Widespread Epigenetic Reprogramming in Aging Human Hematopoietic Stem Cells

by

Emmalee R. Adelman

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular and Cellular Pathology)
in The University of Michigan
2018

Doctoral Committee:

Associate Professor Maria E. Figueroa, University of Miami, Co-chair
Assistant Professor Andrew G. Muntean, Co-chair
Professor Yali Dou
Professor Ivan P. Maillard, University of Pennsylvania
Professor Richard A. Miller

adelmane@umich.edu

ORCID iD: 0000-0003-3429-0803

DEDICATION

"They gathered sticks, and kindled a fire, and left it burning"
-Quaker proverb

To my Grandma Lee, who read *Science* just for fun.

ACKNOWLEDGEMENTS

No scientist is an island, and throughout my graduate experience, I have been surrounded by a community that has helped to make this work possible. Firstly, I would like to thank my advisor, Dr. Maria “Ken” Figueroa. Long ago, when I told Ken that I would like to join her lab, she leaped up, shouted in triumph, and gave me a hug. Five years and one cross-country move later, she still has maintained that same level of enthusiasm. No matter if an RO1 deadline is looming, or if we are on separate continents, she has always taken the time to advise. From writing a catchy abstract title, to performing bioinformatic analysis of DNA methylation data, the skills that she has taught me are innumerable. She has always demanded the best and provided the advice and resources to achieve it. During my graduate education, she has been my biggest supporter in science and truly lives up to the role of mentor. I would also like to thank my committee members, Drs. Yali Dou, Ivan Maillard, Richard Miller, and Andrew Muntean, whose experience and insight were a constant guidance. As I was writing this dissertation I was impressed by their foresight in the challenges I would encounter and the solutions that would best address them.

In addition, I would like to thank the University of Michigan MCP program, especially Dr. Zaneta Nikolovska-Coleska and Laura Labut. I sincerely appreciate the resources and opportunities MCP has provided and the gift of the education that I have

received. I would also like to thank the University of Michigan Biology of Aging training program for challenging me to be a better scientist and providing a collegial environment to learn about biogerontology.

My time in graduate school was also greatly shaped by the members of the Figueroa lab, both past and present. They have been the first responders during my PhD training. When an experiment went wrong, they were the first to help, and if it worked, the first to celebrate in its success. Over the years, I've been lucky to call them my friends. I would especially like to thank Antonio Colaprico for performing the ChIP-seq permutation analysis, Kristen Meldi for being a role-model in lab and life, Miguel Torres-Martin for always providing a listening ear and advice (and helping to escape from hurricanes), Emily Huang for her contribution to this work and for dragging me out of the lab and to the beach, and Jian Shi, whose ever-present smile and calm demeanor brightened up the lab.

Finally, I would like to thank my friends and family. A special shout-out to Caitlin, Zach, Kerry, Matt, Meaghan, Kevin, and Claire, who can always make me laugh and share in my love of pizza. To Sensei Iha, who instilled in me to "always make it better", to my grandparents, four formative pillars of my life, to Jacob and Sara, my oldest friends, and finally to my parents, who gave me the freedom to be whoever I wanted to be and the unwavering support to do so, I am eternally thankful.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ALL ..... Acute lymphoblastic leukemia
AML ..... Acute myeloid leukemia
ARCH ..... Age-related clonal hematopoiesis
Bal-HSC ..... Balanced hematopoietic stem cell
BER ..... Base excision repair
CHIP ..... Clonal hematopoiesis of in determinant potential
ChIP-seq ..... Chromatin immunoprecipitation with sequencing
CLP ..... Common lymphoid progenitor
CMP ..... Common myeloid progenitor
DDR ..... DNA damage response
DHMR ..... Differentially hydroxymethylated region
DMR ..... Differentially methylated region
ERRBS ..... Enhanced reduced representation bisulfite sequencing
GMP ..... Granulocyte-monocyte progenitors
hmC ..... Hydroxymethyl-cytosine
hmeDIP-seq ..... Hydroxymethylated DNA immunoprecipitation sequencing
HSC ..... Hematopoietic stem cell
HSCe ..... Hematopoietic stem cell enriched
HSPC ..... Hematopoietic stem and progenitor cells
ICUS ..... Isolated cytopenia of undetermined significance
Ly-HSC ..... Lymphoid biased hematopoietic stem cell
LT-HSC ..... Long-term hematopoietic stem cell
mC ..... Methyl cytosine
MDS ..... Myelodysplastic syndromes
MEP ..... Megakaryocyte-erythroid progenitors
My-HSC ..... Myeloid biased hematopoietic stem cell
NHEJ ..... Non-homologous end joining
OHPOS ..... Oxidative phosphorylation
PTM ..... Post-translational modification
ST-HSC ..... Short-term hematopoietic progenitor
sc-RNA-seq ..... Single-cell RNA-seq
T-ALL ..... T-cell acute lymphoblastic leukemia

ABSTRACT

At the root of the hematopoietic hierarchy reside the Janus-faced hematopoietic stem cells (HSC), capable of both self-renewal and differentiation. Aging impairs HSC function, leading to increased self-renewal, reduced homing ability and a myeloid differentiation bias. In addition, hematopoietic cells acquire somatic mutations as they age, frequently affecting epigenetic modifier genes. In this dissertation work, I provide a comprehensive characterization of epigenomic changes during normal human HSC aging and demonstrate that aged HSCs undergo widespread reduction in H3K27ac, H3K4me1 and H3K4me3, with little change in H3K27me3. Age-associated loss of enrichment of the activating histone marks H3K27ac and H3K4me3 was particularly prominent at active enhancers and bivalent promoters, respectively. Functional annotation of enhancers lost with age suggests that enhancer deregulation may contribute to HSC myeloid bias and the immune impairments observed in older individuals. Focal changes in DNA methylation were also observed with age, affecting WNT and cadherin associated pathways, and at regions that may predispose to leukemogenesis. DNA 5-hydroxymethylation displayed age-related gains, targeting GATA and KLF binding sites. Concurrent with these epigenetic changes were transcriptional downregulation and mis-splicing of epigenetic modifiers, spliceosome components, transcription factors, including many in the KLF family, and *LMNA*, which is mutated in Hutchinson-Gilford progeria syndrome. Together, these

results establish that multiple levels of epigenetic deregulation with age converge on key hematopoietic regulatory genes and pathways contributing to aged HSC dysfunction.

CHAPTER 1

Introduction

Hematopoietic stem cells are impaired with age

Hematopoietic stem cells

HSCs give rise to the entire hematopoietic system, yet are extremely rare; only 1 in 3,000,000 human bone marrow cells is transplantable[1]. They are mostly quiescent and reside primarily in the bone marrow niche, close to sinusoids. In steady-state murine hematopoiesis, it appears that HSC may divide to self-renew only 4-5 times during a mouse's lifetime[2-4]. However, in times of stress, HSC can rapidly be induced to proliferate[5-7]. Advances in flow cytometry and identification of numerous cell surface markers have allowed for the isolation of fairly pure HSC. However, the criteria for the panel of cell surface antigens that define a "HSC" is frequently refined, making some comparisons across studies difficult. HSCs include both long-term HSCs (LT-HSC; in mouse (Lineage-, CD150+, CD48-, CD41-) and in human (Lineage-, CD34+, CD38-, CD90+, CD45RA-)) which are capable of permanent long-term reconstitution of the bone marrow, as well as short-term HSCs (ST-HSC) which provide only transient bone marrow reconstitution. While the HSC is the only cell capable of stable long-term reconstitution, other progenitors such as the multipotent progenitor (MPP) and ST-HSC most likely are

the main cells contributing to steady-state hematopoiesis[4,8]. However, HSC are the only hematopoietic stem cell capable of both self-renewal and differentiation.

