylation, non-promoter-focused CpG methylation (BSPP)), but all three identified similar trends in terms of the proportions and types of genetic abnormalities present in these subgroups. Perhaps it should be expected that higher-risk genetic abnormalities would be accompanied by activation of different gene expression pathways and epigenetic changes, however an interesting question arising from this finding is whether one is a consequence of the other (e.g. altered epigenome causing altered gene expression, or vice versa), and how these changes correlate with the differentiation state of the cell populations used to generate the data. Future studies using single-cell approaches could be fruitful for determining the contributions of different cell types to the signatures observed in each of the molecular subgroups identified in these studies.

Epigenetic Heterogeneity in TET2-Mutant MDS

One unexpected finding of our studies in Chapter 2 was the lack of a differential DNAm signature associated with *TET2* mutation. We observed only 21 total differentially methylated CpGs out of ~300,000 when comparing *TET2*-mutant vs. wild-type patients, and zero differences when comparing subsets of mutant vs. wild-type *within each cluster* as well. Given the known effects of *TET2* loss in model organisms and several studies of chronic myelo-monocytic leukemia patients which reported DNA hypermethylation in *TET2*-mutant cases, we thought we might find a similar phenotype in our patients (Meldi et al. 2015; Yamazaki et al. 2015). Interestingly, in the study initially identifying the relationship between *TET2* mutation and impaired hydroxymethylation in myeloid malignancies, there was a fairly large proportion of MDS cases that had global loss of 5hmC even in the absence of *TET2* mutation, which suggests that there may be alternative mechanisms in MDS which phenocopy *TET2* loss (Ko et al. 2010). In our study we did observe that Cluster D, which was enriched for *TET2*-mutant patients, displayed 5mC changes similar to those seen in CMML and mouse models of *TET2* loss, and notably, this cluster had a lower co-mutation burden compared with *TET2*-mutants in other clusters. There is some evidence in the literature that *TET2*-mutation in cooperation with co-mutations such as *NRAS*, *DNMT3A*, and others, can influence 5mC patterns, and in some contexts counteract the changes associated with *TET2* mutation alone (Muto et al. 2013; X. Zhang et al. 2016; Kunimoto et al. 2018). It is therefore likely that the highly co-mutated *TET2* mutants were diluting the global signal associated with *TET2* mutation across all

clusters in our cohort. It is plausible that the Cluster-D patients were in earlier stages of their disease where *TET2* mutation may still have a large influence on the epigenome, while patients in Clusters B and C were in later disease stages where cooperating lesions lead to further remodeling of the epigenome.

In future work studying the role of *TET2* in the disease course of MDS, it may be enlightening to study 5hmC and 5mC patterns in mutant and wild type patients as the disease progresses. One possible experiment to answer the question of where and when 5hmC/5mC alterations were most significant would be to identify patients who have clonal hematopoiesis, and determine their 5mC and 5hmC profiles at multiple time points as their disease progresses. Most patients with clonal hematopoiesis never develop myeloid malignancy, so a retrospective study of patients who progressed from clonal hematopoiesis to MDS or AML, would be most appropriate. It would be interesting to observe how 5mC and 5hmC patterns evolve as patients acquire further co-mutations. The sequence of genetic vs. epigenetic changes would also be interesting, as it is known that the epigenome changes with age, and epigenome changes could very well precede genetic changes. Since the majority of patients with clonal hematopoiesis harboring mutations in *DNMT3A* and *TET2* never progress to frank malignancy, one can imagine a scenario where an early mutation in these genes leads to 5mC alterations accumulating over time, which may set the stage for when a second mutation strikes to produce pre-malignancy. If we understood and were able to predict which epigenetic changes were associated with progression, we may be

able to target these pre-malignant 5mC changes using DNA methyltransferase inhibitors to actually *prevent* progression to malignancy. Further research is certainly warranted into the time-course of epigenetic changes in clonal hematopoiesis and MDS because inactivation of epigenetic regulating genes appear to be the initiators of these disorders, and disease prevention could be much more effective than treatment at more advanced stages.

Role of CD93 in MDS

A novel finding of our work in Chapter 2 was the discovery of the association of *CD93* methylation and expression with overall survival in our study as well as in a validation study (Gerstung et al. 2015). Notably we were able to relate methylation to expression by validating the relationships in publicly available datasets of both methylation and expression in the same samples which were derived from AML tumors in The Cancer Genome Atlas LAML study cohort. Now that we have identified the relevance of *CD93* in MDS outcome, it begs the question of what role it is playing in the disease. CD93, also known as the complement component 1 subcomponent Q receptor 1 (C1qR), is reportedly expressed on myeloid lineage cells including granulocytes, monocytes, and platelets, and also on hematopoietic stem and progenitor cells (both CD34+ and CD34- progenitors), as well as in endothelial cells (Kao et al. 2012). *CD93* has been reported to have functions in enhancing phagocytosis (Tenner 1998), promoting angiogenesis (Langenkamp et al. 2015; Lugano et al. 2018), and is associated with inflammatory diseases such as rheumatoid arthritis and various

allergies (Sigari et al. 2016). Perhaps most relevant to MDS, CD93 reportedly marks a subset of non-quiescent leukemia stem cells in MLL-rearranged AML and suppresses expression of the tumor suppressor *CDKN2B* (Iwasaki et al. 2015). Because CD93 is expressed on immature CD34+ and CD34- progenitors, we wondered if its association with outcome may be because it is a surrogate marker of the percentage of bone marrow blasts, which is a known predictor of outcome in MDS, however we did not observe a correlation between % blasts and CD93 expression. It is possible that similar to AML, MDS patients with high CD93 have higher amounts of non-quiescent HSPCs which produce larger amounts of immature cells, enhancing their disease risk, however that was not seen in the bone marrow blast counts in these patients. It is also possible that the role of CD93 in inflammation may be relevant, as autoimmunity is increasingly becoming recognized as a significant contributor to MDS pathobiology (Wolach and Stone 2016). Finally, it could very well be that differences in expression within the mononuclear cell compartment are due to different proportions of cells differentiating along the granulocytic-monocytic lineages which have high CD93 expression. It may be that these cells have a block in differentiation along this lineage and immature cells along this lineage are contributing to disease risk in some manner. This question could be addressed simply by performing flow-cytometric sorting of hematopoietic cell populations while simultaneously staining for cell-surface CD93 expression. One could determine the proportion of cells falling into each hematopoietic compartment and then quantify the amount of

CD93 positive cells falling into each of those compartments to determine which population was contributing the most to the CD93 signal in the bone marrow. There remain more questions unanswered than answered by our discovery, however CD93 may be an exciting new target to study for novel therapies and biomarkers in MDS.

Role of TET2 in Response to Azacitidine

In chapter 3 we explored how *TET2* loss may sensitize MDS patients towards DNA methyltransferase inhibitors like azacitidine (5-Aza), and our cell line studies provided several insights into potential mechanisms. One of the key findings of this work was that TET2KO cells down-regulate erythroid differentiation gene expression signatures at baseline, and 5-Aza treatment can selectively and disproportionately induce these genes to correct the baseline down-regulation. It is also important to note that even though the TET2KO cells down-regulated erythroid differentiation signatures at baseline, upon erythroid differentiation induction with a strong inducing agent, erythropoietin, TET2KO cells had no significant loss of differentiation capacity compared with WT cells, indicating that the effects of *TET2* loss on differentiation are subtle. This is somewhat in line with mouse models of *TET2* inactivation which reported subtle changes, such as myeloid biased differentiation and expansion of HSC populations, but never complete loss of differentiation capacity and outright malignancy (Moran-Crusio et al. 2011).

As mentioned in previous chapters, *TET2* mutation is an imperfect predictor of response to 5-Aza due to co-mutations which can have a negative impact on association with response (Bejar, Lord, et al. 2014a). Co-mutations can strongly influence the epigenetic and differentiation phenotype initiated by *TET2* loss as has been shown in animal models, and correlated with our results (for DNAm patterns) in Chapter 2 (Kunimoto et al. 2018; X. Zhang et al. 2016). We have shown a plausible mechanism whereby TET2 inactivation leads to down-regulation of differentiation-associated gene expression which can be targeted by 5-Aza, however the majority of TET2-mutated patients have co-mutations in a variety of other genes, and it remains to be seen how co-mutations impact the signatures we observed in our study. A particularly interesting co-mutation to study in future experiments would be *TET2* co-mutated with *ASXL1*, which was shown to predict resistance to 5-Aza in MDS patients. It would be interesting to observe how *ASXL1* impacted the DNA methylome, particularly at enhancers which we showed to be important for gene expression changes after 5-Aza. In Chapter 2 we showed that patients co-mutated for *ASXL1* and *TET2* were much more likely to be enriched in Cluster C, which had global hypomethylation compared to other clusters, and differential methylation changes were enriched at enhancer regions. It may be that *ASXL1* mutations remove the enhancer hypermethylation phenotype associated with *TET2* loss, thus making these patients resistant to 5-Aza therapy. This hypothesis could be further tested in studies of 5mC profiles in patients treated with 5-Aza and comparing the long term 5mC changes at

enhancers in *TET2*-alone vs. *TET2* and *ASXL1* co-mutant, which are both quite frequent in MDS patient populations. In general, if we can better understand the 5mC signatures that occur in responding patients before and after treatment, it may be easier to discover a signature that can be used as a biomarker capable of predicting response even in patients with diverse co-mutations. With the work in Chapter 2 we provide some insights into where these 5mC alterations are likely to occur, namely within enhancers of differentiation-associated genes. It is also quite possible that mechanisms other than 5mC and 5hmC alterations contribute to responses in a subset of patients, such as the amount of drug actually reaching the disease cells after metabolism, and it may be effective to have multiple predictors of response built into a single model capable of higher accuracy predictions.

An interesting question we were unable to answer with the studies in Chapter 3 was what role the population doubling time lengthening in TET2KO cells after 5-Aza exposure may have in response. Several studies of MDS patient clonal dynamics during 5-Aza treatment report that *TET2* mutant clones remain a proportional contributor to the bone marrow HSPC population during response to therapy, so we know that response is not driven by cytotoxic sensitivity to AZA. What remains to be seen is what proportion of mature blood cells arise from the *TET2* mutant vs. wild-type HSPCs. It is possible that 5-Aza disproportionately slows *TET2* mutant clone expansion in the bone marrow, allowing wild-type HSPCs to repopulate the bone marrow with differentiating progenitors that would

otherwise be crowded out by dysplastic *TET2* mutant progenitors, leading to enhanced response in patients harboring *TET2* mutation. This could also potentially simultaneously explain why *TET2* mutant clones are resistance to cytotoxic effects of 5-Aza, which requires active cell cycling for its maximal toxic effects. These questions could potentially be addressed in patient samples by performing targeted sequencing of HSPCs from bone marrow samples as well as mature cells in the peripheral blood in a time-series during treatment, and calculating the contribution of *TET2* mutated clones to the mature blood cell population during a response to 5-Aza. If the *TET2* mutant clone in HSPCs remains constant, but blood counts increase and have a lower proportion of *TET2* mutant allele burden, then we could conclude that the association with response is driven by the delayed or slowed expansion of mutant HSPCs in the bone marrow as opposed to differentiation induction of the mutant clone.

A major limitation of our studies in Chapter 3 on 5mC and expression dynamics during 5-Aza treatment is that our methylation profiling technique, BSPP, did not have sufficient coverage of all erythroid-differentiation gene enhancers to make definitive conclusions regarding differences in 5mC dynamics at these regions. Future studies on this topic should be performed using targeted methods that specifically target differentiation-associated enhancers or more comprehensive methods like whole-genome bisulfite sequencing, so that we may achieve a more accurate understanding of the dynamic 5mC changes at these

locations during treatment and how they may relate to differentiation and treatment response.

References

Abdel-Wahab, Omar, Ann Mullally, Cyrus Hedvat, Guillermo Garcia-Manero, Jay Patel, Martha Wadleigh, Sebastien Malinge, JinJuan Yao, Outi Kilpivaara, Rukhmi Bhat, Kety Huberman, Sabrena Thomas, Igor Dolgalev, Adriana Heguy, Elisabeth Paietta, Michelle M Le Beau, Miloslav Beran, Martin S Tallman, Benjamin L Ebert, Hagop M Kantarjian, Richard M Stone, D Gary Gilliland, John D Crispino, and Ross L Levine. 2009. “Genetic Characterization of TET1, TET2, and TET3 Alterations in Myeloid Malignancies.” *Blood* 114 (1): 144–47. https://doi.org/10.1182/blood-2009-03-210039.

Adès, Lionel, Raphael Itzykson, and Pierre Fenaux. 2014. “Myelodysplastic Syndromes.” *The Lancet* 383 (9936): 2239–52. https://doi.org/10.1016/S0140-6736(13)61901-7.

Adolfsson, Jörgen, Robert Månsson, Natalija Buza-Vidas, Anne Hultquist, Karina Liuba, Christina T. Jensen, David Bryder, Liping Yang, Ole-Johan Borge, Lina A. M. Thoren, Kristina Anderson, Ewa Sitnicka, Yutaka Sasaki, Mikael Sigvardsson, and Sten Eirik W. Jacobsen. 2005. “Identification of Flt3+ Lympho-Myeloid Stem Cells Lacking Erythro-Megakaryocytic Potential: A Revised Road Map for Adult Blood Lineage Commitment.” *Cell* 121 (2): 295–306. https://doi.org/10.1016/j.cell.2005.02.013.

Agathocleous, Michalis, Corbin E. Meacham, Rebecca J. Burgess, Elena Piskounova, Zhiyu Zhao, Genevieve M. Crane, Brianna L. Cowin, Emily Bruner, Malea M. Murphy, Weina Chen, Gerald J. Spangrude, Zeping Hu, Ralph J. DeBerardinis, and Sean J. Morrison. 2017. “Ascorbate Regulates Haematopoietic Stem Cell Function and Leukaemogenesis.” *Nature* 549 (7673): 476. https://doi.org/10.1038/nature23876.

Akashi, Koichi, David Traver, Toshihiro Miyamoto, and Irving L. Weissman. 2000. “A Clonogenic Common Myeloid Progenitor That Gives Rise to All Myeloid Lineages.” *Nature* 404 (6774): 193. https://doi.org/10.1038/35004599.

Álvarez-Errico, Damiana, Roser Vento-Tormo, Michael Sieweke, and Esteban Ballestar. 2015. “Epigenetic Control of Myeloid Cell Differentiation, Identity and Function.” *Nature Reviews Immunology* 15 (1): 7–17. https://doi.org/10.1038/nri3777.

Andersson, Robin, Claudia Gebhard, Irene Miguel-Escalada, Ilka Hoof, Jette Bornholdt, Mette Boyd, Yun Chen, Xiaobei Zhao, Christian Schmidl, Takahiro Suzuki, Evgenia Ntini, Erik Arner, Eivind Valen, Kang Li, Lucia Schwarzfischer, Dagmar Glatz, Johanna Raithel, Berit Lilje, Nicolas Rapin, Frederik Otzen Bagger, Mette Jørgensen, Peter Refsing Andersen, Nicolas Bertin, Owen Rackham, A Maxwell Burroughs, J Kenneth Baillie, Yuri Ishizu, Yuri Shimizu, Erina Furuhata, Shiori Maeda, Yutaka Negishi,

Christopher J Mungall, Terrence F Meehan, and Timo Lassmann. 2014. “An Atlas of Active Enhancers across Human Cell Types and Tissues.” *Nature* 507. https://doi.org/10.1038/nature12787.

Azad, Nilofer S., Anthony el-Khoueiry, Jun Yin, Ann L. Oberg, Patrick Flynn, Douglas Adkins, Anup Sharma, Daniel J. Weisenberger, Thomas Brown, Prakriti Medvari, Peter A. Jones, Hariharan Easwaran, Ihab Kamel, Nathan Bahary, George Kim, Joel Picus, Henry C. Pitot, Charles Erlichman, Ross Donehower, Hui Shen, Peter W. Laird, Richard Piekarz, Stephen Baylin, and Nita Ahuja. 2017. “Combination Epigenetic Therapy in Metastatic Colorectal Cancer (MCRC) with Subcutaneous 5-Azacitidine and Entinostat: A Phase 2 Consortium/Stand Up 2 Cancer Study.” *Oncotarget* 8 (21): 35326–38. https://doi.org/10.18632/oncotarget.15108.

Barretina, Jordi, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A. Margolin, Sungjoon Kim, Christopher J. Wilson, Joseph Lehár, Gregory V. Kryukov, Dmitriy Sonkin, Anupama Reddy, Manway Liu, Lauren Murray, Michael F. Berger, John E. Monahan, Paula Morais, Jodi Meltzer, Adam Korejwa, Judit Jané-Valbuena, Felipa A. Mapa, Joseph Thibault, Eva Bric-Furlong, Pichai Raman, Aaron Shipway, Ingo H. Engels, Jill Cheng, Guoying K. Yu, Jianjun Yu, Peter Aspesi, Melanie de Silva, Kalpana Jagtap, Michael D. Jones, Li Wang, Charles Hatton, Emanuele Palescandolo, Supriya Gupta, Scott Mahan, Carrie Sougnez, Robert C. Onofrio, Ted Liefeld, Laura MacConaill, Wendy Winckler, Michael Reich, Nanxin Li, Jill P. Mesirov, Stacey B. Gabriel, Gad Getz, Kristin Ardlie, Vivien Chan, Vic E. Myer, Barbara L. Weber, Jeff Porter, Markus Warmuth, Peter Finan, Jennifer L. Harris, Matthew Meyerson, Todd R. Golub, Michael P. Morrissey, William R. Sellers, Robert Schlegel, and Levi A. Garraway. 2012. “The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity.” *Nature* 483 (7391): 603–7. https://doi.org/10.1038/nature11003.

Bauer, Daniel E, Matthew C Canver, and Stuart H Orkin. 2015. “Generation of Genomic Deletions in Mammalian Cell Lines via CRISPR/Cas9.” *Journal of Visualized Experiments : JoVE*, no. 95 (January). https://doi.org/10.3791/52118.

Bejar, Rafael. 2017. “Implications of Molecular Genetic Diversity in Myelodysplastic Syndromes.” *Current Opinion in Hematology*. https://doi.org/10.1097/MOH.0000000000000313.

Bejar, Rafael, Ross Levine, and Benjamin L. Ebert. 2011. “Unraveling the Molecular Pathophysiology of Myelodysplastic Syndromes.” *Journal of Clinical Oncology*. https://doi.org/10.1200/JCO.2010.31.1175.

Bejar, Rafael, Allegra Lord, Kristen Stevenson, Michal Bar-Natan, Albert Pérez-Ladaga, Jacques Zaneveld, Hui Wang, Bennett Caughey, Petar Stojanov, Gad Getz, Guillermo Garcia-Manero, Hagop Kantarjian, Rui Chen, Richard M Stone, Donna Neuberg, David P Steensma, Benjamin L Ebert, International Vidaza High-Risk MDS Survival Study Group, DP. Steensma, MR. Baer, JL. Slack, P. Fenaux, L. Ades, G. Garcia-Manero, Groupe Francophone des Myelodysplasies(GFM), Eastern Cooperative Oncology Group (ECOG) and North American Leukemia intergroup, R. Bejar, K. Stevenson, O. Abdel-Wahab, Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium, T. Haferlach, Y. Nagata, V. Grossmann, Myelodysplastic Syndrome-003 Study Investigators, C. Stresemann, F. Lyko, M. Ko, Y. Huang, AM. Jankowska, J. Yamazaki, R. Taby, A. Vasanthakumar, Groupe Francophone des Myelodysplasies (GFM), DM. Murphy, R. Bejar, K. Stevenson, H. Li, R. Durbin, MA. DePristo, E. Banks, R. Poplin, K. Cibulskis, MS. Lawrence, SL. Carter, K. Moran-Crusio, L. Reavie, A. Shih, R. Bejar, KE. Stevenson, BA. Caughey, K. Yoshida, M. Sanada, Y. Shiraishi, L. Shen, H. Kantarjian, Y. Guo, O. Abdel-Wahab, M. Adli, LM. LaFave, H. Kunimoto, Y. Fukuchi, M. Sakurai, Z. Li, X. Cai, CL. Cai, M. Jädersten, L. Saft, A. Pellagatti, MJ. Walter, D. Shen, L. Ding, MJ. Walter, D. Shen, J. Shao, A. Valencia, E. Masala, A. Rossi, RZ. Mahfouz, A. Jankowska, Q. Ebrahem, Y. Sauntharajah, T. Qin, R. Castoro, S. El Ahdab, C. Bally, L. Adès, A. Renneville, C. Alhan, TM. Westers, LH. van der Helm, A. Valencia, E. Masala, A. Rossi, M. Breccia, G. Loglisci, and L. Cannella. 2014a. “TET2 Mutations Predict Response to Hypomethylating Agents in Myelodysplastic Syndrome Patients.” *Blood* 124 (17): 2705–12. https://doi.org/10.1182/blood-2014-06-582809.

Bejar, Rafael, Allegra Lord, Kristen Stevenson, Michal Bar-Natan, Albert Pérez-Ladaga, Jacques Zaneveld, Hui Wang, Bennett Caughey, Petar Stojanov, Gad Getz, Guillermo Garcia-Manero, Hagop Kantarjian, Rui Chen, Richard M Stone, Donna Neuberg, David P Steensma, Benjamin L Ebert, International Vidaza High-Risk MDS Survival Study Group, DP. Steensma, MR. Baer, JL. Slack, P. Fenaux, L. Ades, G. Garcia-Manero, Groupe Francophone des Myelodysplasies(GFM), Eastern Cooperative Oncology Group (ECOG) and North American Leukemia intergroup, R. Bejar, K. Stevenson, O. Abdel-Wahab, Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium, T. Haferlach, Y. Nagata, V. Grossmann, Myelodysplastic Syndrome-003 Study Investigators, C. Stresemann, F. Lyko, M. Ko, Y. Huang, AM. Jankowska, J. Yamazaki, R. Taby, A. Vasanthakumar, Groupe Francophone des Myelodysplasies (GFM), DM. Murphy, R. Bejar, K. Stevenson, H. Li, R. Durbin, MA. DePristo, E. Banks, R. Poplin, K. Cibulskis, MS. Lawrence, SL. Carter, K. Moran-Crusio, L. Reavie, A. Shih, R. Bejar, KE. Stevenson, BA. Caughey, K. Yoshida, M. Sanada, Y. Shiraishi, L. Shen, H. Kantarjian, Y. Guo, O. Abdel-

Wahab, M. Adli, LM. LaFave, H. Kunimoto, Y. Fukuchi, M. Sakurai, Z. Li, X. Cai, CL. Cai, M. Jädersten, L. Saft, A. Pellagatti, MJ. Walter, D. Shen, L. Ding, MJ. Walter, D. Shen, J. Shao, A. Valencia, E. Masala, A. Rossi, RZ. Mahfouz, A. Jankowska, Q. Ebrahem, Y. Sauntharajah, T. Qin, R. Castoro, S. El Ahdab, C. Bally, L. Adès, A. Renneville, C. Alhan, TM. Westers, LH. van der Helm, A. Valencia, E. Masala, A. Rossi, M. Breccia, G. Loglisci, and L. Cannella. 2014b. “TET2 Mutations Predict Response to Hypomethylating Agents in Myelodysplastic Syndrome Patients.” *Blood* 124 (17): 2705–12. https://doi.org/10.1182/blood-2014-06-582809.

Bejar, Rafael, Kristen Stevenson, Omar Abdel-Wahab, Naomi Galili, Björn Nilsson, Guillermo Garcia-Manero, Hagop Kantarjian, Azra Raza, Ross L Levine, Donna Neuberg, and Benjamin L Ebert. 2011. “Clinical Effect of Point Mutations in Myelodysplastic Syndromes.” *The New England Journal of Medicine* 364 (26): 2496–2506. https://doi.org/10.1056/NEJMoa1013343.

Bejar, Rafael, Kristen E. Stevenson, Bennett Caughey, R. Coleman Lindsley, Brenton G. Mar, Petar Stojanov, Gad Getz, David P. Steensma, Jerome Ritz, Robert Soiffer, Joseph H. Antin, Edwin Alyea, Philippe Armand, Vincent Ho, John Koreth, Donna Neuberg, Corey S. Cutler, and Benjamin L. Ebert. 2014. “Somatic Mutations Predict Poor Outcome in Patients with Myelodysplastic Syndrome after Hematopoietic Stem-Cell Transplantation.” *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 32 (25): 2691–98. https://doi.org/10.1200/JCO.2013.52.3381.

Benjamini, Yoav, and Hochberg Yosef. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society* 57 (1): 289–300.

Bröske, Ann-Marie, Lena Vockentanz, Shabnam Kharazi, Matthew R. Huska, Elena Mancini, Marina Scheller, Christiane Kuhl, Andreas Enns, Marco Prinz, Rudolf Jaenisch, Claus Nerlov, Achim Leutz, Miguel A. Andrade-Navarro, Sten Eirik W. Jacobsen, and Frank Rosenbauer. 2009. “DNA Methylation Protects Hematopoietic Stem Cell Multipotency from Myeloerythroid Restriction.” *Nature Genetics* 41 (11): 1207–15. https://doi.org/10.1038/ng.463.

Brunet, J P, P Tamayo, T R Golub, and J P Mesirov. 2004. “Metagenes and Molecular Pattern Discovery Using Matrix Factorization.” *Proc Natl Acad Sci U S A* 101 (12): 4164–69. https://doi.org/10.1073/pnas.0308531101.

Canver, Matthew C, Daniel E Bauer, Abhishek Dass, Yvette Y Yien, Jacky Chung, Takeshi Masuda, Takahiro Maeda, Barry H Paw, and Stuart H Orkin. 2014. “Characterization of Genomic Deletion Efficiency Mediated by Clustered

Regularly Interspaced Palindromic Repeats (CRISPR)/Cas9 Nuclease System in Mammalian Cells." *The Journal of Biological Chemistry* 289 (31): 21312–24. https://doi.org/10.1074/jbc.M114.564625.

Cazzola, M., M. G. Della Porta, and L. Malcovati. 2013. "The Genetic Basis of Myelodysplasia and Its Clinical Relevance." *Blood* 122 (25): 4021–34. https://doi.org/10.1182/blood-2013-09-381665.

Chen, Lei-Lei, Huai-Peng Lin, Wen-Jie Zhou, Chen-Xi He, Zhi-Yong Zhang, Zhou-Li Cheng, Jun-Bin Song, Peng Liu, Xin-Yu Chen, Yu-Kun Xia, Xiu-Fei Chen, Ren-Qiang Sun, Jing-Ye Zhang, Yi-Ping Sun, Lei Song, Bing-Jie Liu, Rui-Kai Du, Chen Ding, Fei Lan, Sheng-Lin Huang, Feng Zhou, Suling Liu, Yue Xiong, Dan Ye, and Kun-Liang Guan. 2018. "SNIP1 Recruits TET2 to Regulate C-MYC Target Genes and Cellular DNA Damage Response." *Cell Reports* 25 (6): 1485-1500.e4. https://doi.org/10.1016/j.celrep.2018.10.028.

Chen, Pao-Yang, Suhua Feng, Jong Wha Joanne Joo, Steve E. Jacobsen, and Matteo Pellegrini. 2011. "A Comparative Analysis of DNA Methylation across Human Embryonic Stem Cell Lines." *Genome Biology* 12 (7): R62. https://doi.org/10.1186/gb-2011-12-7-r62.

Chen, Tong, Yanan You, Hua Jiang, and Zack Z. Wang. 2017. "Epithelial-Mesenchymal Transition (EMT): A Biological Process in the Development, Stem Cell Differentiation, and Tumorigenesis." *Journal of Cellular Physiology* 232 (12): 3261–72. https://doi.org/10.1002/jcp.25797.

Cilloni, Daniela, Enrico Gottardi, Francesca Messa, Milena Fava, Patrizia Scaravaglio, Marilena Bertini, Mauro Girotto, Carlo Marinone, Dario Ferrero, Andrea Gallamini, Alessandro Levis, Giuseppe Saglio, and Piedmont Study Group on Myleodysplastic Syndromes. 2003. "Significant Correlation between the Degree of WT1 Expression and the International Prognostic Scoring System Score in Patients with Myelodysplastic Syndromes." *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 21 (10): 1988–95. https://doi.org/10.1200/JCO.2003.10.503.

Cimmino, Luisa, Igor Dolgalev, Yubao Wang, Akihide Yoshimi, Gaëlle H. Martin, Jingjing Wang, Victor Ng, Bo Xia, Matthew T. Witkowski, Marisa Mitchell-Flack, Isabella Grillo, Sofia Bakogianni, Delphine Ndiaye-Lobry, Miguel Torres Martín, Maria Guillamot, Robert S. Banh, Mingjiang Xu, Maria E. Figueroa, Ross A. Dickins, Omar Abdel-Wahab, Christopher Y. Park, Aristotelis Tsirigos, Benjamin G. Neel, and Iannis Aifantis. 2017. "Restoration of TET2 Function Blocks Aberrant Self-Renewal and Leukemia Progression." *Cell*. https://doi.org/10.1016/j.cell.2017.07.032.

Connolly, Roisin M., Huili Li, Rachel C. Jankowitz, Zhe Zhang, Michelle A. Rudek, Stacie C. Jeter, Shannon A. Slater, Penny Powers, Antonio C. Wolff, John H. Fetting, Adam Brufsky, Richard Piekarz, Nita Ahuja, Peter W. Laird, Hui Shen, Daniel J. Weisenberger, Leslie Cope, James G. Herman, George Somlo, Agustin A. Garcia, Peter A. Jones, Stephen B. Baylin, Nancy E. Davidson, Cynthia A. Zahnow, and Vered Stearns. 2017. “Combination Epigenetic Therapy in Advanced Breast Cancer with 5-Azacitidine and Entinostat: A Phase II National Cancer Institute/Stand Up to Cancer Study.” *Clinical Cancer Research* 23 (11): 2691–2701. https://doi.org/10.1158/1078-0432.CCR-16-1729.

Corces-Zimmerman, M Ryan, Wan-Jen Hong, Irving L Weissman, Bruno C Medeiros, and Ravindra Majeti. 2014. “Preleukemic Mutations in Human Acute Myeloid Leukemia Affect Epigenetic Regulators and Persist in Remission.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (7): 2548–53. https://doi.org/10.1073/pnas.1324297111.

Delhommeau, François, Sabrina Dupont, Véronique Della Valle, Chloé James, Severine Trannoy, Aline Massé, Olivier Kosmider, Jean-Pierre Le Couedic, Fabienne Robert, Antonio Alberdi, Yann Lécluse, Isabelle Plo, François J. Dreyfus, Christophe Marzac, Nicole Casadevall, Catherine Lacombe, Serge P. Romana, Philippe Dessen, Jean Soulier, Franck Viguié, Michaela Fontenay, William Vainchenker, and Olivier A. Bernard. 2009. “Mutation in TET2 in Myeloid Cancers.” *New England Journal of Medicine* 360 (22): 2289–2301. https://doi.org/10.1056/NEJMoa0810069.