Previously, it was thought that an "HSC is an HSC is an HSC", but both murine and human studies have now shown that HSCs are diverse in their lineage potential. At least three types of HSC have been identified: lymphoid-biased (Ly-HSC), myeloid-biased (My-HSC) and balanced HSC (Bal-HSC). All three types are capable of differentiating into myeloid, lymphoid, or erythroid cells, but Ly-HSC give rise to a higher proportion of lymphoid/myeloid cells *in vitro* and *in vivo* and the opposite is true of My-HSC. True to their name, Bal-HSC have no predilection towards either lineage[9-13]. Recently, a number of studies have suggested that HSC can also be biased towards the megakaryocytic lineage. von Willebrand factor (vWF) is a gene that is highly expressed in megakaryocytes. HSC that highly express vWF give rise to more platelets and myeloid cells when transplanted compared to HSC with scant vWF[14,15]. Single-cell RNA-seq also identified a subset of HSC that highly express vWF, and cluster with the megakaryocyte-erythroid progenitors (MEPs), common myeloid progenitors (CMPs), and granulocyte-monocyte progenitors (GMPs)[16]. Finally, using genetic lineage tracing[17], a subset of LT-HSC that is restricted to the megakaryocytic/platelet lineage was identified. In the bone marrow, myeloid and platelet biased HSCs reside close to sinusoidal endothelial cells and megakaryocytes, whereas Ly-HSCs are found closer to Nestin-GFP+ perivascular cells[18], highlighting the importance of the stem cell niche in regulating HSC fate.

Loss of HSC function with age

Aging is progressive decline leading to an increased likelihood of death. It is the main risk factor for numerous diseases, including cancer[19]. As the world's population ages, research into aging and its associated pathologies is vital[20]. Studies in, *C. elegans*, Drosophila, and mice have shown that aging is plastic; select genetic, chemical, and dietary interventions can increase lifespan and healthspan[21,22]. However, it is unlikely that any one gene is the Philosopher's stone, as aging is a complex process in which it can be difficult to differentiate the causes from the symptoms. López-Otín and colleagues have identified nine hallmarks of aging: stem cell exhaustion, altered intercellular communication, genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, and cellular senescence[19]. Many of these features have been observed in aged HSC and contribute to loss of function with age.

Much of what is known about HSC aging has been derived from murine studies, and while strain-to-strain variations have been observed, it is clear that aged murine HSC have altered reconstitution and lineage potential. Initial studies with flow cytometric analysis found that there is a 3-7-fold increase in the frequency of HSC in the bone marrow of aged (22-25 mo) C57BL/6J and BALB/cByJ mice compared to young (2-7 mo)[23,24]. Additional experiments using the cobblestone area-forming cell (CAFC) assay, in which colony development can be used to estimate stemness and primitiveness, showed more strain-to-strain variation with age. Measuring HSC frequency using the CAFC assay, there is an increase in HSC with age (>20 mo) in C57BL/6J mice, but a decrease in HSC number

in DBA/2 mice[25-27]. Notably, while many other strains have increased HSC frequency with age, the degree to which it changes seems to be strain dependent[25,28].

Even though the number of HSCs increases with age, their function declines. Aged HSC are one fourth as efficient as young HSC at homing to the bone marrow when transplanted into irradiated mice[23,29]. Aging also increases HSC self-renewal under transplant conditions, impairing their ability to differentiate and reconstitute the marrow long-term[30-32]. The advent of refined HSC cell markers and the competitive transplant assay (in which HSC are co-transplanted with a fixed number of "normal" cells) allowed for a more detailed examination of aged HSC function. These assays showed that aged HSCs have decreased B-cell output with increased myeloid cell frequency as measured by flow cytometry of peripheral blood[30]. Aged HSCs also have an increase in gene expression of myeloid associated genes, with a concomitant decrease in lymphoid gene expression[33]. This myeloid skewing with age may be caused in part by early alterations in HSC differentiation, as the granulocyte-megakaryocyte progenitors (GMP) are increased with age, whereas there is a decrease with age of common lymphoid progenitors (CLP)[30]. Notably, while the ability of Ly-HSCs or My-HSCs to generate their respective linages does not change with age, the composition of the HSC pool does. With age, there is an increase in My-HSCs, and over a 50-fold increase in platelet biased HSCs, with a decrease in Ly-HSCs and Bal-HSCs[9,10,18]. This suggests that with age, myeloid and platelet biased HSCs clones become dominant. Intriguingly, the more times that an HSC has divided, the less multipotent and more skewed towards myeloid differentiation it becomes[4]. This hints that epigenetic alterations may be at play in HSCs becoming less potent with age.

Whether human HSCs have the same functional impairments with age is unclear. HSCs from aged donors do have reduced transplantation success in bone marrow transplants, suggesting there is loss of function with age[34]. Similar to what has been observed in mice, there is an increase in HSC frequency in human bone marrow with age[35-37]. However, NSG xenotransplant studies of young and aged human HSC into humanized mice have generated conflicting results, that are not always in concordance with the aged murine HSC phenotype. Unexpectedly, no decrease in HSC engraftment or donor chimerism was observed in transplanted aged human HSC compared to young[35,36]. Yet like the murine studies, one group found that the bone marrow myeloid/lymphoid ratio was increased in mice transplanted with aged HSC (Lin-, CD34+, CD38-, CD90+, CD45RA-) compared to young. This age-associated myeloid skewing was confirmed *in vitro* using a stromal co-culture system[35]. However, in another study, where less pure HSC (CD34+, CD38-) were used, mice transplanted with aged HSC displayed no deficiencies in lymphoid output and actually had fewer bone marrow myeloid cells compared to mice that received young HSCs. Colony assays confirmed the decreased myeloid output in aged HSCs and showed that while aged individuals had fewer lymphoid progenitors, B-cell (CD19+) production was not impaired with age. Perhaps the conflicting findings are due to differences in the purity of "HSCs" used in these studies. However, in a recent study that utilized a stromal co-culture system and a highly pure HSC population (Lin-,CD34+, CD90+, CD45-, CD123-), similar to that used by the researchers who observed myeloid skewing, no difference in myeloid/lymphoid output with age was found[37]. Given the conflicting results and the now known heterogeneity within the HSC compartment, single-cell assays

using highly pure HSC will be needed in order to determine if aged human HSC are truly skewed towards myeloid differentiation with age.

DNA damage with HSC aging

Impaired DNA damage response (DDR) has long been implicated in cellular decline with aging. Quiescence protects HSC from DNA damage during cell division, but HSC in this state rely on error-prone non-homologous end-joining (NHEJ) for DNA damage repair[38]. The DDR is clearly important for HSC function. However, whether increased DNA damage with age contributes to HSC loss of function is somewhat controversial. Transplant experiments have shown that mutations in genes associated with NHEJ ($Ku80^{-/-}$ mouse model), nucleotide excision repair (XPD^{TTD} mouse model), mismatch repair ($MSH2^{-/-}$ mouse model), or telomere maintenance ($mTR^{-/-}$ mouse model) cause decreased HSC self-renewal and insufficient long-term reconstitution[39-43]. Similar to what has been found with normal aging in C57/B6 mice, mutations in $Ku80^{-/-}$, $mTR^{-/-}$, or XPD^{TTD} also result in a higher frequencies of HSC in the bone marrow (irrespective of mouse age)[40].

Some of the most compelling evidence for increased DNA damage with HSC aging is from studies measuring phosphorylated histone H2AX (γ–H2AX), which is a marker for unresolved DNA double stranded breaks (DSB)[40,44-46]. Over 80% of aged (122 weeks) murine HSCs have γ–H2AX foci, compared to <20% of young (10 weeks) HSCs. Of the aged HSCs with foci, 70% of cells had multiple foci[40]. This phenomena has been observed in human HSC aging, albeit with a more modest difference between young and aged HSCs[47]. Use of the alkaline comet assay, which is a single cell gel electrophoresis assay used to

measure DNA single stranded and double stranded breaks, has also shown an increase in DNA damage with age[44,48]. After exposure to gamma radiation, highly purified aged HSC had significantly more DNA damage than young HSC, although 30% of aged HSC had no DSB. Importantly, this DNA damage was repaired once HSC exited from quiescence[44]. Another group used heterozygosity at microsatellite repeats as a surrogate marker of DNA mutations and found that there is 2-3 fold increase in the number of mutations in aged hematopoietic progenitors compared to young[46]. Increased DNA damage with age may contribute to decreased lymphopoiesis. A recent study found that period circadian gene 2 (Per2) mediates the induction of DNA replication stress signaling in HSCs. Expression of Per2 was increased with radiation and in non-irradiated aged Ly-HSCs (but not in My-HSCs). Genetic deletion of Per2 ablated the age-associated decrease in Ly-HSCs and B-cells[49]. Thus Ly-HSCs may be more sensitive to DNA damage than My-HSCs. However, more studies will be needed to identify the role of DNA damage in controlling the lineage potential of HSCs with aging.

Recent work has challenged whether the DDR contributes to HSC loss of function with age. It was found that in response to radiation, young and aged HSCs do not differ in their DDR, and both age-groups bypassed the G1-S cell-cycle checkpoint, as has also been observed in embryonic stem cells[46]. Additionally, no co-localization of DNA damage response proteins and γ–H2AX in HSC has been observed, and γ–H2AX has been shown to be associated with silenced ribosomal genes in HSC, indicating that increased γ–H2AX with age may not be due to increased DNA damage[45].

Metabolism and oxidative stress in HSC aging

The free radical theory of aging, that endogenous metabolic processes generate free radicals which contribute to aging, was first proposed in 1956[50]. In regards to HSCs, alterations of reactive oxygen species (ROS) and cellular metabolism do contribute to loss of function, although how they are involved in HSC aging is less clear. The hypoxic niche protects HSCs from oxidative stress, as they are extremely sensitive to ROS[51-53]. Increased levels of ROS induce decreased colony forming ability *in vitro* and reduced reconstitution *in vivo*[53]. Studies using mice with a genetic deletion of Ataxia telangiectasia mutated (ATM) a kinase that is vital for maintaining genomic stability in response to DNA damage, or conditional knockouts of FOXO1, FOXO3, and FOXO4, transcription factors that regulate stress resistance, apoptosis, and cell cycle arrest, have shown the influence of ROS on HSC differentiation and survival. Conditional deletion of FOXO1, FOXO3, and FOXO4 ($FoxO1/3/4^{L/L}$), causes myeloid expansion and apoptosis of the LT-HSC compartment. This phenotype is mediated in part by ROS, as treatment with the antioxidant N-acetyl-L-cysteine (NAC) restores HSC numbers and differentiation[54]. Similarly, $ATM^{-/-}$ HSCs have decreased self-renewal and repopulating ability, but treatment with NAC abolishes these defects[55]. Additionally, when HSCs from young mice are FACS sorted by ROS activity, ROS-high HSCs display decreased reconstitution ability, increased mTOR activity, and myeloid-bias, all characteristics of older HSCs.[56] An initial study of hematopoietic progenitors (Lineage-, Sca1+, cKit+) found that young HSCs have very low levels of ROS, but that intercellular ROS increases with age[53]. However, a more

recent study using a more purified HSC population, reported that aged HSCs have decreased levels of ROS compared to young[57].

Unlike differentiated cells, HSC utilize anaerobic glycolysis rather than oxidative phosphorylation (OXPHOS)[58-60]. While not as energetic as OXPHOS, glycolysis generates less ROS and may be an adaptation to lower oxygen availability in the niche[61]. However, in order to differentiate, HSC must switch to OXPHOS, which generates more ATP[62]. With age, murine HSCs show signs of mitochondrial damage, but are more metabolically active than young HSCs. HSCs from aged (24-28 months) have increased NAPDH and ATP levels with more OXPHOS and less glycolysis compared to young (6-12 weeks) HSC[57]. Autophagy is necessary to clear mitochondria and preserve HSC regenerative capacity. HSC from mice lacking autophagy-related 12 (Atg12) have a metabolic profile similar to aged HSCs, are myeloid biased, and have decreased engraftment. Furthermore, about 33% of aged HSCs have high autophagy levels and are functionally similar to young HSCs[57]. Of note, we found that loss of autophagy drives these metabolic changes through epigenetic alteration of cytosines. We identified differentially methylated regions, consisting of 162 hypermethylated and 783 hypomethylated, in Atg12 knockout mice compared to control. Notably, regions that became hypomethylated with loss of Atg12 were associated with metabolic pathways (Figure S1)[57].

Studies of the metabolic mammalian target of rapamycin (mTOR) signaling pathway also indicate the importance of metabolism in HSC aging. mTOR signaling is increased in HSCs from elderly mice, and inhibition of mTOR signaling improves aged HSC function. HSCs from mice treated with rapamycin, an mTOR inhibitor, show increased

reconstitution ability, and decreased myeloid bias, resembling HSCs from younger animals[63]. Conversely, increased mTOR signaling, caused by deletion of tumor sclerosis complex (TSC), phenocopies HSC aging[64,65]. However, a recent study using wild-type HSCs of the same purity and mouse strain as the previous study, found no increase in mTOR signaling with HSC age. Additionally, they observed decreased protein synthesis with age[57]. In summary, while increased levels of ROS do cause aging phenotypes, and HSCs are more metabolically active with age, there is still controversy whether ROS increases in aged HSCs and if age-associated changes involve increased mTOR signaling.

There is mounting evidence that metabolism may contribute to HSC lineage fate and myeloid bias with age. My-HSCs have decreased expression of glycolytic enzymes compared to Ly-HSCs, and Ly-HSCs and MY-HSCs can be isolated based on their expression of GRP78, the receptor for Cripto, a protein known to regulate glycolysis[10,66]. Additionally, it appears that Ly-HSCs may be more sensitive to mitochondrial stresses. Murine studies of mitofusion2 (Mfn2), which is involved in mitochondrial fusion, showed that perterbutations of Mfn2 largely affect only HSC of the lymphoid potential[67].

Altered WNT signaling and loss of polarity

WNT signaling is critical for hematopoiesis. Modulation of WNT signaling alters HSC self-renewal and reconstitution ability[68,69]. HSCs express canonical Wnt3a, as well as a number of other WNT proteins[70]. With age, levels of Wnt3a do not change in murine HSCs, however there is increased expression of Wnt5a in middle-aged (10 mo) and old (20-24 mo) HSCs compared to young[71]. Overexpression of Wnt5a in young HSCs results

in an aged phenotype, HSCs have a decreased regenerative capacity and myeloid skewing. Wnt5a mediates these effects through the Rho GTPase CDC42, which is critical for maintaining cell polarity, or the distribution of proteins and organelles.[71,72] Like Wnt5a, expression of CDC42 is increased with age[72]. Increased non-canonical WNT signaling with age may contribute to epigenetic remodeling of HSCs. Aged HSCs have uneven distribution of H4K16ac within the nucleus. This histone modification is involved in regulating replicative aging through the NAD+ dependent histone deacetylase SIRT2[73]. Inhibition of the Wnt5a target CDC42 in old HSC restores H4K16ac polarity and corresponds with improved function[72]. This suggests that there is loss of epigenetic regulation in aged HSC, which contributes to impaired function.

In summary, HSCs are highly quiscent cells that reside in the bone-marrow niche and can rapidly be induced to proliferate under times of stress. While all HSCs are capable of both self-renewal and differentiation, a growing amount of evidence suggests that HSCs can be "biased" towards different cell fates. With age, HSCs display increased self-renewal, decreased reconstitution ability, and an increased predispositon to develop into myeloid, rather than lymphoid cells. Aged HSCs also become more metabolically active and may have impairments in the DNA damage reponse (Figure 1.1). However, most studies of HSC aging have utilized murine models, which have strain-to-strain variations of aging phenotypes, and it is not clear if mouse aging fully recapitulates human HSC aging.

The aging hematopoietic system

Impaired immunity with aging

Older individuals do not respond as vigorously to immune challenges as their younger counterparts. With age, there is an increase in the frequency of bacterial infections from species such as *Mycobacterium tuberculosis* and *Streptococcus pneumonia* and an increase in susceptibility to viral infections such as *herpes zoster* and influenza[74-78]. Furthermore, the mortality rate for some of these infections is 2-3 times higher in the elderly compared to younger individuals[74-78]. Prevention of infection in older people is further complicated because the elderly have a reduced response to vaccination[79]. This deficient immune response is due in part to altered production and function of lymphocytes with age.

Perhaps the most macroscopic change with immune aging is the reduction in the size of the thymus, or age-related thymic involution. The thymus is composed of two main compartments, the thymic epithelial space (TES), which is the primary site for thymopoiesis, or the production of mature T-cell lymphocytes, and the perivascular space (PVS), which does not contain developing lymphocytes. In humans, thymic involution begins as early as 1-year old and accelerates around the onset of puberty[80]. However, the overall size of the human thymus is not as drastically reduced with age compared to that of the mouse, as adipocytes accumulate in the PVS with age[80]. Yet the TES decreases at rate of about 3% per year through middle age (45 yo) and 1% a year thereafter[80,81]. Even though the thymus size is reduced with age, it is still capable of producing healthy T-cells, although at a much lower rate than that of a younger individual[82,83].