Deng, Jie, Robert Shoemaker, Bin Xie, Athurva Gore, Emily M LeProust, Jessica Antosiewicz-Bourget, Dieter Egli, Nimet Maherali, In-Hyun Park, Junying Yu, George Q Daley, Kevin Eggan, Konrad Hochedlinger, James Thomson, Wei Wang, Yuan Gao, and Kun Zhang. 2009. “Targeted Bisulfite Sequencing Reveals Changes in DNA Methylation Associated with Nuclear Reprogramming.” *Nature Biotechnology* 27 (4): 353–60. https://doi.org/10.1038/nbt.1530.

Diep, Dinh, Nongluk Plongthongkum, Athurva Gore, Ho-Lim Fung, Robert Shoemaker, and Kun Zhang. 2012. “Library-Free Methylation Sequencing with Bisulfite Padlock Probes.” *Nature Methods* 9 (3): 270–72. https://doi.org/10.1038/nmeth.1871.

Doi, Akiko, In-Hyun Park, Bo Wen, Peter Murakami, Martin J Aryee, Rafael Irizarry, Brian Herb, Christine Ladd-Acosta, Junsung Rho, Sabine Loewer, Justine Miller, Thorsten Schlaeger, George Q Daley, and Andrew P Feinberg. 2009. “Differential Methylation of Tissue- and Cancer-Specific CpG Island Shores Distinguishes Human Induced Pluripotent Stem Cells, Embryonic Stem

Cells and Fibroblasts.” *Nature Genetics* 41 (12): 1350–53. https://doi.org/10.1038/ng.471.

Fandy, Tamer E, James G Herman, Patrick Kerns, Anchalee Jiemjit, Elizabeth A Sugar, Si-Ho Choi, Allen S Yang, Timothy Aucott, Tianna Dauses, Rosalie Odchimar-Reissig, Jonathan Licht, Melanie J McConnell, Chris Nasrallah, Marianne K H Kim, Weijia Zhang, Yezou Sun, Anthony Murgo, Igor Espinoza-Delgado, Katharine Oteiza, Ibitayo Owoeye, Lewis R Silverman, Steven D Gore, and Hetty E Carraway. 2009. “Early Epigenetic Changes and DNA Damage Do Not Predict Clinical Response in an Overlapping Schedule of 5-Azacytidine and Entinostat in Patients with Myeloid Malignancies.” *Blood* 114 (13): 2764–73. https://doi.org/10.1182/blood-2009-02-203547.

Farlik, Matthias, Florian Halbritter, Fabian Müller, Fizzah A. Choudry, Peter Ebert, Johanna Klughammer, Samantha Farrow, Antonella Santoro, Valerio Ciaurro, Anthony Mathur, Rakesh Uppal, Hendrik G. Stunnenberg, Willem H. Ouwehand, Elisa Laurenti, Thomas Lengauer, Mattia Frontini, and Christoph Bock. 2016. “DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation.” *Cell Stem Cell* 19 (6): 808–22. https://doi.org/10.1016/j.stem.2016.10.019.

Fenaux, Pierre, and Lionel Ades. 2009. “Review of Azacitidine Trials in Intermediate-2-and High-Risk Myelodysplastic Syndromes.” *Leukemia Research*. https://doi.org/10.1016/S0145-2126(09)70227-9.

Fenaux, Pierre, Ghulam J Mufti, Eva Hellstrom-Lindberg, Valeria Santini, Carlo Finelli, Aristoteles Giagounidis, Robert Schoch, Norbert Gattermann, Guillermo Sanz, Alan List, Steven D Gore, John F Seymour, John M Bennett, John Byrd, Jay Backstrom, Linda Zimmerman, David McKenzie, Cl Beach, and Lewis R Silverman. 2009. “Efficacy of Azacitidine Compared with That of Conventional Care Regimens in the Treatment of Higher-Risk Myelodysplastic Syndromes: A Randomised, Open-Label, Phase III Study.” *The Lancet. Oncology* 10 (3): 223–32. https://doi.org/10.1016/S1470-2045(09)70003-8.

Figueroa, Maria E, Omar Abdel-Wahab, Chao Lu, Patrick S Ward, Jay Patel, Alan Shih, Yushan Li, Neha Bhagwat, Aparna Vasanthakumar, Hugo F Fernandez, Martin S Tallman, Zhuoxin Sun, Kristy Wolniak, Justine K Peeters, Wei Liu, Sung E Choe, Valeria R Fantin, Elisabeth Paietta, Bob Löwenberg, Jonathan D Licht, Lucy A Godley, Ruud Delwel, Peter J M Valk, Craig B Thompson, Ross L Levine, and Ari Melnick. 2010. “Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation.” *Cancer Cell* 18 (6): 553–67. https://doi.org/10.1016/j.ccr.2010.11.015.

Figueroa, Maria E, Sanne Lugthart, Yushan Li, Claudia Erpelinck-Verschueren, Xutao Deng, Paul J Christos, Elizabeth Schifano, James Booth, Wim van Putten, Lucy Skrabanek, Fabien Campagne, Madhu Mazumdar, John M Greally, Peter J M Valk, Bob Löwenberg, Ruud Delwel, and Ari Melnick. 2010a. “DNA Methylation Signatures Identify Biologically Distinct Subtypes in Acute Myeloid Leukemia.” *Cancer Cell* 17 (1): 13–27. https://doi.org/10.1016/j.ccr.2009.11.020.

———. 2010b. “DNA Methylation Signatures Identify Biologically Distinct Subtypes in Acute Myeloid Leukemia.” *Cancer Cell* 17 (1): 13–27. https://doi.org/10.1016/j.ccr.2009.11.020.

Figueroa, Maria E, Lucy Skrabanek, Yushan Li, Anchalee Jiemjit, Tamer E Fandy, Elisabeth Paietta, Hugo Fernandez, Martin S Tallman, John M Greally, Hetty Carraway, Jonathan D Licht, Steven D Gore, Ari Melnick, C. Plass, JG. Herman, SB. Baylin, PA. Jones, SB. Baylin, M. Tessema, F. Langer, J. Dingemann, A. Ganser, H. Kreipe, U. Lehmann, K. Brakensiek, F. Langer, B. Schlegelberger, H. Kreipe, U. Lehmann, DH. Christiansen, MK. Andersen, J. Pedersen-Bjergaard, A. Aggerholm, MS. Holm, P. Guldberg, LH. Olesen, P. Hokland, E. Griffiths, S. Gore, LR. Silverman, EP. Demakos, BL. Peterson, LR. Silverman, DR. McKenzie, BL. Peterson, P. Fenaux, P. Fenaux, GJ. Mufti, E. Hellstrom-Lindberg, Y. Jiang, A. Dunbar, LP. Gondek, K. Raj, A. John, A. Ho, SD. Gore, S. Baylin, E. Sugar, H. Kantarjian, Y. Oki, G. Garcia-Manero, AR. Karpf, BC. Moore, TO. Ririe, DA. Jones, W-G. Zhu, T. Hileman, Y. Ke, SS. Palii, BO. Van Emburgh, UT. Sankpal, KD. Brown, KD. Robertson, HK. Kuo, JD. Griffith, KN. Kreuzer, B. Khulan, R. Thompson, K. Ye, ME. Figueroa, M. Reimers, RF. Thompson, RR. Selzer, TA. Richmond, NJ. Pofahl, RF. Thompson, M. Reimers, B. Khulan, R. Edgar, M. Domrachev, AE. Lash, M. Ehrich, MR. Nelson, P. Stanssens, R. Gentleman, V. Carey, D. Bates, AC. Culhane, J. Thioulouse, G. Perriere, DG. Higgins, A. Smit, R. Hubley, P. Green, O. Elemento, N. Slonim, S. Tavazoie, D. Karolchik, RM. Kuhn, R. Baertsch, P. Polak, E. Domany, LR. Silverman, A. Verma, R. Odchimar-Reissig, TE. Fandy, JG. Herman, K. Kerns, C. Stresemann, I. Bokelmann, U. Mahlknecht, F. Lyko, FS. Harman, CJ. Nicol, HE. Marin, JM. Ward, FJ. Gonzalez, JM. Peters, EE. Cameron, KE. Bachman, S. Myohanen, JG. Herman, SB. Baylin, C. Mund, B. Hackanson, C. Stresemann, M. Lubbert, and F. Lyko. 2009. “MDS and Secondary AML Display Unique Patterns and Abundance of Aberrant DNA Methylation.” *Blood* 114 (16): 3448–58. https://doi.org/10.1182/blood-2009-01-200519.

Folia haematol (Leipz). 1907. Akademische Verlagsgesellschaft Geest & Portig k.-G.

Garcia-Manero, Guillermo. 2008. “Demethylating Agents in Myeloid Malignancies.” *Current Opinion in Oncology*. https://doi.org/10.1097/CCO.0b013e328313699c.

Garcia-Manero, Guillermo, Steven D Gore, Christopher Cogle, Renee Ward, Tao Shi, Kyle J Macbeth, Eric Laille, Heidi Giordano, Sarah Sakoian, Elias Jabbour, Hagop Kantarjian, and Barry Skikne. 2011. “Phase I Study of Oral Azacitidine in Myelodysplastic Syndromes, Chronic Myelomonocytic Leukemia, and Acute Myeloid Leukemia.” *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 29 (18): 2521–27. https://doi.org/10.1200/JCO.2010.34.4226.

Gelsi-Boyer, Véronique, Virginie Trouplin, José Adélaïde, Julien Bonansea, Nathalie Cervera, Nadine Carbuccia, Arnaud Lagarde, Thomas Prebet, Meyer Nezri, Danielle Sainty, Sylviane Olschwang, Luc Xerri, Max Chaffanet, Marie-Joëlle Mozziconacci, Norbert Vey, and Daniel Birnbaum. 2009. “Mutations of Polycomb-Associated Gene ASXL1 in Myelodysplastic Syndromes and Chronic Myelomonocytic Leukaemia.” *British Journal of Haematology* 145 (6): 788–800. https://doi.org/10.1111/j.1365-2141.2009.07697.x.

Gerstung, Moritz, Andrea Pellagatti, Luca Malcovati, Aristoteles Giagounidis, Matteo G. Della Porta, Martin Jädersten, Hamid Dolatshad, Amit Verma, Nicholas C.P. Cross, Paresh Vyas, Sally Killick, Eva Hellström-Lindberg, Mario Cazzola, Elli Papaemmanuil, Peter J. Campbell, and Jacqueline Boultwood. 2015. “Combining Gene Mutation with Gene Expression Data Improves Outcome Prediction in Myelodysplastic Syndromes.” *Nature Communications* 6. https://doi.org/10.1038/ncomms6901.

Giagounidis, A. a. N., A. Kulasekararaj, U. Germing, R. Radkowski, S. Haase, P. Petersen, G. Göhring, G. Büsche, C. Aul, G. J. Mufti, and U. Platzbecker. 2012. “Long-Term Transfusion Independence in Del(5q) MDS Patients Who Discontinue Lenalidomide.” *Leukemia* 26 (4): 855–58. https://doi.org/10.1038/leu.2011.268.

Greenberg, Peter L, Heinz Tuechler, Julie Schanz, Guillermo Sanz, Guillermo Garcia-Manero, Francesc Solé, John M Bennett, David Bowen, Pierre Fenaux, Francois Dreyfus, Hagop Kantarjian, Andrea Kuendgen, Alessandro Levis, Luca Malcovati, Mario Cazzola, Jaroslav Cermak, Christa Fonatsch, Michelle M Le Beau, Marilyn L Slovak, Otto Krieger, Michael Luebbert, Jaroslaw Maciejewski, Silvia M M Magalhaes, Yasushi Miyazaki, Michael Pfeilstöcker, Mikkael Sekeres, Wolfgang R Sperr, Reinhard Stauder, Sudhir Tauro, Peter Valent, Teresa Vallespi, Arjan A van de Loosdrecht, Ulrich Germing, and Detlef Haase. 2012. “Revised International

Prognostic Scoring System for Myelodysplastic Syndromes.” *Blood* 120 (12): 2454–65. https://doi.org/10.1182/blood-2012-03-420489.

Grövdal, M, M Karimi, M Tobiasson, L Reinius, M Jansson, K Ekwall, J Ungerstedt, J Kere, D Greco, and E Hellström-Lindberg. 2014. “Azacitidine Induces Profound Genome-Wide Hypomethylation in Primary Myelodysplastic Bone Marrow Cultures but May Also Reduce Histone Acetylation.” *Leukemia* 28 (2): 411–13. https://doi.org/10.1038/leu.2013.265.

Gu, Tianpeng, Xueqiu Lin, Sean M Cullen, Min Luo, Mira Jeong, Marcos Estecio, Jianjun Shen, Swanand Hardikar, Deqiang Sun, Jianzhong Su, Danielle Rux, Anna Guzman, Minjung Lee, Lei Stanley Qi, Jia-Jia Chen, Michael Kyba, Yun Huang, Taiping Chen, Wei Li, and Margaret A Goodell. 2018. “DNMT3A and TET1 Cooperate to Regulate Promoter Epigenetic Landscapes in Mouse Embryonic Stem Cells.” *Genome Biology* 19 (1): 88. https://doi.org/10.1186/s13059-018-1464-7.

Guo, Shicheng, Dinh Diep, Nongluk Plongthongkum, Ho-Lim Fung, Kang Zhang, and Kun Zhang. 2017. “Identification of Methylation Haplotype Blocks Aids in Deconvolution of Heterogeneous Tissue Samples and Tumor Tissue-of-Origin Mapping from Plasma DNA.” *Nature Genetics*, March. https://doi.org/10.1038/ng.3805.

Haferlach, T, Y Nagata, V Grossmann, Y Okuno, U Bacher, G Nagae, S Schnittger, M Sanada, A Kon, T Alpermann, K Yoshida, A Roller, N Nadarajah, Y Shiraishi, Y Shiozawa, K Chiba, H Tanaka, H P Koeffler, H-U Klein, M Dugas, H Aburatani, A Kohlmann, S Miyano, C Haferlach, W Kern, and S Ogawa. 2014. “Landscape of Genetic Lesions in 944 Patients with Myelodysplastic Syndromes.” *Leukemia* 28 (2): 241–47. https://doi.org/10.1038/leu.2013.336.

Hannum, Gregory, Justin Guinney, Ling Zhao, Li Zhang, Guy Hughes, Srini Vas Sadda, Brandy Klotzle, Marina Bibikova, Jian Bing Fan, Yuan Gao, Rob Deconde, Menzies Chen, Indika Rajapakse, Stephen Friend, Trey Ideker, and Kang Zhang. 2013. “Genome-Wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates.” *Molecular Cell* 49 (2): 359–67. https://doi.org/10.1016/j.molcel.2012.10.016.

Hark, Amy T., Christopher J. Schoenherr, David J. Katz, Robert S. Ingram, John M. Levorse, and Shirley M. Tilghman. 2000. “CTCF Mediates Methylation-Sensitive Enhancer-Blocking Activity at the H19/Igf2 Locus.” *Nature* 405 (6785): 486. https://doi.org/10.1038/35013106.

Hasegawa, N, M Oshima, G Sashida, H Matsui, S Koide, A Saraya, C Wang, T Muto, K Takane, A Kaneda, K Shimoda, C Nakaseko, K Yokote, and A

Iwama. 2017. “Impact of Combinatorial Dysfunctions of Tet2 and Ezh2 on the Epigenome in the Pathogenesis of Myelodysplastic Syndrome.” *Leukemia* 31 (4): 861–71. https://doi.org/10.1038/leu.2016.268.

He, Y.-F., B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.-X. Song, K. Zhang, C. He, and G.-L. Xu. 2011a. “Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA.” *Science* 333 (6047): 1303–1307. https://doi.org/10.1126/science.1210944.

———. 2011b. “Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA.” *Science* 333 (6047): 1303–7. https://doi.org/10.1126/science.1210944.

Hermann, Andrea, Rachna Goyal, and Albert Jeltsch. 2004. “The Dnmt1 DNA-(Cytosine-C5)-Methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites.” *Journal of Biological Chemistry* 279 (46): 48350–59. https://doi.org/10.1074/jbc.M403427200.

Hon, Gary C., Chun Xiao Song, Tingting Du, Fulai Jin, Siddarth Selvaraj, Ah Young Lee, Chia An Yen, Zhen Ye, Shi Qing Mao, Bang An Wang, Samantha Kuan, Lee E. Edsall, Boxuan Simen Zhao, Guo Liang Xu, Chuan He, and Bing Ren. 2014. “5mC Oxidation by Tet2 Modulates Enhancer Activity and Timing of Transcriptome Reprogramming during Differentiation.” *Molecular Cell* 56 (2): 286–97. https://doi.org/10.1016/j.molcel.2014.08.026.

Irizarry, Rafael A, Christine Ladd-Acosta, Bo Wen, Zhijin Wu, Carolina Montano, Patrick Onyango, Hengmi Cui, Kevin Gabo, Michael Rongione, Maree Webster, Hong Ji, James Potash, Sarven Sabunciyan, and Andrew P Feinberg. 2009. “The Human Colon Cancer Methylome Shows Similar Hypo- and Hypermethylation at Conserved Tissue-Specific CpG Island Shores.” *Nature Genetics* 41 (2): 178–86. https://doi.org/10.1038/ng.298.

Issa, Jean Pierre J. 2013. “The Myelodysplastic Syndrome as a Prototypical Epigenetic Disease.” *Blood*. https://doi.org/10.1182/blood-2013-02-451757.

Issa, Jean Pierre J, and Hagop M. Kantarjian. 2009. “Targeting DNA Methylation.” *Clinical Cancer Research*. https://doi.org/10.1158/1078-0432.CCR-08-2783.

Issa, Jean-Pierre J., Vazganush Gharibyan, Jorge Cortes, Jaroslav Jelinek, Gail Morris, Srdan Verstovsek, Moshe Talpaz, Guillermo Garcia-Manero, and Hagop M. Kantarjian. 2005. “Phase II Study of Low-Dose Decitabine in Patients With Chronic Myelogenous Leukemia Resistant to Imatinib Mesylate.” *Journal of Clinical Oncology* 23 (17): 3948–56. https://doi.org/10.1200/JCO.2005.11.981.

Ito, Shinsuke, Ana C. D'Alessio, Olena V. Taranova, Kwonho Hong, Lawrence C. Sowers, and Yi Zhang. 2010a. "Role of Tet Proteins in 5mC to 5hmC Conversion, ES-Cell Self-Renewal and Inner Cell Mass Specification." *Nature* 466 (7310): 1129–1133. https://doi.org/10.1038/nature09303.

Ito, Shinsuke, Ana C. D'Alessio, Olena V. Taranova, Kwonho Hong, Lawrence C. Sowers, and Yi Zhang. 2010b. "Role of Tet Proteins in 5mC to 5hmC Conversion, ES-Cell Self-Renewal and Inner Cell Mass Specification." *Nature* 466 (7310): 1129–33. https://doi.org/10.1038/nature09303.

Ito, Shinsuke, Li Shen, Qing Dai, Susan C. Wu, Leonard B. Collins, James A. Swenberg, Chuan He, and Yi Zhang. 2011. "Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine." *Science (New York, N.Y.)* 333 (6047): 1300–1303. https://doi.org/10.1126/science.1210597.

Itzykson, R, O Kosmider, T Cluzeau, V Mansat-De Mas, F Dreyfus, O Beyne-Rauzy, B Quesnel, N Vey, V Gelsi-Boyer, S Raynaud, C Preudhomme, L Adès, P Fenaux, and M Fontenay. 2011. "Impact of TET2 Mutations on Response Rate to Azacitidine in Myelodysplastic Syndromes and Low Blast Count Acute Myeloid Leukemias." *Leukemia* 25 (7): 1147–52. https://doi.org/10.1038/leu.2011.71.

Iwasaki, Masayuki, Michaela Liedtke, Andrew J Gentles, and Michael L Cleary. 2015. "CD93 Marks a Non-Quiescent Human Leukemia Stem Cell Population and Is Required for Development of MLL-Rearranged Acute Myeloid Leukemia." *Cell Stem Cell* 17 (4): 412–21. https://doi.org/10.1016/j.stem.2015.08.008.

Jabbour, Elias, Nicholas J. Short, Guillermo Montalban-Bravo, Xuelin Huang, Carlos Bueso-Ramos, Wei Qiao, Hui Yang, Chong Zhao, Tapan Kadia, Gautam Borthakur, Naveen Pemmaraju, Koji Sasaki, Zeev Estrov, Jorge Cortes, Farhad Ravandi, Yesid Alvarado, Rami Komrokji, Mikkael A. Sekeres, David P. Steensma, Amy DeZern, Gail Roboz, Hagop Kantarjian, and Guillermo Garcia-Manero. 2017. "Randomized Phase 2 Study of Low-Dose Decitabine vs Low-Dose Azacitidine in Lower-Risk MDS and MDS/MPN." *Blood* 130 (13): 1514–22. https://doi.org/10.1182/blood-2017-06-788497.

Jaiswal, Siddhartha, Pierre Fontanillas, Jason Flannick, Alisa Manning, Peter V. Grauman, Brenton G. Mar, R. Coleman Lindsley, Craig H. Mermel, Noel Burtt, Alejandro Chavez, John M. Higgins, Vladislav Moltchanov, Frank C. Kuo, Michael J. Kluk, Brian Henderson, Leena Kinnunen, Heikki A. Koistinen, Claes Ladenvall, Gad Getz, Adolfo Correa, Benjamin F. Banahan, Stacey Gabriel, Sekar Kathiresan, Heather M. Stringham, Mark I.

McCarthy, Michael Boehnke, Jaakko Tuomilehto, Christopher Haiman, Leif Groop, Gil Atzmon, James G. Wilson, Donna Neuberg, David Altshuler, and Benjamin L. Ebert. 2014. “Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes.” *New England Journal of Medicine* 371 (26): 2488–98. https://doi.org/10.1056/NEJMoa1408617.

Jiang, Yanan, Lin Liu, Jinhuan Wang, Zeng Cao, and Zhigang Zhao. 2018. “The Wilms’ Tumor Gene-1 Is a Prognostic Factor in Myelodysplastic Syndrome: A Meta Analysis.” *Oncotarget* 9 (22): 16205–12. https://doi.org/10.18632/oncotarget.23671.

Jiang, Ying, Andrew Dunbar, Lukasz P. Gondek, Sanjay Mohan, Manjot Rataul, Christine O’Keefe, Mikkael Sekeres, Yogen Saunthararajah, and Jaroslaw P. Maciejewski. 2009. “Aberrant DNA Methylation Is a Dominant Mechanism in MDS Progression to AML.” *Blood* 113 (6): 1315–25. https://doi.org/10.1182/blood-2008-06-163246.

Jones, P. A., and S. M. Taylor. 1980. “Cellular Differentiation, Cytidine Analogs and DNA Methylation.” *Cell* 20 (1): 85–93. https://doi.org/10.1016/0092-8674(80)90237-8.

Juergens, R. A., J. Wrangle, F. P. Vendetti, S. C. Murphy, M. Zhao, B. Coleman, R. Sebree, K. Rodgers, C. M. Hooker, N. Franco, B. Lee, S. Tsai, I. E. Delgado, M. A. Rudek, S. A. Belinsky, J. G. Herman, S. B. Baylin, M. V. Brock, and C. M. Rudin. 2011. “Combination Epigenetic Therapy Has Efficacy in Patients with Refractory Advanced Non-Small Cell Lung Cancer.” *Cancer Discovery* 1 (7): 598–607. https://doi.org/10.1158/2159-8290.CD-11-0214.

Kafer, Georgia Rose, Xuan Li, Takuro Horii, Isao Suetake, Shoji Tajima, Izuho Hatada, and Peter Mark Carlton. 2016. “5-Hydroxymethylcytosine Marks Sites of DNA Damage and Promotes Genome Stability.” *Cell Reports* 14 (6): 1283–92. https://doi.org/10.1016/j.celrep.2016.01.035.

Kantarjian, Hagop M., Xavier G. Thomas, Anna Dmoszynska, Agnieszka Wierzbowska, Grzegorz Mazur, Jiri Mayer, Jyh-Pyng Gau, Wen-Chien Chou, Rena Buckstein, Jaroslav Cermak, Ching-Yuan Kuo, Albert Oriol, Farhad Ravandi, Stefan Faderl, Jacques Delaunay, Daniel Lysák, Mark Minden, and Christopher Arthur. 2012. “Multicenter, Randomized, Open-Label, Phase III Trial of Decitabine Versus Patient Choice, With Physician Advice, of Either Supportive Care or Low-Dose Cytarabine for the Treatment of Older Patients With Newly Diagnosed Acute Myeloid Leukemia.” *Journal of Clinical Oncology* 30 (21): 2670–77. https://doi.org/10.1200/JCO.2011.38.9429.

Kao, Yuan-Chung, Shinn-Jong Jiang, Wen-An Pan, Kuan-Chieh Wang, Po-Ku Chen, Hsi-Ju Wei, Wei-Sheng Chen, Bi-Ing Chang, Guey-Yueh Shi, and Hua-Lin Wu. 2012. “The Epidermal Growth Factor-like Domain of CD93 Is a Potent Angiogenic Factor.” *PLOS ONE* 7 (12): e51647. https://doi.org/10.1371/journal.pone.0051647.

Kim, Jong Wook, Omar O. Abudayyeh, Huwate Yeerna, Chen-Hsiang Yeang, Michelle Stewart, Russell W. Jenkins, Shunsuke Kitajima, David J. Konieczkowski, Kate Medetgul-Ernar, Taylor Cavazos, Clarence Mah, Stephanie Ting, Eliezer M. Van Allen, Ofir Cohen, John Mcdermott, Emily Damato, Andrew J. Aguirre, Jonathan Liang, Arthur Liberzon, Gabriella Alexe, John Doench, Mahmoud Ghandi, Francisca Vazquez, Barbara A. Weir, Aviad Tsherniak, Aravind Subramanian, Karina Meneses-Cime, Jason Park, Paul Clemons, Levi A. Garraway, David Thomas, Jesse S. Boehm, David A. Barbie, William C. Hahn, Jill P. Mesirov, and Pablo Tamayo. 2017. “Decomposing Oncogenic Transcriptional Signatures to Generate Maps of Divergent Cellular States.” *Cell Systems* 5 (2): 105-118.e9. https://doi.org/10.1016/j.cels.2017.08.002.

Kim, Jong Wook, Olga B. Botvinnik, Omar Abudayyeh, Chet Birger, Joseph Rosenbluh, Yashaswi Shrestha, Mohamed E. Abazeed, Peter S. Hammerman, Daniel Dicara, David J. Konieczkowski, Cory M. Johannessen, Arthur Liberzon, Amir Reza Alizad-Rahvar, Gabriela Alexe, Andrew Aguirre, Mahmoud Ghandi, Heidi Greulich, Francisca Vazquez, Barbara A. Weir, Eliezer M. Van Allen, Aviad Tsherniak, Diane D. Shao, Travis I. Zack, Michael Noble, Gad Getz, Rameen Beroukhim, Levi A. Garraway, Masoud Ardakani, Chiara Romualdi, Gabriele Sales, David A. Barbie, Jesse S. Boehm, William C. Hahn, Jill P. Mesirov, and Pablo Tamayo. 2016. “Characterizing Genomic Alterations in Cancer by Complementary Functional Associations.” *Nature Biotechnology* 34 (5): 539–46. https://doi.org/10.1038/nbt.3527.

Kitamura, Toshio, Tsuyoshi Tange, Takashi Terasawa, Shigeru Chiba, Tomoaki Kuwaki, Kiyoshi Miyagawa, Yun-Feng Piao, Kohei Miyazono, Akio Urabe, and Fumimaro Takaku. 1989. “Establishment and Characterization of a Unique Human Cell Line That Proliferates Dependently on GM-CSF, IL-3, or Erythropoietin.” *Journal of Cellular Physiology* 140 (2): 323–34. https://doi.org/10.1002/jcp.1041400219.

Klutstein, Michael, Deborah Nejman, Razi Greenfield, and Howard Cedar. 2016. “DNA Methylation in Cancer and Aging.” *Cancer Research* 76 (12): 3446–50. https://doi.org/10.1158/0008-5472.CAN-15-3278.

Ko, Myunggon, Hozefa S. Bandukwala, Jungeun An, Edward D. Lamperti, Elizabeth C. Thompson, Ryan Hastie, Angeliki Tsangaratou, Klaus

Rajewsky, Sergei B. Koralov, and Anjana Rao. 2011. “Ten-Eleven-Translocation 2 (TET2) Negatively Regulates Homeostasis and Differentiation of Hematopoietic Stem Cells in Mice.” *Proceedings of the National Academy of Sciences* 108 (35): 14566–71. https://doi.org/10.1073/pnas.1112317108.

Ko, Myunggon, Yun Huang, Anna M Jankowska, Utz J Pape, Mamta Tahiliani, Hozefa S Bandukwala, Jungeun An, Edward D Lamperti, Kian Peng Koh, Rebecca Ganetzky, X Shirley Liu, L Aravind, Suneet Agarwal, Jaroslaw P Maciejewski, and Anjana Rao. 2010. “Impaired Hydroxylation of 5-Methylcytosine in Myeloid Cancers with Mutant TET2.” *Nature* 468 (7325): 839–43. https://doi.org/10.1038/nature09586.

Kobayashi, Sumiko, Yasunori Ueda, Yasuhito Nannya, Hirohiko Shibayama, Hideto Tamura, Kiyoyuki Ogata, Yoshiki Akatsuka, Kensuke Usuki, Yoshikazu Ito, Masaya Okada, Takahiro Suzuki, Tomoko Hata, Akira Matsuda, Kaoru Tohyama, Keiji Kakumoto, Daisuke Koga, Kinuko Mitani, Tomoki Naoe, Haruo Sugiyama, and Fumimaro Takaku. 2016. “Prognostic Significance of Wilms Tumor 1 MRNA Expression Levels in Peripheral Blood and Bone Marrow in Patients with Myelodysplastic Syndromes.” *Cancer Biomarkers : Section A of Disease Markers* 17 (1): 21–32. https://doi.org/10.3233/CBM-160612.

Kondo, Motonari, Irving L. Weissman, and Koichi Akashi. 1997. “Identification of Clonogenic Common Lymphoid Progenitors in Mouse Bone Marrow.” *Cell* 91 (5): 661–72. https://doi.org/10.1016/S0092-8674(00)80453-5.

Kunimoto, Hiroyoshi, Cem Meydan, Abbas Nazir, Justin Whitfield, Kaitlyn Shank, Franck Rapaport, Rebecca Maher, Elodie Pronier, Sara C. Meyer, Francine E. Garrett-Bakelman, Martin Tallman, Ari Melnick, Ross L. Levine, and Alan H. Shih. 2018. “Cooperative Epigenetic Remodeling by TET2 Loss and NRAS Mutation Drives Myeloid Transformation and MEK Inhibitor Sensitivity.” *Cancer Cell* 33 (1): 44-59.e8. https://doi.org/10.1016/j.ccell.2017.11.012.