While thymic production of T-cells is reduced with age, in adult humans most T-cell production is dependent on the peripheral proliferation of T-cells[84-86]. However, there is alteration of both the CD4+ and CD8+ T-cell populations with age. By measuring T-cell receptor beta (TCRB) sequences, it was found that there is a 2-5-fold decrease in the naïve T-cell repertoire with aging[87]. Furthermore, with age there is an oligoclonal expansion of the naïve and memory CD8+ populations[87-89]. This reduced diversity within the T-cell pool may limit an aged individual's response to newly encountered viruses.

Not only is the T-cell lineage affected with age, but so too is the B-cell lineage. B-cells isolated from human peripheral blood of aged donors show reduced class-switch recombination compared to those from younger donors[90]. Additionally, antibodies produced by B-cells from elderly individuals have reduced specificity and affinity for their target antigens[90,91]. Like CD8+ cells, B-cells also show an increase in clonality with aging[92]. Taken together, impairments in lymphocyte function contribute to loss of adaptive immunity with age.

Increase in cytopenias and clonal hematopoiesis

With age, there is an increase in the rate of clonal hematopoiesis and cytopenias. Specifically, aged individuals have increased rates of anemia. Data generated from 39,695 Americans revealed that 10% of men and 11% of women over 65 years old are anemic, with the frequency increasing to greater than 20% in people over 85 years old[93]. Anemia can pose a serious health risk; elderly anemic individuals have increased rates of "frailty" and mortality compared to non-anemic individuals[94,95]. Notably, while many of the

incidences of anemia in the elderly are due to nutritional deficiency, up to 33% of the cases have unknown causes[93,96].

Currently, cytopenias of unknown causes (such as anemia) that cannot be classified as a hematological disorder or myelodysplastic syndrome (MDS) are categorized as idiopathic cytopenia of undermined significance (ICUS)[97]. A recent prospective study profiled 154 patients with ICUS (median age=53) and found that 25% went on to develop a myeloid neoplasm[98]. Notably, examination of somatic mutations showed that 36% of patients with ICUS had one or more mutations, which has previously been defined as clonal cytopenia of undetermined significance (CCUS)[97,99]. The epigenetic modifiers *TET2*, *ASXL1*, and *DNMT3A* were some of the genes most frequently mutated in these individuals, and they were older than those with ICUS alone (median age=68).

A number of recent studies have shown that this phenotype of clonal hematopoiesis (one progenitor giving rise to a disproportionately high number of mature cells) with somatic mutations in epigenetic modifiers is quite frequent with normal aging[100-104]. In a study by Jaiswal et al., whole exome sequencing was performed using peripheral blood DNA from 17,182 individuals ranging from 20-108 years old. They found a low incidence of mutations in people <40 years old, but an increase in the frequency of mutations with age. Amongst individuals aged 70-79, 9.5% had a clonal mutation, with the majority of variants occurring in *DNMT3A*, *TET2*, and *ASXL1*[101]. These findings have been recapitulated in a number of other studies, and the phenomena dubbed as clonal hematopoiesis of indeterminate potential, or CHIP. Similarly, it is also referred to as age related clonal hematopoiesis (ARCH)[105]. It is postulated that the incidence of CHIP is even

higher than initially reported due to the moderate sequencing depths of the initial studies[100,103,104,106]. While individuals with CHIP do not have any diagnosable hematological ailment, they have 11 times the risk of developing a hematological cancer, although the rate of progression is low (0.5-1% per year)[101]. However, people with CHIP have twice the risk of developing coronary heart disease compared to those who do not have CHIP[102]. Thus, while CHIP is fairly common with age and does not pose an immediate hematologic risk, it is still of clinical significance.

Myeloid malignancies in aged individuals

Advanced age is associated with an increase incidence of myeloid malignancies, whereas cancers of the lymphoid lineages are more frequent in children and young adults (Figure 1.2). Acute myeloid leukemia (AML) is the most common acute leukemia in adults, and the average AML patient age at time of diagnosis is 68 years old[107,108]. Cytogenetic abnormalities and mutations in *NRAS*, *FLT3*, *TP53*, and the epigenetic modifiers *DNMT3A*, *TET2*, *ASXL1*, and *IDH1/2* are frequent in AML. For AML diagnosis, a patient must have >20% blasts in the bone marrow, or one of the following cytogenetic abnormalities t(15;17), t(8;21), inv(16), or t(16;16)[109]. Chemotherapy is the standard treatment for AML, however older individuals have higher rates of treatment related mortality and a poorer prognosis compared to people diagnosed at a younger age[109]. Treatment of AML is especially difficult due to the existence of leukemic stem cells (LSC), rare pathogenic hematopoietic progenitors that harbor mutations and are not completely eradicated with standard chemotherapy[110].

Another age related myeloid malignancy is myelodysplastic syndrome (MDS). In the United States, the incidence of MDS in people under 40 years old is 1 case per 100,000 individuals, but this rate increases to 20 in 100,000 in people aged 70-79 years old[111]. MDS is a heterogeneous disease, in which patients present with peripheral cytopenias, hypocellular dysplastic bone marrow, and possible cytogenetic or mutational abnormalities[112]. Ultimately, this disease progresses to bone marrow failure or AML. Mutations in over 40 genes have been observed in MDS, with the spliceosome components *SRSF2, SF3B1*, and *U2AF1*, and epigenetic modifiers *DNMT3A*, *TET2, ASXL1* being the most often mutated[113]. Notably, aberrant modification of cytosine bases is found in MDS and AML (see "Alterations in cytosine modifications in myeloid malignancies").

In summary, with age there is an increase in clonal hematopoiesis and unexplained cytopenias, impaired adaptive immunity, and an increased frequency to develop myeloid malignancies (Figure 1.3). While both humans and mice have a decline in immune function with age, de novo CHIP, MDS, or AML have not been reported in mice, highlighting the differences in aging hematopoiesis amongst species. Although it is unclear how aging predisposes for myeloid malignancies, it is likely that age-associated HSC myeloid bias is a driving factor. Additionally, given the recent evidence that some of the same epigenetic modifiers are mutated with age and in myeloid neoplasms, it is possible that changes in the epigenetic landscape with age primes cells for myeloid pathogenesis.

Epigenetics

A primer on epigenetics

What determines cell identity? A HSC, cardiomyocyte and β-cell all contain the same genetic information, yet only one of them is capable of hematopoiesis. The answer to this cell-fate riddle is epigenetics, heritable modifications to the genome that do not involve alteration of the DNA sequence itself[114]. The term epigenetics was coined in 1942 by C.H. Waddington as a means to describe how development and genes interact to produce a phenotype[115]. In its contemporary form, epigenetics encompasses an intricate multi-level system of gene regulation; non-coding RNA, and chemical modifications of the proteins that package DNA, or even the DNA bases themselves, can all alter gene expression.

In order to accommodate the large amount of DNA within each diploid cell (up to 2 meters!), there is a complex system of DNA organization[116]. DNA molecules are wrapped around a protein octamer known as the nucleosome core. Each nucleosome is composed of two each of histones H2A, H2B, H3, and H4. There is also a fifth histone type, linker H1, which binds to the outside of the nucleosome core. About 147 base pairs of DNA are wrapped around each nucleosome, compacting the DNA 7-fold[117-120]. The core histones are highly conserved across species, highlighting their importance for cell viability[121]. In addition, there are numerous histone variants, such as H3.3, which play a role in development and cancer[122]. Together, the histones and DNA compose chromatin, which can be highly condensed (heterochromatin) or open (euchromatin). Heterochromatin is often associated with the nuclear lamina, although localization to the lamina does not

always equate with gene silencing[123-126]. Chromatin is further organized into topological associated domains (TADs), large megabase sized regions of interactive chromatin[127-129].

The histone code

Both the histone globular domain and N-terminal tail can be altered through the addition of post-transcriptional modifications (PTM). These modifications include acetylation, methylation, ubiquitinylation, propionylation, butyrylation, sumoylation, phosphorylation, ADP ribosylation, and deamination, and taken together generate the "histone code"[130-138]. This code serves two purposes, to alter chromatin structure and to recruit other regulatory factors. Though there are many histone PTM, for the purpose of this study, the focus will be on lysine modifications of histone 3 (H3) which are maintained by histone "writers" and "erasers".

In regards to acetylation, histone acetyl transferases (HATs) deposit acetyl groups, whereas histone deactelyases (HDACs) remove the modification. There are two major classes of HATs, the more nuclear Type-A HATs and the predominantly cytoplasmic Type-B HATs. The Type-A HATs can be subdivided into three families based on sequence homology: MYST, GNAT, and CBP/p300[139]. The HDACs are a diverse group of proteins that can be divided into four classes, practically named Class I, II, III, and IV. Class III HDACs include the NAD+ dependent Sirtuin family[140]. It is important to note that many histone-modifying enzymes exist in large multi-protein complexes, and perturbations of one component can affect another[141,142]. There are 9 lysine residues on human H3 that have been shown to be modified, K4, K9, K14, K18, K27, K36, K56, and K79. All but K56 and K79

are located on the histone tail. In general, histone acetylation is associated with active gene transcription, however there are distinct roles for each acetyl modification depending on which residue it is present[143-148]. For example, acetylation of lysine 56 of histone 3 (H3K56) is important for reassembly of nucleosomes during replication, and H3K14ac is actually associated with gene repression when co-localized with H3K4me1 or H3K9me3[149-151]. One of the most well studied PTM is H3K27ac, which is found at active enhancers (see below)[152].

While a lysine can only contain one acetyl group, mono, di, or tri-methylation of histone lysine residues is possible. Histone methylation is accomplished by histone methyltransferases (KMTs). All but one of the KMTs contains a Su(var)3–9 Enhancer of Zeste and Trithorax (SET) domain[153,154]. The H3K79 methyltransferase DOT1L is the only KMT lacking a SET domain[155-157]. The first lysine demethylase (KMD), lysine-specific demethylase 1 (LSD1), was not discovered until 2004[158]. In the following years, many more KMDs were discovered. Unlike LSD1, the other KMDs have a Jumonji C (JmjC) domain and use Fe(II) and α-ketoglutarate as cofactors[159,160]. Like the HATs and HDACs, many KMTs and KDM exist in multi-protein complexes[161-165]. Depending on which amino acid of H3 is modified, lysine methylation can correspond with increased or decreased gene transcription. For instance, tri-methylation of lysine 27 (H3K27me3) by the KMT enhancer of Zeste protein (EZH2) and the polycomb repressive complex (PRC2) is associated with gene silencing at tissue specific sites (facultative chromatin)[164,166,167]. Like H3K27me3, tri-methylation of lysine 9 (H3K9me3) by SETDB1, SUV39H1, or SUV39H2, is a marker of facultative heterochromatin. However, H3K9me3 is also a marker of

constitutive heterochromatin, regions that are silenced across multiple cell types[153,168,169]. In contrast, H3K4me2 and H3K4me3 are enriched at promoters of genes that are actively transcribed, while H3K4me1 is associated with poised enhancers[145,170-173]. Finally, tri-methylation of H3K36, seems to be important for both transcription initiation and elongation[174].

After a lysine has been modified, the "readers" or effector proteins either initiate changes in chromatin accessibility or recruit other regulators. A number of different domains can recognize methylated or acetylated lysines, including the bromo, PHD, chromo, tudor, and PWWP domains[175-185]. Readers can recruit proteins for gene transcription, RNA processing, and chromatin remodeling[186]. For example, the transcription factor TFIID is recruited to H3K4me3 via TATA-binding polypeptide factor 3 (TAF3), leading to transcription inititation[187,188].

Bivalent Promoters

Histone modifications often work in concert to control gene expression. One such example is bivalent promoters, which paradoxically are marked by both H3K4me3 and H3K27me3. These regions were first observed in embryonic stem cells (ESC) using sequential ChIP-seq experiments. Bernstein et al. found that while the majority of ESC transcription start sites (TSS) contained H3K4me3 alone, a fraction was marked by both H3K4me3 and H3K27me3. Genes with bivalent promoters tended to be lowly expressed developmental transcription factors. Differentiation of ESC along the neural pathway showed that bivalent genes lost either H3K4me3 or H3K27me3, indicating that bivalency

represents a “poised” state where genes can be turned off or on with differentiation. This study originally concluded that because most genes that are bivalently marked in ESC are not marked as such in differentiated cells, that bivalent promoters are unique to pluripotent cells[189]. However, later studies have shown this genomic feature is present in other cell types[190-192].

Given the role of bivalent promoters in silencing transcription factors, they play an important role in development. Most evidence for this has come from studies of PRC2 deficient ESC. Disruption of the PRC2 complex alters H3K27me3 at bivalent promoters, increases expression of lineage-specific genes, and causes aberrant ESC differentiation[193-198]. In non-pluripotent cells, loss of bivalency can contribute to cancer pathogenesis. For instance, ectopic expression of the transcription factor Lim-only 1 (LMO1) is known to cause T-ALL[199]. In normal T-cells, the LMO1 promoter is bivalently marked and there is low expression of LMO1. However, in T-ALL, there is loss of H3K27me3 at the LMO1 promoter, correlating with increased LMO1 expression[200]. Thus, H3K27me3 acts as a brake at bivalent promoters, and loss of this mark can lead to increased expression of oncogenes.

Enhancers

Mounting research over the past few decades has shown the importance of non-promoter regions in regulating gene expression. Enhancers are long-range cis-regulatory elements that unlike promoters can control gene expression from great distances. Enhancers were first documented in 1981, when it was found that cis viral SV-40 genes

could increase beta-globin gene expression by 200-fold[201]. Later studies in fish and mice showed the long-range regulatory abilities of enhancers when a gene regulatory element of sonic hedgehog (Shh) was identified that controlled Shh expression from over 1 Mb away[202,203].

Current research has revealed that there are different types of enhancers that can be characterized by the histone modifications present, transcription factors bound, and production or lack thereof enhancer RNA (eRNA). Poised enhancers were first identified in ES and are marked by H3K4me1 and H3K27me3[152,204]. Both PCR2 and the acetyltransferase p300 are bound at poised enhancers[204,205]. As their name suggests, poised enhancers are not highly expressed. Thus, they seem to serve as a mechanism of priming gene transcription. If a poised enhancer is demethylated by UTX, and then acetylated by a p300, it becomes an active enhancer. Active enhancers are distinguished from poised enhancers by the presence of H3K27ac and lack of H3K27me3[152,204]. Presence of p300 is also used as a way to identify active enhancers[204,206-208]. Unlike poised enhancers, target genes of active enhancers are highly transcribed[173,209]. Recently a third type of enhancers "super-enhancers" has been characterized. Super-enhancers are large enhancer regions marked by H3K27ac that are bound by the Mediator complex and cell-type specific transcription factors. They have been identified in multiple cell types and seem to mark cell-identity genes[210,211]. However, whether they are really a distinct class of enhancers or rather just very potent active enhancers has set to be determined[212].

Given the findings that enhancers can actually encode RNA, there has been a movement to define active enhancers based on their ability to produce eRNA[206,213,214]. eRNAs are short (<2 kb) non-coding RNAs that are bi-directionally transcribed[215]. Of active enhancers defined using chromatin marks, about 25% are bound by RNA Pol II, and 17% encode RNA[215]. However, given the low-abundance of eRNA, it is possible that some transcripts are undetectable, and this frequency is actually higher[216]. Transcription of eRNA can be induced by various stimuli. It is estimated that there are 40,000-65,000 eRNA in the human genome[217]. Importantly, some eRNAs have been shown to facilitate the interaction of enhancers and their target gene promoters and be necessary for target gene transcription[218-222].