Kwok, Brian, Jeff M Hall, John S Witte, Yin Xu, Prashanti Reddy, Keming Lin, Rachel Flamholz, Bashar Dabbas, Aine Yung, Jenan Al-Hafidh, Emily Balmert, Christine Vaupel, Carlos El Hader, Matthew J McGinniss, Shareef A Nahas, Julie Kines, Rafael Bejar, JL. Liesveld, MA. Lichtman, JW. Vardiman, J. Thiele, DA. Arber, P. Font, J. Loscertales, C. Benavente, P. Font, J. Loscertales, C. Soto, DP. Steensma, P. Valent, BJ. Bain, JM. Bennett, F. Wimazal, C. Fonatsch, R. Thalhammer, T. Schroeder, L. Ruf, A. Bernhardt, L. Busque, JP. Patel, ME. Figueroa, R. Bejar, Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium, T. Haferlach, Y. Nagata, V. Grossmann, K. Yoshida, M. Sanada, Y. Shiraishi,

Cancer Genome Atlas Research Network, M. Meggendorfer, A. Roller, T. Haferlach, JP. Patel, M. Gönen, ME. Figueroa, F. Damm, E. Mylonas, A. Cosson, European Research Initiative on CLL (ERIC), M. Neumann, S. Vosberg, C. Schlee, L. Couronné, C. Bastard, OA. Bernard, S. Jaiswal, P. Fontanillas, J. Flannick, G. Genovese, AK. Kähler, RE. Handsaker, M. Xie, C. Lu, J. Wang, CC. Laurie, CA. Laurie, K. Rice, KB. Jacobs, M. Yeager, W. Zhou, DP. Steensma, R. Bejar, S. Jaiswal, A. Nazha, MA. Sekeres, SD. Gore, AM. Zeidan, J. Hall, J. Al Hafidh, E. Balmert, B. Kwok, P. Reddy, K. Lin, AK. Wong, B. Fang, L. Zhang, X. Guo, S. Lee, R. Schreck, Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium and of the Associazione Italiana per la Ricerca sul Cancro Gruppo Italiano Malattie Mieloproliferative, Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium, MA. Sekeres, M. Mallo, E. Luño, C. Sanzo, LJ. Zhang, ES. Shin, ZX. Yu, SB. Li, R. Bejar, KE. Stevenson, and BA. Caughey. 2015. “MDS-Associated Somatic Mutations and Clonal Hematopoiesis Are Common in Idiopathic Cytopenias of Undetermined Significance.” *Blood* 126 (21): 2355–61. https://doi.org/10.1182/blood-2015-08-667063.

Labun, Kornel, Tessa G. Montague, James A. Gagnon, Summer B. Thyme, and Eivind Valen. 2016. “CHOPCHOP v2: A Web Tool for the next Generation of CRISPR Genome Engineering.” *Nucleic Acids Research* 44 (W1): W272–76. https://doi.org/10.1093/nar/gkw398.

Langenkamp, Elise, Lei Zhang, Roberta Lugano, Hua Huang, Tamador Elsir Abu Elhassan, Maria Georganaki, Wesam Bazzar, Johan Lööf, George Trendelenburg, Magnus Essand, Fredrik Pontén, Anja Smits, and Anna Dimberg. 2015. “Elevated Expression of the C-Type Lectin CD93 in the Glioblastoma Vasculature Regulates Cytoskeletal Rearrangements That Enhance Vessel Function and Reduce Host Survival.” *Cancer Research* 75 (21): 4504–16. https://doi.org/10.1158/0008-5472.CAN-14-3636.

Laurenti, Elisa, and Berthold Göttgens. 2018. “From Haematopoietic Stem Cells to Complex Differentiation Landscapes.” *Nature* 553 (7689): 418–26. https://doi.org/10.1038/nature25022.

Lee, Minjung, Jia Li, Yi Liang, Guolin Ma, Jixiang Zhang, Lian He, Yuliang Liu, Qian Li, Minyong Li, Deqiang Sun, Yubin Zhou, and Yun Huang. 2017. “Engineered Split-TET2 Enzyme for Inducible Epigenetic Remodeling.” *Journal of the American Chemical Society* 139 (13): 4659–62. https://doi.org/10.1021/jacs.7b01459.

Lee, Stanley Chun-Wei, Heidi Dvinge, Eunhee Kim, Hana Cho, Jean-Baptiste Micol, Young Rock Chung, Benjamin H Durham, Akihide Yoshimi, Young Joon Kim, Michael Thomas, Camille Lobry, Chun-Wei Chen, Alessandro

Pastore, Justin Taylor, Xujun Wang, Andrei Krivtsov, Scott A Armstrong, James Palacino, Silvia Buonamici, Peter G Smith, Robert K Bradley, and Omar Abdel-Wahab. 2016. “Modulation of Splicing Catalysis for Therapeutic Targeting of Leukemia with Mutations in Genes Encoding Spliceosomal Proteins.” *Nature Medicine* 22 (6): 672–78. https://doi.org/10.1038/nm.4097.

Lee, Stanley Chun-Wei, Khrystyna North, Eunhee Kim, Eunjung Jang, Esther Obeng, Sydney X Lu, Bo Liu, Daichi Inoue, Akihide Yoshimi, Michelle Ki, Mirae Yeo, Xiao Jing Zhang, Min Kyung Kim, Hana Cho, Young Rock Chung, Justin Taylor, Benjamin H Durham, Young Joon Kim, Alessandro Pastore, Sebastien Monette, James Palacino, Michael Seiler, Silvia Buonamici, Peter G Smith, Benjamin L Ebert, Robert K Bradley, and Omar Abdel-Wahab. 2018. “Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations.” *Cancer Cell* 34 (2): 225-241.e8. https://doi.org/10.1016/j.ccell.2018.07.003.

Li, E., T. H. Bestor, and R. Jaenisch. 1992. “Targeted Mutation of the DNA Methyltransferase Gene Results in Embryonic Lethality.” *Cell* 69 (6): 915–26. https://doi.org/10.1016/0092-8674(92)90611-f.

Lister, Ryan, Mattia Pelizzola, Robert H. Dowen, R. David Hawkins, Gary Hon, Julian Tonti-Filippini, Joseph R. Nery, Leonard Lee, Zhen Ye, Que-Minh Ngo, Lee Edsall, Jessica Antosiewicz-Bourget, Ron Stewart, Victor Ruotti, A. Harvey Millar, James A. Thomson, Bing Ren, and Joseph R. Ecker. 2009. “Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences.” *Nature* 462 (7271): 315–22. https://doi.org/10.1038/nature08514.

Losman, Julie-Aurore, Ryan E Looper, Peppi Koivunen, Sungwoo Lee, Rebekka K Schneider, Christine McMahon, Glenn S Cowley, David E Root, Benjamin L Ebert, William G Kaelin, and Jr. 2013. “(R)-2-Hydroxyglutarate Is Sufficient to Promote Leukemogenesis and Its Effects Are Reversible.” *Science (New York, N.Y.)* 339 (6127): 1621–25. https://doi.org/10.1126/science.1231677.

Lübbert, Michael, Stefan Suciu, Liliana Baila, Björn Hans Rüter, Uwe Platzbecker, Aristoteles Giagounidis, Dominik Selleslag, Boris Labar, Ulrich Germing, Helmut R Salih, Filip Beeldens, Petra Muus, Karl-Heinz Pflüger, Corneel Coens, Anne Hagemeijer, Hans Eckart Schaefer, Arnold Ganser, Carlo Aul, Theo de Witte, and Pierre W Wijermans. 2011. “Low-Dose Decitabine versus Best Supportive Care in Elderly Patients with Intermediate- or High-Risk Myelodysplastic Syndrome (MDS) Ineligible for Intensive Chemotherapy: Final Results of the Randomized Phase III Study of the European Organisation for Research and Treatment of Cancer Leukemia Group and the German MDS Study Group.” *Journal of Clinical Oncology :*

Official Journal of the American Society of Clinical Oncology 29 (15): 1987–96. https://doi.org/10.1200/JCO.2010.30.9245.

Lugano, Roberta, Kalyani Vemuri, Di Yu, Michael Bergqvist, Anja Smits, Magnus Essand, Staffan Johansson, Elisabetta Dejana, and Anna Dimberg. 2018. “CD93 Promotes B1 Integrin Activation and Fibronectin Fibrillogenesis during Tumor Angiogenesis.” *The Journal of Clinical Investigation* 128 (8): 3280–97. https://doi.org/10.1172/JCI97459.

Madzo, Jozef, Hui Liu, Alexis Rodriguez, Aparna Vasanthakumar, Sriram Sundaravel, Donne Bennett D. Caces, Timothy J. Looney, Li Zhang, Janet B. Lepore, Trisha Macrae, Robert Duszynski, Alan H. Shih, Chun-Xiao Song, Miao Yu, Yiting Yu, Robert Grossman, Brigitte Raumann, Amit Verma, Chuan He, Ross L. Levine, Don Lavelle, Bruce T. Lahn, Amittha Wickrema, and Lucy A. Godley. 2014. “Hydroxymethylation at Gene Regulatory Regions Directs Stem/Early Progenitor Cell Commitment during Erythropoiesis.” *Cell Reports* 6 (1): 231–44. https://doi.org/10.1016/j.celrep.2013.11.044.

Maiti, Atanu, and Alexander C. Drohat. 2011a. “Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine.” *Journal of Biological Chemistry* 286 (41): 35334–35338. https://doi.org/10.1074/jbc.C111.284620.

———. 2011b. “Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine.” *Journal of Biological Chemistry* 286 (41): 35334–38. https://doi.org/10.1074/jbc.C111.284620.

Malcovati, Luca, Anna Gallì, Erica Travaglino, Ilaria Ambaglio, Ettore Rizzo, Elisabetta Molteni, Chiara Elena, Virginia Valeria Ferretti, Silvia Catricalà, Elisa Bono, Gabriele Todisco, Antonio Bianchessi, Elisa Rumi, Silvia Zibellini, Daniela Pietra, Emanuela Boveri, Clara Camaschella, Daniela Toniolo, Elli Papaemmanuil, Seishi Ogawa, and Mario Cazzola. 2017. “Clinical Significance of Somatic Mutation in Unexplained Blood Cytopenia.” *Blood* 129 (25): 3371–78. https://doi.org/10.1182/blood-2017-01-763425.

Mayer, Wolfgang, Alain Niveleau, Jörn Walter, Reinald Fundele, and Thomas Haaf. 2000. “Demethylation of the Zygotic Paternal Genome.” *Nature* 403 (6769): 501. https://doi.org/10.1038/35000656.

Meggendorfer, Manja, Andreas Roller, Torsten Haferlach, Christiane Eder, Frank Dicker, Vera Grossmann, Alexander Kohlmann, Tamara Alpermann, Kenichi Yoshida, Seishi Ogawa, H Phillip Koeffler, Wolfgang Kern, Claudia Haferlach, and Susanne Schnittger. 2012. “SRSF2 Mutations in 275 Cases

with Chronic Myelomonocytic Leukemia (CMML).” *Blood* 120 (15): 3080–88. https://doi.org/10.1182/blood-2012-01-404863.

Meldi, Kristen, Tingting Qin, Francesca Buchi, Nathalie Droin, Jason Sotzen, Jean-Baptiste Micol, Dorothée Selimoglu-Buet, Erico Masala, Bernardino Allione, Daniela Gioia, Antonella Poloni, Monia Lunghi, Eric Solary, Omar Abdel-Wahab, Valeria Santini, Maria E. Figueroa, JM Bennett, JW Vardiman, NL Harris, RD Brunning, MJ Walter, A Tefferi, O Abdel-Wahab, V Gelsi-Boyer, SN Khan, T Ernst, H Makishima, G Nikoloski, ME Figueroa, Y Jiang, H Kantarjian, LR Silverman, K Ghoshal, K Patel, J Dickson, S Din, K Macleod, D Jodrell, B Ramsahoye, SS Palii, BO Van Emburgh, UT Sankpal, KD Brown, KD Robertson, A Cihak, JW Weiss, HC Pitot, EA Griffiths, SD Gore, R Itzykson, SD Gore, M Daskalakis, H Kantarjian, C Mund, B Hackanson, C Stresemann, M Lubbert, F Lyko, W Blum, L Shen, MY Follo, JP Issa, TE Fandy, R Bejar, AM Jankowska, O Kosmider, K Yoshida, MM Patnaik, R Bejar, F Traina, R Itzykson, T Braun, BD Cheson, R Itzykson, MM Patnaik, M Meggendorfer, A Kohlmann, A Akalin, ME Figueroa, ME Figueroa, L Bullinger, M Ehrich, A Sanyal, BR Lajoie, G Jain, J Dekker, Y Park, ME Figueroa, LS Rozek, MA Sartor, D Aran, S Sabato, A Hellman, C Cortes, V Vapnik, A Valencia, A Subramanian, SM Graham, JK Vass, TL Holyoake, GJ Graham, A Schaffner, P Rhyn, G Schoedon, DJ Schaer, MM Pillai, M Iwata, N Awaya, L Graf, B Torok-Storb, FO Bagger, FO Bagger, M Aivado, C Chen, DT Bowen, AA Giagounidis, B Schlegelberger, S Haase, EG Wright, H Reikvam, H Fredly, AO Kittang, O Bruserud, AO Kittang, K Hatfield, K Sand, H Reikvam, Ø Bruserud, ZC Han, SK Gupta, JP Singh, AZ Dudek, I Nesmelova, K Mayo, CM Verfaillie, S Pitchford, A Slungaard, I Bruns, HC Tsai, T Prebet, J Dekker, K Rippe, M Dekker, N Kleckner, M Simonis, Z Zhao, J Dostie, E Lieberman-Aiden, YF Wong, LM Jakt, S Nishikawa, E Chung, CL Hsu, M Kondo, E Katsoulidis, SL Abrams, LS Steelman, L Zhang, AM Bolger, M Lohse, B Usadel, H Li, R Durbin, MA DePristo, K Wang, M Li, H Hakonarson, F Krueger, SR Andrews, RC Gentleman, T Sing, O Sander, N Beerenwinkel, T Lengauer, L Jiang, C Trapnell, DG Hendrickson, M Sauvageau, L Goff, JL Rinn, L Pachter, S Anders, PT Pyl, W Huber, MD Robinson, DJ McCarthy, and GK Smyth. 2015. “Specific Molecular Signatures Predict Decitabine Response in Chronic Myelomonocytic Leukemia.” *Journal of Clinical Investigation* 125 (5): 1857–72. https://doi.org/10.1172/JCI78752.

Merlevede, Jane, Nathalie Droin, Tingting Qin, Kristen Meldi, Kenichi Yoshida, Margot Morabito, Emilie Chautard, Didier Auboeuf, Pierre Fenaux, Thorsten Braun, Raphael Itzykson, Stéphane de Botton, Bruno Quesnel, Thérèse Commes, Eric Jourdan, William Vainchenker, Olivier Bernard, Noemie Pata-Merci, Stéphanie Solier, Velimir Gayevskiy, Marcel E. Dinger, Mark J. Cowley, Dorothée Selimoglu-Buet, Vincent Meyer, François Artiguenave,

Jean-François Deleuze, Claude Preudhomme, Michael R. Stratton, Ludmil B. Alexandrov, Eric Padron, Seishi Ogawa, Serge Koscielny, Maria Figueroa, and Eric Solary. 2016. “Mutation Allele Burden Remains Unchanged in Chronic Myelomonocytic Leukaemia Responding to Hypomethylating Agents.” *Nature Communications* 7 (February): 10767. https://doi.org/10.1038/ncomms10767.