Gene regulation by enhancers is an intricate process involving multiple transcription factors and proteins that aide in mediating the enhancer and promoter interaction. The predominant theory is that enhancers activate their target genes through chromatin looping (Figure 1.5)[223,224]. If the chromatin surrounding the enhancer region is condensed, "pioneer" transcription factors can bind and recruit chromatin remodeling enzymes and other transcription factors[206,225,226]. Nucleosome repositioning allows for the combinatorial binding of other transcription factors, many of which are lineage specefic[227,228]. Co-activators, such as p300 and the Mediator complex are then required to facilitate the enhancer promoter interaction. Mediator acts as bridge between the enhancer and promoter and is capable of recruiting RNA Pol II to the promoter[229-231]. Cohesin and Ying-Yang 1 (YY1) are also necessary to form the DNA loops[231,232]. Deletion of an enhancer's eRNA can diminish the chromatin looping[220].

Enhancer deregulation in leukemia

Alterations of enhancer elements are present in both developmental disorders and cancer. Indeed, some of these disorders may be thought of as "enhanceropathies". In the hematopoietic system, translocations, amplifications, or mutations within enhancers has been shown to contribute to the pathogenesis of hematological malignancies. In T-cell ALL (T-ALL) at least two translocations that drive expression of oncogenes have been identified[233,234]. The transcription factor TAL1 (or SCL) is a crucial driver of leukemogenesis in T-ALL. The t(1;14)(p33;q11) translocation disrupts the TCR locus and results in ectopic expression of TAL1 driven by the TCR enhancer[233]. Similarly, the t(14;21)(q11.2;q22) translocation leads to overexpression of *BHLHB1* by a TCR enhancer, which possibly contributes to leukemogenesis by inhibiting E2A[234]. Enhancer translocations can also cause myeloid leukemias. In patients with inv(3)/t(3;3) AML, chromosomal rearrangement of the *GATA2* enhancer leads to the abnormal expression of the proto-oncogene *EVI1* and downregulation of *GATA2*[235,236].

Overamplification of enhancers can also drive leukemogenesis. For example, 1.47 Mb downstream of the *MYC* transcription start site is an enhancer that is activated by NOTCH1 and is required for normal thymocyte development. This enhancer region becomes amplified in 5% of T-ALL patients, leading to increased expression of *MYC*[237]. Deletion of this enhancer region prevented leukemia initiation in a Notch1 mouse model. Finally, somatic mutations can cause the formation of oncogenic neo-enhancers. In T-ALL, heterozygous somatic mutations within a non-coding region lead to the formation of a

super-enhancer. This neo-enhancer is bound by MYB, which recruits the leukemogenic transcriptional complex[238].

Cytosine Modifications

Chemical modifications of DNA provide another layer of epigenetic control. Cytosine modifications include 5-methylcytosine (mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxycytosine (5caC). The most prevalent cytosine residue is mC, which can be sequentially oxidized to generate 5hmC, 5fC, and 5caC[239,240]. While recent technological advances have allowed for the profiling of rare 5fC and 5caC, little is known about their function[241,242]. For the purpose of this study, the characteristics and functions of mC and 5hmC will be highlighted.

Since its discovery in 1950, diverse roles of mC have been observed in processes such as X-chromosome-inactivation, imprinting, tumorigenesis, and transcription regulation[243-245]. In mammalian genomes, the majority of mC is located at CpG dinucleotides, and 70-80% of CpG are methylated[246-248]. For many years, research was focused on short (<1kb) regions of CpG repeats found throughout the genome called CpG islands (CpGi)[249,250]. The majority of housekeeping genes have CpGi in their promoter regions, although only 40% of non-constitutively expressed genes do[251]. The regions flanking CpGi (+/- 2kb) are known as CpG shores and have shown to be altered in cancer[252]. Large unmethylated regions, named CpG valleys and canyons, have also been identified[253-255]. Cytosine methylation is performed by the DNA methyltransferases (DNMTs): DNMT1, DNMT2, DNMT3A, DNMT3B, and DNMT3L. S-adenosyl-L-methionine

(SAM) is used as the methyl donor. De novo methylation of unmethylated cytosines is completed by DNMT3A and DNMT3B, whereas DNMT1 preferentially binds hemimethylated DNA and is responsible for maintaining the DNA methylation pattern during replication[256-258]. The other DNMTs are not active in DNA methylation in adults, as DNMT3L mostly functions during embryogenesis, and DNMT2 actually methylates RNA[259-263]. Cytosine methylation is a stable mark; however, it can be removed through passive demethylation during DNA replication[264], or using the base excision repair (BER) pathway[248]. Specifically, the mC oxidized derivatives 5fC and 5caC are recognized by thymine DNA glycoslyase (TDG), triggering BER and replacement of the modified cytosine with an unmodified cytosine (Figure 1.6)[239,265-267].

As far as the function of mC, context is everything. DNA methylation has various roles, depending on whether it is located at promoters, gene bodies, or transposable elements. For many years, it was cannon that methylation of promoters causes gene silencing[268,269]. While this is certainly true for many genes, whole genome bisulfite sequencing approaches have shown that at the genome-wide level, differences in promoter methylation do not always strongly correlate with altered gene expression[270-274]. In contrast, DNA methylation at gene bodies positively correlates with gene expression[247,275,276]. This may be due to the relationship of H3K36me3 and mC. Gene body levels of H3K36me3 and mC positively correlate, and DNMT3A/B contain a PWWP domain that recognizes H3K36me3[277,278]. It appears that mC is required at these sites to maintain transcriptional fidelity and prevent spurious transcription within the gene body[279]. Additionally, DNA methylation is crucial for the repression of retrotransposable elements,

which account for 35% of the genome[280,281]. The methyl-CpG-binding protein 2 (MECP2) protein recognizes mC at these regions and recruits HDACs, leading to chromatin compaction and silencing via H3K9me3[282,283].

Not long after mC was first observed, 5hmC was discovered[284]. Conversion of mC to 5hmC is catalyzed by the 2-oxoglutarate (2OG) and Fe(II)-dependent Ten Eleven Translocation (TET) enzymes. There are three TET proteins, TET1, TET2, and TET3, all of which have been shown to be capable of progressive oxidation of mC to 5caC[239,240,285-287]. Even though 5hmC is an intermediate cytosine modification, studies using isotopic labeling of DNA have suggested that it is stable mark and not just a transient byproduct of DNA demethylation[288]. Additionally, expression levels of 5hmC are tissue specific, indicating that it has biological function besides DNA demethylation[289]. In hematopoiesis, there are dynamic changes in 5hmC with normal erythroid differentiation, and alteration of 5hmC via TET2 mutation causes aberrant erythroid development[290]. At least in ES cells, 5hmC is enriched at enhancers and is important for regulating enhancer activity[291]. Knockdown of Tet2 leads to loss of 5hmC at enhancers, with concomitant hypermethylation and decrease of enhancer target gene expession[292].

Alterations of cytosines in myeloid malignancies

Deregulation of the DNA methylation pathway is a common feature in age associated myeloid malignancies such as AML, MDS, and CMML. Leukemic blasts have vast alteration of DNA methylation compared to normal hematopoietic cells[293-296]. Furthermore, clinically relevant subsets of AML patients can be identified solely based on

their methylation profile[295]. Initial studies of DNA methylation utilized promoter-based assays and found hypermethylation of promoters, including those involved in WNT signaling[294-296]. However, a recent study that utilized a more genome-wide approach, found that gene neighborhood (region 2-50kb from the TSS or TES) methylation predicted AML epigenetic subtype better than promoter methylation and that there is significant hypomethylation of enhancers[297]. Thus, DNA methylation abnormalities may lead to activation of enhancer elements and their target genes in AML.

In addition to the extensive methylation alterations seen in MDS, AML, and CMML, mutations of DNMT3A are frequently found in patients with these diseases. The DNA methyltransferase is mutated in 10% and 22% of MDS and AML patients, respectively[298-300]. A heterozygous mutation of arginine 882 to histidine (R882H) composes up >50% of these mutations in AML and 30% in MDS[298,300]. In AML, the mutation is associated with poor prognsosis[301]. The R882H mutant DNMT3A acts as a dominant negative, preventing wild-type DNMT3A from homodimerizing[302,303]. Studies in mice have also shown that loss of DNMT3A leads to development of myeloid leukemias[304,305]. The high variant allele frequencies in leukemia and the fact that mutations are found even in normal aged hematopoietic cells indicate that DNMT3A mutations may occur in founding clones[100,104,306,307]. However, it should be noted that the majority (83%) of DNMT3A mutations in CHIP do not occur at R882[101] .