Moran-Crusio, Kelly, Linsey Reavie, Alan Shih, Omar Abdel-Wahab, Delphine Ndiaye-Lobry, Camille Lobry, Maria E. Figueroa, Aparna Vasanthakumar, Jay Patel, Xinyang Zhao, Fabiana Perna, Suveg Pandey, Jozef Madzo, Chunxiao Song, Qing Dai, Chuan He, Sherif Ibrahim, Miloslav Beran, Jiri Zavadil, Stephen D. Nimer, Ari Melnick, Lucy A. Godley, Iannis Aifantis, and Ross L. Levine. 2011. “Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation.” *Cancer Cell* 20 (1): 11–24. https://doi.org/10.1016/j.ccr.2011.06.001.

Morimoto, Suguru, Hideki Makishima, Yasunobu Nagata, Niroshan Nadarajah, Constance Baer, Tamara Alpermann, Genta Nagae, Yasuhito Nannya, Yasushi Miyazaki, Kenichi Yoshida, Tetsuichi Yoshizato, Masahiro Marshall Nakagawa, Ryosaku Inagaki, June Takeda, Yoichi Fujii, Yasuhide Takeuchi, Hiroo Ueno, Yuichi Shiraishi, Kenichi Chiba, Hiroko Tanaka, Masashi Sanada, Satoru Miyano, Claudia Haferlach, Wolfgang Kern, Hiroyuki Aburatani, Torsten Haferlach, and Seishi Ogawa. 2018. “DNA Methylation and Genetic Profiles in 320 Patients with Myelodysplastic Syndromes.” *Blood* 132 (Suppl 1): 1799–1799. https://doi.org/10.1182/blood-2018-99-116489.

Mund, Cora, Bodo Brueckner, and Frank Lyko. 2006. “Reactivation of Epigenetically Silenced Genes by DNA Methyltransferase Inhibitors: Basic Concepts and Clinical Applications.” *Epigenetics* 1 (1): 7–13. https://doi.org/10.4161/epi.1.1.2375.

Muto, Tomoya, Goro Sashida, Motohiko Oshima, George R. Wendt, Makiko Mochizuki-Kashio, Yasunobu Nagata, Masashi Sanada, Satoru Miyagi, Atsunori Saraya, Asuka Kamio, Genta Nagae, Chiaki Nakaseko, Koutaro Yokote, Kazuya Shimoda, Haruhiko Koseki, Yutaka Suzuki, Sumio Sugano, Hiroyuki Aburatani, Seishi Ogawa, and Atsushi Iwama. 2013. “Concurrent Loss of Ezh2 and Tet2 Cooperates in the Pathogenesis of Myelodysplastic Disorders.” *Journal of Experimental Medicine* 210 (12): 2627–39. https://doi.org/10.1084/jem.20131144.

Neri, Francesco, Stefania Rapelli, Anna Krepelova, Danny Incarnato, Caterina Parlato, Giulia Basile, Mara Maldotti, Francesca Anselmi, and Salvatore Oliviero. 2017. “Intragenic DNA Methylation Prevents Spurious

Transcription Initiation.” *Nature* 543 (7643): 72–77. https://doi.org/10.1038/nature21373.

Network, The Cancer Genome Atlas Research. 2013. “Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia.” *The New England Journal of Medicine* 22 (368): 2059–74. https://doi.org/doi:10.1056/NEJMoa1301689.

Niavarani, Ahmadreza, Tobias Herold, Yasmin Reyal, Maria C Sauerland, Thomas Buchner, Wolfgang Hiddemann, Stefan K Bohlander, Peter J M Valk, and Dominique Bonnet. 2016. “A 4-Gene Expression Score Associated with High Levels of Wilms Tumor-1 (WT1) Expression Is an Adverse Prognostic Factor in Acute Myeloid Leukaemia.” *British Journal of Haematology* 172 (3): 401–11. https://doi.org/10.1111/bjh.13836.

Nikoloski, Gorica, Saskia M C Langemeijer, Roland P Kuiper, Ruth Knops, Marion Massop, Evelyn R L T M Tönnissen, Adrian van der Heijden, Theresia N Scheele, Peter Vandenberghe, Theo de Witte, Bert A van der Reijden, and Joop H Jansen. 2010. “Somatic Mutations of the Histone Methyltransferase Gene EZH2 in Myelodysplastic Syndromes.” *Nature Genetics* 42 (8): 665–67. https://doi.org/10.1038/ng.620.

Ogawa, Seishi. 2019. “Genetics of MDS.” *Blood*, January, blood-2018-10-844621. https://doi.org/10.1182/blood-2018-10-844621.

Okano, Masaki, Daphne W Bell, Daniel A Haber, and En Li. 1999. “DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development.” *Cell* 99 (3): 247–57. https://doi.org/10.1016/S0092-8674(00)81656-6.

Olsen, Renate S, Mikael Lindh, Emina Vorkapic, Roland E Andersson, Niklas Zar, Sture Löfgren, Jan Dimberg, Andreas Matussek, and Dick Wågsäter. 2015. “CD93 Gene Polymorphism Is Associated with Disseminated Colorectal Cancer.” *International Journal of Colorectal Disease* 30 (7): 883–90. https://doi.org/10.1007/s00384-015-2247-1.

Orlando, David A., Mei Wei Chen, Victoria E. Brown, Snehakumari Solanki, Yoon J. Choi, Eric R. Olson, Christian C. Fritz, James E. Bradner, and Matthew G. Guenther. 2014. “Quantitative ChIP-Seq Normalization Reveals Global Modulation of the Epigenome.” *Cell Reports*. https://doi.org/10.1016/j.celrep.2014.10.018.

Orlanski, Shari, Verena Labi, Yitzhak Reizel, Adam Spiro, Michal Lichtenstein, Rena Levin-Klein, Sergei B. Koralov, Yael Skversky, Klaus Rajewsky, Howard Cedar, and Yehudit Bergman. 2016. “Tissue-Specific DNA

Demethylation Is Required for Proper B-Cell Differentiation and Function." *Proceedings of the National Academy of Sciences of the United States of America* 113 (18): 5018–23. https://doi.org/10.1073/pnas.1604365113.

Papaemmanuil, Elli, Moritz Gerstung, Luca Malcovati, Sudhir Tauro, Gunes Gundem, Peter Van Loo, Chris J Yoon, Peter Ellis, David C Wedge, Andrea Pellagatti, Adam Shlien, Michael John Groves, Simon A Forbes, Keiran Raine, Jon Hinton, Laura J Mudie, Stuart McLaren, Claire Hardy, Calli Latimer, Matteo G Della Porta, Sarah O'Meara, Ilaria Ambaglio, Anna Galli, Adam P Butler, Gunilla Walldin, Jon W Teague, Lynn Quek, Alex Sternberg, Carlo Gambacorti-Passerini, Nicholas C P Cross, Anthony R Green, Jacqueline Boultwood, Paresh Vyas, Eva Hellstrom-Lindberg, David Bowen, Mario Cazzola, Michael R Stratton, and Peter J Campbell. 2013. "Clinical and Biological Implications of Driver Mutations in Myelodysplastic Syndromes." *Blood* 122 (22): 3616–27; quiz 3699. https://doi.org/10.1182/blood-2013-08-518886.

Pappenheim, Artur. 1896. "Ueber Entwickelung und Ausbildung der Erythroblasten." *Archiv für pathologische Anatomie und Physiologie und für klinische Medicin* 145 (3): 587–643. https://doi.org/10.1007/BF01969901.

Pérez, Cristina, Nicolas Martínez-Calle, José Ignacio Martín-Subero, Victor Segura, Eric Delabesse, Marta Fernandez-Mercado, Leire Garate, Sara Alvarez, José Rifon, Sara Varea, Jacqueline Boultwood, James S Wainscoat, Juan Cruz Cigudosa, María José Calasanz, Nicholas C P Cross, Felipe Prósper, and Xabier Agirre. 2012. "TET2 Mutations Are Associated with Specific 5-Methylcytosine and 5-Hydroxymethylcytosine Profiles in Patients with Chronic Myelomonocytic Leukemia." *PloS One* 7 (2): e31605. https://doi.org/10.1371/journal.pone.0031605.

Platzbecker, Uwe. 2019. "Treatment of MDS." *Blood* 133 (10): 1096–1107. https://doi.org/10.1182/blood-2018-10-844696.

Prendergast, G. C., and E. B. Ziff. 1991. "Methylation-Sensitive Sequence-Specific DNA Binding by the c-Myc Basic Region." *Science* 251 (4990): 186–89. https://doi.org/10.1126/science.1987636.

Qin, Taichun, Ryan Castoro, Samih El Ahdab, Jaroslav Jelinek, Xiaodan Wang, Jiali Si, Jingmin Shu, Rong He, Nianxiang Zhang, Woonbok Chung, Hagop M Kantarjian, and Jean-Pierre J Issa. 2011. "Mechanisms of Resistance to Decitabine in the Myelodysplastic Syndrome." Edited by S. K. Batra. *PloS One* 6 (8): e23372. https://doi.org/10.1371/journal.pone.0023372.

Ramsahoye, Bernard H., Detlev Biniszkiewicz, Frank Lyko, Victoria Clark, Adrian P. Bird, and Rudolf Jaenisch. 2000. "Non-CpG Methylation Is Prevalent in

Embryonic Stem Cells and May Be Mediated by DNA Methyltransferase 3a." *Proceedings of the National Academy of Sciences* 97 (10): 5237–42. https://doi.org/10.1073/pnas.97.10.5237.

Ran, F Ann, Patrick D Hsu, Jason Wright, Vineeta Agarwala, David A Scott, and Feng Zhang. 2013. "Genome Engineering Using the CRISPR-Cas9 System." *Nature Protocols* 8 (11): 2281–2308. https://doi.org/10.1038/nprot.2013.143.

Rasmussen, Kasper D, Guangshuai Jia, Jens V Johansen, Marianne T Pedersen, Nicolas Rapin, Frederik O Bagger, Bo T Porse, Olivier A Bernard, Jesper Christensen, and Kristian Helin. 2015. "Loss of TET2 in Hematopoietic Cells Leads to DNA Hypermethylation of Active Enhancers and Induction of Leukemogenesis." *Genes & Development* 29 (9): 910–22. https://doi.org/10.1101/gad.260174.115.

Rasmussen, Kasper Dindler, and Kristian Helin. 2016. "Role of TET Enzymes in DNA Methylation, Development, and Cancer." *Genes & Development* 30 (7): 733–50. https://doi.org/10.1101/gad.276568.115.

Roadmap Epigenomics Consortium, Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha Bilenky, Angela Yen, Alireza Heravi-Moussavi, Pouya Kheradpour, Zhizhuo Zhang, Jianrong Wang, Michael J Ziller, Viren Amin, John W Whitaker, Matthew D Schultz, Lucas D Ward, Abhishek Sarkar, Gerald Quon, Richard S Sandstrom, Matthew L Eaton, Yi-Chieh Wu, Andreas R Pfenning, Xinchen Wang, Melina Claussnitzer, Yaping Liu, Cristian Coarfa, R Alan Harris, Noam Shoresh, Charles B Epstein, Elizabeta Gjoneska, Danny Leung, Wei Xie, R David Hawkins, Ryan Lister, Chibo Hong, Philippe Gascard, Andrew J Mungall, Richard Moore, Eric Chuah, Angela Tam, Theresa K Canfield, R Scott Hansen, Rajinder Kaul, Peter J Sabo, Mukul S Bansal, Annaick Carles, Jesse R Dixon, Kai-How Farh, Soheil Feizi, Rosa Karlic, Ah-Ram Kim, Ashwinikumar Kulkarni, Daofeng Li, Rebecca Lowdon, GiNell Elliott, Tim R Mercer, Shane J Neph, Vitor Onuchic, Paz Polak, Nisha Rajagopal, Pradipta Ray, Richard C Sallari, Kyle T Siebenthall, Nicholas A Sinnott-Armstrong, Michael Stevens, Robert E Thurman, Jie Wu, Bo Zhang, Xin Zhou, Arthur E Beaudet, Laurie A Boyer, Philip L De Jager, Peggy J Farnham, Susan J Fisher, David Haussler, Steven J M Jones, Wei Li, Marco A Marra, Michael T McManus, Shamil Sunyaev, James A Thomson, Thea D Tlsty, Li-Huei Tsai, Wei Wang, Robert A Waterland, Michael Q Zhang, Lisa H Chadwick, Bradley E Bernstein, Joseph F Costello, Joseph R Ecker, Martin Hirst, Alexander Meissner, Aleksandar Milosavljevic, Bing Ren, John A Stamatoyannopoulos, Ting Wang, Manolis Kellis, and Roadmap Epigenomics Consortium. 2015. "Integrative Analysis of 111 Reference Human Epigenomes." *Nature* 518 (7539): 317–30.

Römermann, D, B Hasemeier, K Metzig, G Göhring, B Schlegelberger, F Länger, H Kreipe, and U Lehmann. 2008. “Global Increase in DNA Methylation in Patients with Myelodysplastic Syndrome.” *Leukemia* 22 (10): 1954–56. https://doi.org/10.1038/leu.2008.76.

Rönnerblad, Michelle, Robin Andersson, Tor Olofsson, Iyadh Douagi, Mohsen Karimi, Sören Lehmann, Ilka Hoof, Michiel de Hoon, Masayoshi Itoh, Sayaka Nagao-Sato, Hideya Kawaji, Timo Lassmann, Piero Carninci, Yoshihide Hayashizaki, Alistair R R Forrest, Albin Sandelin, Karl Ekwall, Erik Arner, Andreas Lennartsson, and FANTOM consortium. 2014. “Analysis of the DNA Methylome and Transcriptome in Granulopoiesis Reveals Timed Changes and Dynamic Enhancer Methylation.” *Blood* 123 (17): 79–89. https://doi.org/10.1182/blood-2013-02-482893.

Rousseeuw, Peter J. 1987. “Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis.” *Journal of Computational and Applied Mathematics*. https://doi.org/10.1016/0377-0427(87)90125-7.

Santi, D V, A Norment, and C E Garrett. 1984. “Covalent Bond Formation between a DNA-Cytosine Methyltransferase and DNA Containing 5-Azacytosine.” *Proceedings of the National Academy of Sciences of the United States of America* 81 (22): 6993–97.

Saxonov, Serge, Paul Berg, and Douglas L. Brutlag. 2006. “A Genome-Wide Analysis of CpG Dinucleotides in the Human Genome Distinguishes Two Distinct Classes of Promoters.” *Proceedings of the National Academy of Sciences* 103 (5): 1412–17. https://doi.org/10.1073/pnas.0510310103.

Schmidl, Christian, Maja Klug, Tina J. Boeld, Reinhard Andreesen, Petra Hoffmann, Matthias Edinger, and Michael Rehli. 2009. “Lineage-Specific DNA Methylation in T Cells Correlates with Histone Methylation and Enhancer Activity.” *Genome Research* 19 (7): 1165–74. https://doi.org/10.1101/gr.091470.109.

Schultz, Matthew D., Yupeng He, John W. Whitaker, Manoj Hariharan, Eran A. Mukamel, Danny Leung, Nisha Rajagopal, Joseph R. Nery, Mark A. Urich, Huaming Chen, Shin Lin, Yiing Lin, Inkyung Jung, Anthony D. Schmitt, Siddarth Selvaraj, Bing Ren, Terrence J. Sejnowski, Wei Wang, and Joseph R. Ecker. 2015a. “Human Body Epigenome Maps Reveal Noncanonical DNA Methylation Variation.” *Nature*. https://doi.org/10.1038/nature14465.