Mutations in enzymes necessary for DNA hydroxymethylation have also been observed in myeloid neoplasms. Over 700 mutations in TET2 in leukemia have been detected[308,309]. Mutations in TET2 have been found in MDS (19-26% of patients), CMML

(20-58% of patients), and AML (24-27% of patients)[309-313]. Like DNTM3A mutations, TET2 mutations may also originate in founder clones[311]. Though there are numerous TET2 mutations described, the majority (~87% in MDS and CMML) are predicted to significantly alter TET2 structure and function[314]. Global decreases and focal alterations of 5hmC have been detected in AML patients with TET2 mutations[294,315-318]. However, the biological consequences of reduced 5hmC in leukemia are still being elucidated. The hydroxymethylation pathway is also impaired through mutation of isocitrate dehydrogenases (IDHs), the enzymes that generate α-ketoglutarate, a necessary co-factor of TET2. Mutations in IDH1 and IDH2 are found in up to 20% of AML patients[299,319-321]. When mutated, IDH1 produces 2-hydroxyglutarate (2-HG) instead of α-ketoglutarate. The oncometabolite 2-HG inhibits the activity of TET2 enzymes[322,323]. Notably, IDH1/2 and TET2 mutations are mutually exclusive, indicating that the mutations phenocopy each other[294].

Perhaps the greatest evidence of the importance of DNA methylation and hydroxymethylation in myeloid malignancies is the therapeutic benefit of epigenetic drugs in these diseases. The DNA methyltransferase inhibitors (DNMTi), azacytidine and decitabine are used in the treatment of MDS[112,324,325]. Both therapies improve patient response, and while their mechanism of action is not fully understood, it appears that they hypomethylate retrotransposable elements, inducing interferon signaling and viral defense genes[326-328]. Additionally, this past year, the first inhibitor of mutant IDH, enasidenib, was FDA approved[329,330]. The first phase 1/2 clinical study showed that 40%

of patients with refractory/relapsed AML and an IDH2 mutation showed an overall response, making IDH inhibition a landmark therapy in AML treatment.

In all, epigenetics provides for complex multi-level regulation of the genome. Histones can be post-transcriptionally modified, leading to recruitment of transcription factors, transcription machinery, or chromatin remodeling complexes. Combinations of histone modifications, such as H3K4me3 and H3K27me3 at promoter regions fine tune gene expression. Over the past decade, research into the function of enhancer regions, long-range elements that control gene expression, has exploded. Enhancer deregulation can contribute to leukemogenesis through aberrant activation of oncogenes. In addition to histone modifications, cytosine modifications also provide another level of genetic control. Like enhancers, DNA methylation can also contribute to oncogenesis, and aberrant DNA methylation is hallmark of age-associated myeloid neoplasms.

Epigenetic alterations with aging

Histone composition changes with age

An emerging feature of aging is alteration of both the type and number of histones. In yeast, there is a decrease in protein levels of histone 2 (H2), 3 (H3) and 4 (H4) with replicative aging and reduced levels of histones are found in the short-lived yeast strains *asf1* and *rtt109.* Strikingly, overexpression of H2A/H2B but not H3 or H4, extends the lifespan of both *asf1* and wild-type cells[331]. In a later study the authors observed that there is ~50% loss of nucleosomes with yeast replicative aging, which contributes to genomic instability[332]. In human fibroblasts, there is also a decrease in the total levels of

H3 and H4 with age[333]. This reduction is caused by a ~50% reduction in the biosynthesis of H3 and H4 with age. This loss seems to be replication dependent, as fibroblasts that were forced to proliferate displayed a similar phenotype[333]. However, whether this phenomenon is universal feature of aging will need to be determined by quantifying histones in more tissues with more biological replicates.

Accompanying this change in total histone levels with age is a shift in the expression of histone variants. Canonical histone 3.1 (H3.1) is replication dependent, whereas the variant H3.3 is expressed throughout the cell cycle[334,335]. While originally identified as a histone enriched at pericentric heterochromatin, H3.3 also plays a role in gene activation, can be post-transcriptionally modified, and appears to be functionally distinct from H3.1[336,337]. A growing amount of research also shows that transcriptional alteration and mutation of H3.3 and other histone variants contributes to oncogenesis[122]. In multiple mammalian tissues H3.3 accumulates to near saturation with age[338,339]. A study of H3.3 in mouse liver, kidney, heart, and brain, found that increased levels of H3.3 with age correlated with increases in specific histone modifications[339]. How accumulation of H3.3 may alter chromatin structure and gene expression with age and whether it is also amasses in aged quiescent stem cells is still to be determined.

Balance of repressive and activating histone modifications in aging and longevity

Imbalance of histone activating and silencing marks may contribute to aging. This is primarily supported by research using C. elegans and Drosophila. In these model organisms, perturbations of histone modifiers have been shown to alter lifespan. In C.

elegans, there is a decrease in H3K27me3 with age. Knockdown of UTX-1, a H3K27me3 specific demethylase increased levels of H3K27me3 and worm lifespan in a FoxO dependent manner[340]. Studies of H3K4me3 methyltransferases and demethylases in C. elegans have suggested that maintenance of this modification is also important in aging. Deficiencies in the H3K4me3 ASH-2 methyltransferase complex components ASH-2, WDR-5, and SET-2 extended C. elegans lifespan. The H3K4me3 methyltransferase RBR-2 may also regulate lifespan in C. elegans, although there are conflicting reports about whether it extends or shortens lifespan[341,342]. Similar to the phenotype seen with UTX-1 knockdown, inactivation of E(z), the catalytic subunit of PRC2 in Drosophila, increased fly lifespan by 33%[343]. Co-mutation of the Trithorax gene trx largely diminished the increase in lifespan of the E(z) mutant[343]. Additionally, overexpression of the RBR-2 homolog *lid* increased fly lifespan[344]. Studies of the Sirtuin family of HDACs have shown the ability of HDACs to regulate lifespan and suggested that alteration of activating acetyl groups may contribute to aging[345,346].

While maintenance of activating and repressive histone modifications is clearly involved in modulating lifespan, discrepancies between the model organisms make it difficult to extrapolate how these marks are altered in human aging. For instance, while there is a global loss of silencing histone modifications with age in C. elegans, in flies there is an increase in H3K9me3 and loss of the activating mark H3K4me3[347]. Similar to flies, in *Nothobranchius furzeri* (African Killifish), another organism used to study aging, there is actually an increase in H3K27me3 with age[348]. Thus, more studies using mammalian

models of aging are needed in order to resolve how global and local levels of these modifications change with age and what the biological consequence is.

Changes in cytosine modifications with age

Alterations in DNA methylation with age were first discovered in salmon, and since then have been found in many species[349]. DNA methylation profiles can accurately predicate an individual's age, regardless of tissue type or donor sex[350]. Early studies found global hypomethylation with age in rats and mice[351,352]. However, more recent studies have not found a significant difference in total mC levels with age in either mouse or human tissues[353,354]. Additionally, in a longitudinal study where individuals were followed over time, only 30% of participants had a methylation change of >10% over time, and there was an equal amount of cases of hypomethylation and hypermethylation with age.[355] While global levels of methylation may not drastically change with age, a number of studies utilizing array-based technologies have identified focal regions of DNA hypermethylation with age[356-360]. Both in mice and humans, CpG that are located within regions bound by Polycomb group proteins (PcG) become hypermethylated with age[357,358]. Hypermethylation of bivalent promoters has also been observed in human CD4+ T-cells and monocytes[356]. In contrast, repetitive elements appear to be hypomethylated with age[361]. As most studies of methylation in aging have used promoter-based assays, little is known how DNA methylation changes with age at enhancer regions. One study using human whole blood did find that similarly to AML, regions that were hypomethylated with age were enriched at putative enhancers[362]. Overall, with aging,

there is hypermethylation of PcG and bivalent promoters, which may lead to silencing of genes, and hypomethylation of repetitive elements which may contribute to genomic instability (Figure 1.7).