———. 2015b. “Human Body Epigenome Maps Reveal Noncanonical DNA Methylation Variation.” *Nature*. https://doi.org/10.1038/nature14465.

Sheaffer, Karyn L, Rinho Kim, Reina Aoki, Ellen N Elliott, Jonathan Schug, Lukas Burger, Dirk Schübeler, and Klaus H Kaestner. 2014a. “DNA Methylation Is Required for the Control of Stem Cell Differentiation in the Small Intestine.” *Genes & Development* 28 (6): 652–64. https://doi.org/10.1101/gad.230318.113.

———. 2014b. “DNA Methylation Is Required for the Control of Stem Cell Differentiation in the Small Intestine.” *Genes & Development* 28 (6): 652–64. https://doi.org/10.1101/gad.230318.113.

Shen, Lanlan, Hagop Kantarjian, Yi Guo, E Lin, Jianqin Shan, Xuelin Huang, Donald Berry, Saira Ahmed, Wei Zhu, Sherry Pierce, Yutaka Kondo, Yasuhiro Oki, Jaroslav Jelinek, Hussain Saba, Eli Estey, and Jean-Pierre J Issa. 2010a. “DNA Methylation Predicts Survival and Response to Therapy in Patients with Myelodysplastic Syndromes.” *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 28 (4): 605–13. https://doi.org/10.1200/JCO.2009.23.4781.

———. 2010b. “DNA Methylation Predicts Survival and Response to Therapy in Patients with Myelodysplastic Syndromes.” *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 28 (4): 605–13. https://doi.org/10.1200/JCO.2009.23.4781.

Shiozawa, Yusuke, Luca Malcovati, Anna Gallì, Andrea Pellagatti, Mohsen Karimi, Aiko Sato-Otsubo, Yusuke Sato, Hiromichi Suzuki, Tetsuichi Yoshizato, Kenichi Yoshida, Yuichi Shiraishi, Kenichi Chiba, Hideki Makishima, Jacqueline Boultwood, Eva Hellström-Lindberg, Satoru Miyano, Mario Cazzola, and Seishi Ogawa. 2017. “Gene Expression and Risk of Leukemic Transformation in Myelodysplasia.” *Blood* 130 (24): 2642–53. https://doi.org/10.1182/blood-2017-05-783050.

Shlyueva, Daria, Gerald Stampfel, and Alexander Stark. 2014. “Transcriptional Enhancers: From Properties to Genome-Wide Predictions.” *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3682.

Sigari, Naseh, Ali Jalili, Laili Mahdawi, Ebrahim Ghaderi, and Mohammadi Shilan. 2016. “Soluble CD93 as a Novel Biomarker in Asthma Exacerbation.” *Allergy, Asthma & Immunology Research* 8 (5): 461–65. https://doi.org/10.4168/aair.2016.8.5.461.

Silverman, Lewis R., Pierre Fenaux, Ghulam J. Mufti, Valeria Santini, Eva Hellström-Lindberg, Norbert Gattermann, Guillermo Sanz, Alan F. List, Steven D. Gore, and John F. Seymour. 2011. “Continued Azacitidine Therapy beyond Time of First Response Improves Quality of Response in

Patients with Higher-Risk Myelodysplastic Syndromes.” *Cancer* 117 (12): 2697–2702. https://doi.org/10.1002/cncr.25774.

Smith, Zachary D., and Alexander Meissner. 2013. “DNA Methylation: Roles in Mammalian Development.” *Nature Reviews Genetics* 14 (3): 204–20. https://doi.org/10.1038/nrg3354.

Sorm, F., A. Píškala, A. Cihák, and J. Veselý. 1964. “5-Azacytidine, a New, Highly Effective Cancerostatic.” *Experientia* 20 (4): 202–3. https://doi.org/10.1007/bf02135399.

Spencer, David H, David A Russler-Germain, Shamika Ketkar, Nichole M Helton, Tamara L Lamprecht, Robert S Fulton, Catrina C Fronick, Michelle O’Laughlin, Sharon E Heath, Marwan Shinawi, Peter Westervelt, Jacqueline E Payton, Lukas D Wartman, John S Welch, Richard K Wilson, Matthew J Walter, Daniel C Link, John F DiPersio, and Timothy J Ley. 2017. “CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression.” *Cell* 168 (5): 801-816.e13. https://doi.org/10.1016/j.cell.2017.01.021.

Sperling, Adam S., Christopher J. Gibson, and Benjamin L. Ebert. 2017. “The Genetics of Myelodysplastic Syndrome: From Clonal Haematopoiesis to Secondary Leukaemia.” *Nature Reviews Cancer*. https://doi.org/10.1038/nrc.2016.112.

Stahl, Maximilian, Michelle DeVeaux, Pau Montesinos, Raphael Itzykson, Ellen K. Ritchie, Mikkael A. Sekeres, John D. Barnard, Nikolai A. Podoltsev, Andrew M. Brunner, Rami S. Komrokji, Vijaya R. Bhatt, Aref Al-Kali, Thomas Cluzeau, Valeria Santini, Amir T. Fathi, Gail J. Roboz, Pierre Fenaux, Mark R. Litzow, Sarah Perreault, Tae Kon Kim, Thomas Prebet, Norbert Vey, Vivek Verma, Ulrich Germing, Juan Miguel Bergua, Josefina Serrano, Steven D. Gore, and Amer M. Zeidan. 2018. “Hypomethylating Agents in Relapsed and Refractory AML: Outcomes and Their Predictors in a Large International Patient Cohort.” *Blood Advances* 2 (8): 923–32. https://doi.org/10.1182/BLOODADVANCES.2018016121.

Steensma, David P, Maria R Baer, James L Slack, Rena Buckstein, Lucy A Godley, Guillermo Garcia-Manero, Maher Albitar, Julie S Larsen, Sujata Arora, Michael T Cullen, and Hagop Kantarjian. 2009. “Multicenter Study of Decitabine Administered Daily for 5 Days Every 4 Weeks to Adults with Myelodysplastic Syndromes: The Alternative Dosing for Outpatient Treatment (ADOPT) Trial.” *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 27 (23): 3842–48. https://doi.org/10.1200/JCO.2008.19.6550.

Stresemann, Carlo, and Frank Lyko. 2008. “Modes of Action of the DNA Methyltransferase Inhibitors Azacytidine and Decitabine.” *International Journal of Cancer. Journal International Du Cancer* 123 (1): 8–13. https://doi.org/10.1002/ijc.23607.

Subramanian, Aravind, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, and Jill P Mesirov. 2005. “Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (43): 15545–50. https://doi.org/10.1073/pnas.0506580102.

Tahiliani, Mamta, Kian Peng Koh, Yinghua Shen, William A Pastor, Hozefa Bandukwala, Yevgeny Brudno, Suneet Agarwal, Lakshminarayan M Iyer, David R Liu, L Aravind, and Anjana Rao. 2009a. “Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1.” *Science (New York, N.Y.)* 324 (5929): 930–5. https://doi.org/10.1126/science.1170116.

———. 2009b. “Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1.” *Science (New York, N.Y.)* 324 (5929): 930–35. https://doi.org/10.1126/science.1170116.

Tefferi, Ayalew, and James W Vardiman. 2009. “Myelodysplastic Syndromes.” *The New England Journal of Medicine* 361 (19): 1872–85. https://doi.org/10.1056/NEJMra0902908.

Tenner, A. J. 1998. “C1q Receptors: Regulating Specific Functions of Phagocytic Cells.” *Immunobiology* 199 (2): 250–64. https://doi.org/10.1016/S0171-2985(98)80031-4.

Thol, Felicitas, Sofia Kade, Carola Schlarmann, Patrick Löffeld, Michael Morgan, Jürgen Krauter, Marcin W Wlodarski, Britta Kölking, Martin Wichmann, Kerstin Görlich, Gudrun Göhring, Gesine Bug, Oliver Ottmann, Charlotte M Niemeyer, Wolf-Karsten Hofmann, Brigitte Schlegelberger, Arnold Ganser, and Michael Heuser. 2012. “Frequency and Prognostic Impact of Mutations in SRSF2, U2AF1, and ZRSR2 in Patients with Myelodysplastic Syndromes.” *Blood* 119 (15): 3578–84. https://doi.org/10.1182/blood-2011-12-399337.

Traina, F, V Visconte, P Elson, A Tabarroki, A M Jankowska, E Hasrouni, Y Sugimoto, H Szpurka, H Makishima, C L O’Keefe, M A Sekeres, A S Advani, M Kalaycio, E A Copelan, Y Sauntharajah, S T Olalla Saad, J P Maciejewski, and R V Tiu. 2014. “Impact of Molecular Mutations on

Treatment Response to DNMT Inhibitors in Myelodysplasia and Related Neoplasms.” *Leukemia* 28 (1): 78–87. https://doi.org/10.1038/leu.2013.269.

Trowbridge, Jennifer J., Jonathan W. Snow, Jonghwan Kim, and Stuart H. Orkin. 2009. “DNA Methyltransferase 1 Is Essential for and Uniquely Regulates Hematopoietic Stem and Progenitor Cells.” *Cell Stem Cell* 5 (4): 442–49. https://doi.org/10.1016/j.stem.2009.08.016.

Tsai, Hsing-Chen, Huili Li, Leander Van Neste, Yi Cai, Carine Robert, Feyruz V Rassool, James J Shin, Kirsten M Harbom, Robert Beaty, Emmanouil Pappou, James Harris, Ray-Whay Chiu Yen, Nita Ahuja, Malcolm V Brock, Vered Stearns, David Feller-Kopman, Lonny B Yarmus, Yi-Chun Lin, Alana L Welm, Jean-Pierre Issa, Il Minn, William Matsui, Yoon-Young Jang, Saul J Sharkis, Stephen B Baylin, and Cynthia A Zahnow. 2012. “Transient Low Doses of DNA-Demethylating Agents Exert Durable Antitumor Effects on Hematological and Epithelial Tumor Cells.” *Cancer Cell* 21 (3): 430–46. https://doi.org/10.1016/j.ccr.2011.12.029.

Unnikrishnan, Ashwin, Elli Papaemmanuil, Dominik Beck, Jason W H Wong, Peter J Campbell, and John E Pimanda Correspondence. 2017a. “Integrative Genomics Identifies the Molecular Basis of Resistance to Azacitidine Therapy in Myelodysplastic Syndromes.” *Cell Reports* 20 (3): 572–585. https://doi.org/10.1016/j.celrep.2017.06.067.

———. 2017b. “Integrative Genomics Identifies the Molecular Basis of Resistance to Azacitidine Therapy in Myelodysplastic Syndromes.” *Cell Reports* 20 (3): 572–85. https://doi.org/10.1016/j.celrep.2017.06.067.

Vardiman, James W, Jüergen Thiele, Daniel A Arber, Richard D Brunning, Michael J Borowitz, Anna Porwit, Nancy Lee Harris, Michelle M Le Beau, Eva Hellström-Lindberg, Ayalew Tefferi, Clara D Bloomfield, ES. Jaffe, NL. Harris, H. Stein, JW. Vardiman, SH. Swerdlow, E. Campo, NL. Harris, NL. Harris, ES. Jaffe, H. Stein, JW. Vardiman, NL. Harris, RD. Brunning, GJ. Mufti, JM. Bennett, J. Goasguen, A. Orazi, M. Albitar, NA. Heerema, S. Haskins, RS. Neiman, J. Cools, DJ. DeAngelo, J. Gotlib, C. James, V. Ugo, JP. Le Couedic, R. Kralovics, F. Passamonti, AS. Buser, RL. Levine, M. Wadleigh, J. Cools, A. Tefferi, TL. Lasho, G. Gilliland, AD. Pardanani, RL. Levine, T. Lasho, A. Pardanani, C. Akin, P. Valent, ML. Loh, S. Vattikuti, S. Schubbert, K. Stephens, M. Weaver, KA. Leppig, M. Tartaglia, CM. Niemeyer, A. Fragale, CL. Willman, C. Hirsch-Ginsberg, AC. LeMaistre, H. Kantarjian, RA. Padua, G. Carter, D. Hughes, K. Mrozek, CD. Bloomfield, K. Mrozek, G. Marcucci, P. Paschka, SP. Whitman, CD. Bloomfield, P. Paschka, G. Marcucci, AS. Ruppert, B. Falini, C. Mecucci, E. Tiacci, K. Dohner, RF. Schlenk, M. Habdank, ME. Greene, G. Mundschau, J. Wechsler, RL. Levine, DG. Gilliland, AV. Jones, S. Kreil, K. Zoi, EJ. Baxter,

LM. Scott, PJ. Campbell, LM. Scott, W. Tong, RL. Levine, J. Thiele, HM. Kvasnicka, A. Orazi, E. Lengfelder, A. Hochhaus, U. Kronawitter, A. Regev, P. Stark, D. Blickstein, M. Lahav, S. Sacchi, G. Vinci, L. Gugliotta, BS. Wilkins, WN. Erber, D. Bareford, J. Thiele, HM. Kvasnicka, AM. Florena, C. Tripodo, E. Iannitto, R. Porcasi, S. Ingrao, V. Franco, U. Gianelli, C. Vener, PR. Raviele, TR. Golub, GF. Barker, M. Lovett, DG. Gilliland, P. Keene, B. Mendelow, MR. Pinto, BJ. Bain, SH. Fletcher, EJ. Steer, NC. Cross, D. Macdonald, A. Reiter, NC. Cross, LV. Abruzzo, ES. Jaffe, JD. Cotelingam, J. Whang-Peng, V. Del Duca, LJ. Medeiros, G. Metzgeroth, C. Walz, J. Score, H. Nagata, AS. Worobec, CK. Oh, RL. Levine, M. Loriaux, BJ. Huntly, DP. Steensma, GW. Dewald, TL. Lasho, K. Sugimoto, N. Hirano, H. Toyoshima, U. Germing, N. Gattermann, C. Strupp, M. Aivado, C. Aul, U. Germing, C. Strupp, S. Knipp, E. Hellstrom-Lindberg, JM. Raya, L. Arenillas, A. Domingo, M. Boissinot, R. Garand, M. Hamidou, S. Hermouet, AF. Remacha, JF. Nomdedeu, G. Puget, JW. Vardiman, D. Bowen, D. Culligan, S. Jowitt, GA. Kennedy, TD. Kay, DW. Johnson, R. Brunning, R. Mckenna, PA. Kouides, JM. Bennett, BJ. Bain, JP. Maciejewski, A. Risitano, EM. Sloand, O. Nunez, NS. Young, DP. Steensma, GW. Dewald, JM. Hodnefield, A. Tefferi, CA. Hanson, V. Gupta, C. Brooker, JA. Tooze, RV. Pierre, HC. Hoagland, MG. Della Porta, L. Malcovati, R. Invernizzi, SJ. Kussick, BL. Wood, M. Stetler-Stevenson, DC. Arthur, N. Jabbour, F. Wimazal, C. Fonatsch, R. Thalhammer, C. Strupp, N. Gattermann, A. Giagounidis, G. Lambertenghi-Deliliers, A. Orazi, R. Luksch, C. Annaloro, D. Soligo, N. Tuzuner, C. Cox, JM. Rowe, D. Watrous, JM. Bennett, DP. Steensma, CA. Hanson, L. Letendre, A. Tefferi, MT. Elghetany, SD. Hudnall, FH. Gardner, J. Barrett, Y. Sauntharajah, J. Molldrem, H. Hasle, CM. Niemeyer, JM. Chessells, LM. Kelly, DG. Gilliland, RF. Schlenk, K. Dohner, J. Krauter, C. Thiede, S. Koch, E. Creutzig, A. Zelent, F. Guidez, A. Melnick, S. Waxman, JD. Licht, C. Meyer, B. Schneider, S. Jakob, LY. Shih, DC. Liang, JF. Fu, MA. Caligiuri, MP. Strout, D. Lawrence, MA. Bitter, ME. Neilly, MM. Le Beau, MG. Pearson, JD. Rowley, LM. Secker-Walker, A. Mehta, B. Bain, MG. Pearson, JW. Vardiman, MM. Le Beau, CP. Soupir, JA. Vergilio, P. Dal Cin, YK. Keung, M. Beaty, BL. Powell, I. Molnar, D. Buss, M. Pettenati, E. Paietta, J. Racevskis, JM. Bennett, DA. Arber, AS. Stein, NH. Carter, D. Ikle, SJ. Forman, ML. Slovak, M. Yanada, M. Suzuki, K. Kawashima, OK. Weinberg, M. Seetharam, L. Ren, H. Wandt, U. Schakel, F. Kroschinsky, T. Haferlach, C. Schoch, H. Loffler, SM. Smith, MM. Le Beau, D. Huo, ZN. Singh, D. Huo, J. Anastasi, MK. Andersen, RA. Larson, N. Mauritzson, S. Schnittger, SC. Jhanwar, J. Pedersen-Bjergaard, CD. Bloomfield, KJ. Archer, K. Mrozek, JD. Rowley, HJ. Olney, ML. Slovak, V. Bedell, L. Popplewell, DA. Arber, C. Schoch, R. Slater, SA. Pileri, S. Ascani, MC. Cox, S. Gurbuxani, P. Vyas, JD. Crispino, DS. Brink, GV. Massey, A. Zipursky, MN. Chang, M. Cella, D. Jarrossay, F. Facchetti, L. Chaperot, N. Bendriss, O. Manches, T. Petrella, MR. Comeau, M. Maynadie, ME.

Pilichowska, MD. Fleming, JL. Pinkus, GS. Pinkus, E. Matutes, R. Morilla, N. Farahat, CA. Hanson, M. Abaza, S. Sheldon, CW. Ross, B. Schnitzer, LM. Stoolman, LE. Sulak, CN. Clare, BA. Morale, KL. Hansen, MM. Montiel, D. Campana, TE. Hansen-Hagge, E. Matutes, F. Brito-Babapulle, H. Pullon, DM. Layton, A. Cuneo, A. Ferrant, JL. Michaux, B. Johansson, F. Mertens, F. Mitelman, F. Mittelman, A. Borkhardt, G. Cazzaniga, S. Viehmann, CD. Bloomfield, AI. Goldman, G. Alimena, CH. Pui, LL. Robison, AT. Look, SA. Armstrong, and AT. Look. 2009. “The 2008 Revision of the World Health Organization (WHO) Classification of Myeloid Neoplasms and Acute Leukemia: Rationale and Important Changes.” *Blood* 114 (5): 937–51. https://doi.org/10.1182/blood-2009-03-209262.

Velten, Lars, Simon F. Haas, Simon Raffel, Sandra Blaszkiewicz, Saiful Islam, Bianca P. Hennig, Christoph Hirche, Christoph Lutz, Eike C. Buss, Daniel Nowak, Tobias Boch, Wolf-Karsten Hofmann, Anthony D. Ho, Wolfgang Huber, Andreas Trumpp, Marieke A. G. Essers, and Lars M. Steinmetz. 2017. “Human Haematopoietic Stem Cell Lineage Commitment Is a Continuous Process.” *Nature Cell Biology* 19 (4): 271. https://doi.org/10.1038/ncb3493.

Verma, Nipun, Heng Pan, Louis C. Doré, Abhijit Shukla, Qing V. Li, Bobbie Pelham-Webb, Virginia Teijeiro, Federico González, Andrei Krivtsov, Chan-Jung Chang, Eirini P. Papapetrou, Chuan He, Olivier Elemento, and Danwei Huangfu. 2018. “TET Proteins Safeguard Bivalent Promoters from de Novo Methylation in Human Embryonic Stem Cells.” *Nature Genetics* 50 (1): 83–95. https://doi.org/10.1038/s41588-017-0002-y.

Waddington, C. H. 1942. “The Epigenotype.” *Endeavour* 1: 18–20.

Walter, M J, L Ding, D Shen, J Shao, M Grillot, M McLellan, R Fulton, H Schmidt, J Kalicki-Veizer, M O’Laughlin, C Kandoth, J Baty, P Westervelt, J F DiPersio, E R Mardis, R K Wilson, T J Ley, and T A Graubert. 2011. “Recurrent DNMT3A Mutations in Patients with Myelodysplastic Syndromes.” *Leukemia* 25 (7): 1153–58. https://doi.org/10.1038/leu.2011.44.

Walter, M J, D Shen, J Shao, L Ding, B S White, C Kandoth, C A Miller, B Niu, M D McLellan, N D Dees, R Fulton, K Elliot, S Heath, M Grillot, P Westervelt, D C Link, J F DiPersio, E Mardis, T J Ley, R K Wilson, and T A Graubert. 2013. “Clonal Diversity of Recurrently Mutated Genes in Myelodysplastic Syndromes.” *Leukemia* 27 (6): 1275–82. https://doi.org/10.1038/leu.2013.58.

Wang, Fang, Jeremy Travins, Byron DeLaBarre, Virginie Penard-Lacronique, Stefanie Schalm, Erica Hansen, Kimberly Straley, Andrew Kernytsky, Wei Liu, Camelia Gliser, Hua Yang, Stefan Gross, Erin Artin, Veronique Saada,

Elena Mylonas, Cyril Quivoron, Janeta Popovici-Muller, Jeffrey O Saunders, Francesco G Salituro, Shunqi Yan, Stuart Murray, Wentao Wei, Yi Gao, Lenny Dang, Marion Dorsch, Sam Agresta, David P Schenkein, Scott A Biller, Shinsan M Su, Stephane de Botton, and Katharine E Yen. 2013. “Targeted Inhibition of Mutant IDH2 in Leukemia Cells Induces Cellular Differentiation.” *Science (New York, N.Y.)* 340 (6132): 622–26. https://doi.org/10.1126/science.1234769.

Weber, Alain R., Claudia Krawczyk, Adam B. Robertson, Anna Kuśnierczyk, Cathrine B. Vågbø, David Schuermann, Arne Klungland, and Primo Schär. 2016. “Biochemical Reconstitution of TET1-TDG-BER-Dependent Active DNA Demethylation Reveals a Highly Coordinated Mechanism.” *Nature Communications* 7 (March): 10806. https://doi.org/10.1038/ncomms10806.

Weigele, Peter, and Elisabeth A. Raleigh. 2016. “Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses.” *Chemical Reviews* 116 (20): 12655–87. https://doi.org/10.1021/acs.chemrev.6b00114.

Welch, John S., Allegra A. Petti, Christopher A. Miller, Catrina C. Fronick, Michelle O’Laughlin, Robert S. Fulton, Richard K. Wilson, Jack D. Baty, Eric J. Duncavage, Bevan Tandon, Yi-Shan Lee, Lukas D. Wartman, Geoffrey L. Uy, Armin Ghobadi, Michael H. Tomasson, Iskra Pusic, Rizwan Romee, Todd A. Fehniger, Keith E. Stockerl-Goldstein, Ravi Vij, Stephen T. Oh, Camille N. Abboud, Amanda F. Cashen, Mark A. Schroeder, Meagan A. Jacoby, Sharon E. Heath, Kierstin Luber, Megan R. Janke, Andrew Hantel, Niloufer Khan, Madina J. Sukhanova, Randall W. Knoebel, Wendy Stock, Timothy A. Graubert, Matthew J. Walter, Peter Westervelt, Daniel C. Link, John F. DiPersio, and Timothy J. Ley. 2016a. “TP53 and Decitabine in Acute Myeloid Leukemia and Myelodysplastic Syndromes.” *New England Journal of Medicine* 375 (21): 2023–36. https://doi.org/10.1056/NEJMoa1605949.

———. 2016b. “TP53 and Decitabine in Acute Myeloid Leukemia and Myelodysplastic Syndromes.” *New England Journal of Medicine* 375 (21): 2023–36. https://doi.org/10.1056/NEJMoa1605949.

Westers, T M, R Ireland, W Kern, C Alhan, J S Balleisen, P Bettelheim, K Burbury, M Cullen, J A Cutler, M G Della Porta, A M Dräger, J Feuillard, P Font, U Germing, D Haase, U Johansson, S Kordasti, M R Loken, L Malcovati, J G te Marvelde, S Matarraz, T Milne, B Moshaver, G J Mufti, K Ogata, A Orfao, A Porwit, K Psarra, S J Richards, D Subirá, V Tindell, T Vallespi, P Valent, V H J van der Velden, T M de Witte, D A Wells, F Zettl, M C Béné, and A A van de Loosdrecht. 2012. “Standardization of Flow Cytometry in Myelodysplastic Syndromes: A Report from an International Consortium and the European LeukemiaNet Working Group.” *Leukemia* 26 (7): 1730–41. https://doi.org/10.1038/leu.2012.30.

Witte, Theo de, David Bowen, Marie Robin, Luca Malcovati, Dietger Niederwieser, Ibrahim Yakoub-Agha, Ghulam J. Mufti, Pierre Fenaux, Guillermo Sanz, Rodrigo Martino, Emilio Paolo Alessandrino, Francesco Onida, Argiris Symeonidis, Jakob Passweg, Guido Kobbe, Arnold Ganser, Uwe Platzbecker, Jürgen Finke, Michel van Gelder, Arjan A. van de Loosdrecht, Per Ljungman, Reinhard Stauder, Liisa Volin, H. Joachim Deeg, Corey Cutler, Wael Saber, Richard Champlin, Sergio Giralt, Claudio Anasetti, and Nicolaus Kröger. 2017. “Allogeneic Hematopoietic Stem Cell Transplantation for MDS and CMML: Recommendations from an International Expert Panel.” *Blood* 129 (13): 1753–62. https://doi.org/10.1182/blood-2016-06-724500.

Wolach, Ofir, and Richard Stone. 2016. “Autoimmunity and Inflammation in Myelodysplastic Syndromes.” *Acta Haematologica* 136 (2): 108–17. https://doi.org/10.1159/000446062.

Woll, Petter S., Una Kjällquist, Onima Chowdhury, Helen Doolittle, David C. Wedge, Supat Thongjuea, Rikard Erlandsson, Mtakai Ngara, Kristina Anderson, Qiaolin Deng, Adam J. Mead, Laura Stenson, Alice Giustacchini, Sara Duarte, Eleni Giannoulatou, Stephen Taylor, Mohsen Karimi, Christian Scharenberg, Teresa Mortera-Blanco, Iain C. Macaulay, Sally-Ann Clark, Ingunn Dybedal, Dag Josefsen, Pierre Fenaux, Peter Hokland, Mette S. Holm, Mario Cazzola, Luca Malcovati, Sudhir Tauro, David Bowen, Jacqueline Boultwood, Andrea Pellagatti, John E. Pimanda, Ashwin Unnikrishnan, Paresh Vyas, Gudrun Göhring, Brigitte Schlegelberger, Magnus Tobiasson, Gunnar Kvalheim, Stefan N. Constantinescu, Claus Nerlov, Lars Nilsson, Peter J. Campbell, Rickard Sandberg, Elli Papaemmanuil, Eva Hellström-Lindberg, Sten Linnarsson, and Sten Eirik W. Jacobsen. 2014. “Myelodysplastic Syndromes Are Propagated by Rare and Distinct Human Cancer Stem Cells In Vivo.” *Cancer Cell* 25 (6): 794–808. https://doi.org/10.1016/j.ccr.2014.03.036.

Yamamoto, Ryo, Yohei Morita, Jun Ooehara, Sanae Hamanaka, Masafumi Onodera, Karl Lenhard Rudolph, Hideo Ema, and Hiromitsu Nakauchi. 2013. “Clonal Analysis Unveils Self-Renewing Lineage-Restricted Progenitors Generated Directly from Hematopoietic Stem Cells.” *Cell* 154 (5): 1112–26. https://doi.org/10.1016/j.cell.2013.08.007.

Yamazaki, Jumpei, Jaroslav Jelinek, Yue Lu, Matteo Cesaroni, Jozef Madzo, Frank Neumann, Rong He, Rodolphe Taby, Aparna Vasanthakumar, Trisha Macrae, Kelly R. Ostler, Hagop M. Kantarjian, Shoudan Liang, Marcos R. Estecio, Lucy A. Godley, and Jean-Pierre J. Issa. 2015. “TET2 Mutations Affect Non-CpG Island DNA Methylation at Enhancers and Transcription Factor-Binding Sites in Chronic Myelomonocytic Leukemia.” *Cancer*

Research 75 (14): 2833–43. https://doi.org/10.1158/0008-5472.CAN-14-0739.

Yamazaki, Jumpei, Rodolphe Taby, Aparna Vasanthakumar, Trisha Macrae, Kelly R Ostler, Lanlan Shen, Hagop M Kantarjian, Marcos R Estecio, Jaroslav Jelinek, Lucy A Godley, and Jean-Pierre J Issa. 2012. “Effects of TET2 Mutations on DNA Methylation in Chronic Myelomonocytic Leukemia.” *Epigenetics : Official Journal of the DNA Methylation Society* 7 (2): 201–7. https://doi.org/10.4161/epi.7.2.19015.

Yang, Xiaojing, Han Han, Daniel D. De Carvalho, Fides D. Lay, Peter A. Jones, and Gangning Liang. 2014. “Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer.” *Cancer Cell* 26 (4): 577–90. https://doi.org/10.1016/j.ccr.2014.07.028.

Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. “ClusterProfiler: An R Package for Comparing Biological Themes among Gene Clusters.” *Omics : A Journal of Integrative Biology*. https://doi.org/10.1089/omi.2011.0118.

Zhang, Xiaotian, Jianzhong Su, Mira Jeong, Myunggon Ko, Yun Huang, Hyun Jung Park, Anna Guzman, Yong Lei, Yung-Hsin Huang, Anjana Rao, Wei Li, and Margaret A. Goodell. 2016. “DNMT3A and TET2 Compete and Cooperate to Repress Lineage-Specific Transcription Factors in Hematopoietic Stem Cells.” *Nature Genetics* 48 (9): 1014–23. https://doi.org/10.1038/ng.3610.

Zhang, Yang W., Zhihong Wang, Wenbing Xie, Yi Cai, Limin Xia, Hariharan Easwaran, Jianjun Luo, Ray-Whay Chiu Yen, Yana Li, and Stephen B. Baylin. 2017. “Acetylation Enhances TET2 Function in Protecting against Abnormal DNA Methylation during Oxidative Stress.” *Molecular Cell* 65 (2): 323–35. https://doi.org/10.1016/j.molcel.2016.12.013.

Zhao, Xiaoli, Feng Yang, Shuang Li, Meixia Liu, Shaoxu Ying, Xiaodong Jia, and Xiaoqin Wang. 2014. “CpG Island Methylator Phenotype of Myelodysplastic Syndrome Identified through Genome-Wide Profiling of DNA Methylation and Gene Expression.” *British Journal of Haematology* 165 (5): 649–58. https://doi.org/10.1111/bjh.12811.

Ziller, Michael J., Fabian Müller, Jing Liao, Yingying Zhang, Hongcang Gu, Christoph Bock, Patrick Boyle, Charles B. Epstein, Bradley E. Bernstein, Thomas Lengauer, Andreas Gnirke, and Alexander Meissner. 2011. “Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types.” *PLOS Genetics* 7 (12): e1002389. https://doi.org/10.1371/journal.pgen.1002389.