While not as much is known about 5hmC and aging, numerous studies have shown that it is altered in a tissue specific manner with age. In the mouse hippocampus, there is a global increase in 5hmC with age, whereas in human T-cells and whole blood, 5hmC is decreased with age[353,354,363]. In T-cells, age-associated changes of 5hmC may be due to loss of the enzymes that catalyze mC to 5hmC, as reduction of 5hmC correlated with decreased expression *TET1* and *TET3*[363]. In contrast, TET levels are not disrupted in the mouse hippocampus with age[353]. To date, only one study has performed global profiling of 5hmC in an aging context. They found 785 hyper-differentially hydroxymethylated regions (DHMR) and 846 hypo DHMR in aged human mesenchymal stem cells (MSC)[364]. However, this study was limited to analyzing 5hmC within CpG, and the cells underwent numerous passages before being analyzed. Given the global variations of 5hmC with age, future studies will be need to determine exactly where these changes are localized and how they may contribute to aging.

Epigenetic changes with HSC aging

Currently, all studies of epigenetic aging in HSC have used murine models and focused mostly on DNA methylation. Whole genome bisulfite sequencing has shown that in murine C57/BL6 mice, there is slight, but significant global DNA hypermethylation with age[33,365]. Examination of focal differences between young and aged HSC showed that

regions that become hypomethylated with age are associated with regions of open chromatin in erythroid and myeloid cells and that regions that are hypermethylated with age are associated with genes that are active in lymphopoiesis and targets of PRC2[33]. Notably, forced proliferation of young HSCs using 5-FU resulted in a DNA methylation profile similar to that of aged HSCs, suggesting that DNA methylation changes with age are proliferation dependent[33]. Another group observed that regions that are hypomethylated with age are enriched for Ring1b, Scl, Gata1, Ldb1, and Runx1 binding sites. These transcription factors are known to contribute to HSC self-renewal. Additionally, 70% of HSC cell-identity genes were hypomethylated with age[365]. While site-specific changes of 5hmC with HSC aging have not been profiled, a slight decrease in global levels of 5hmC in aged HSCs has been noted[365].

To date, only one study has examined how histone modifications are altered with age in HSCs. Sun et al. performed ChIP-seq for H3K4me3, H3K27me3, and H3K36me3 in purified HSCs from young (4 months) and aged (24 months) C57/BL6 mice (n=1 pool for each age group)[33]. They found that with age there is a 6.3% increase in the number of H3K4me3 peaks with age. Many H3K4me3 peaks were increased in breadth with age, especially those at genes associated with HSC identity. However, the total number of promoters that had significantly different H3K4me3 with age was relatively small, with 267 promoters having an increase in H3K4me3 and 73 promoters having a decrease. Yet they did observe that higher H3K4me3 signal with age correlated with increased gene expression, and genes such as the HoxB cluster had greater H3K4me3 and expression with age. Additionally, genes that had increases in H3K4me3 with age were also

hypomethylated, suggesting a convergence of multiple epigenetic marks with age. Changes in H3K27me3 with age were modest, with 526 genes having altered H3K27me3 at their promoters. Like H3K4me3, H3K27me3 peaks were also broader in aged HSC compared to young. Bivalent promoters were also found to change with age, with 335 bivalent domains disappearing and 1,245 bivalent promoters emerging in aged HSCs. Bivalent promoters that became silenced with age were associated with cell adhesion. Together, these studies suggest that altered DNA methylation of HSC-identity and lineage-specific genes, as well as changes in H3K4me3 and H3K27me3 signal and spreading, may contribute to aging HSC phenotypes.

Histone modifications are altered in other stem cell types with age

Examination of the epigenome in other stem cells suggests that epigenetic deregulation with age may be a common feature of stem cells. Alterations in histone modifications with age have been observed in both skeletal muscle stem cells (MuSC) and mesenchymal stem cells (MSC). Like HSCs, MuSC are mostly quiescent at resting state and can be induced to proliferate upon tissue damage. With age, there is a reduction in the number of quiescent MuSC (qMuSC) and a loss of MuSC function[366]. Liu et al., performed ChIP-seq for H3K4me3 and H3K27me3 in purified murine qMuSC and found that similar to murine HSC aging, few H3K4me3 peaks change with age and there is broadening of H3K27me3 peaks in old qMuSC. In contrast to HSC, there is an increase in the total number of H3K27me3 peaks in aged qMuSC. The peaks that appeared with age were located at intergenic and promoter regions, which overall showed a 4-fold increase in

H3K27me3 in aged qMuSC. Notably, they found that promoters of genes encoding histones were marked by H3K4me3 in young qMuSC but became bivalently marked with age. This corresponded with a 40% decrease in gene expression of genes encoding H1, H3B, H3, and H4[367]. Another group also observed a global increase in H3K27me3 as well as H3K9me3 in aged qMuSC. Additionally, upon activation, aged MuSC had genome-wide increases in activating histone modifications, including at the promoter of *Hoxa9*. This upsurge had functional consequences, as overexpression of Hoxa9 in young animals resulted in an aged phenotype[366].

Although the aging MSC epigenome has not been profiled in great detail, it does appear that epigenetic alterations may contribute to loss of function with age. With age, there is a decrease in total levels of H3K9me3 in MSC, and MSC from a Werner progeria model also have less H3K9me3 compared to normal MSC[368]. Epigenetic alterations may also contribute to the increased adipogenesis and decreased osteogenesis that is found with aging. In MSC, HDAC9c is a key mediator of osteogenesis. Aged MSC have increased expression of EZH2 and total levels of H3K27me3. Epigenetic silencing of HDAC9c by EZH2 leads to an increase in adipocyte differentiation at the expense of the osteoblast lineage[369]. Given the specialized functions of MSC, HSC, and MuSC, the niches they inhabit, and their different aging phenotypes, it is reasonable they do not follow a universal epigenetic shift with age. Furthermore, tissue-specific aging gene signatures have been observed, supporting the idea that alterations in epigenetic marks with age would also be cell type dependent[370]. While epigenetic remodeling with age may be cell-

type specific, it does appear that epigenetic alterations may be a common mechanism contributing to stem cell loss of function with age.

In sum, epigenetic remodeling occurs with age, although the biological consequences of this are still being investigated. Perturbations of histone readers and erasers affect longevity, highlighting the importance of epigenetics in aging. Additionally, an imbalance between repressive marks, such as H3K9me3 and H3K27me3, and activating marks like H3K4me3 appears to be present with aging. However, no studies have examined if enhancers are altered with age. Alterations in DNA methylation with age may contribute to gene silencing and genomic instability. While the aged murine HSC epigenome profile has been profiled, it is unclear if it is homologous to human HSC aging.

Summary of rational and thesis aims

The world's population is aging. By 2050, over 1.5 billion people, or roughly double the current percentage of the population, will be over 65 years old[371]. This change in demographics will increase the need to address the health challenges that aging presents. At the hematopoietic level, aging is associated with an increased rate of anemia and other isolated cytopenias of unknown significance, loss of adaptive immunity, and an increased predisposition to develop myeloid malignancies such as MDS and AML[372-375]. Additionally, with aging there is increasing incidence of clonal hematopoiesis of indeterminate potential (CHIP) with frequent mutations in the epigenetic modifiers *DNMT3A*, *TET2*, and *ASXL1* as well as proteins involved in alternative splicing such as *SF3B1* and *SRSF2*.

While many mechanisms likely contribute to these alterations in the hematopoietic system with age, loss of HSC function is a major driver of hematopoietic decline. Aged HSC have increased self-renewal, decreased homing ability, and are more likely to differentiate into cells of the myeloid and erythroid rather than lymphoid lineage. These age-associated changes have been attributed to diverse mechanisms such as defective DNA damage response, altered metabolism and autophagy, and non-canonical WNT signaling. Disruption of the epigenome has emerged as a possible feature of aging. Across multiple species, changes in DNA methylation with age have been shown to affect PcG targets, bivalent promoters, and repetitive elements. While not as well studied, alterations in histone modifications have been observed with age, and epigenetic readers and erasers can modulate longevity.

Given the growing evidence that DNA enhancer elements are altered in cancer, it would be beneficial to examine if they also change with aging. Though select histone modifications and DNA methylation have been examined in murine HSC aging, the role of enhancers and 5hmC in aging has not yet been addressed. Furthermore, given the strain-to-strain variabilities in aging HSC phenotypes in mice, the discrepancy in certain aging phenotypes between mice and humans, and the poor homology between select epigenetic modifications across species, it would behoove the community to understand how the epigenome changes with human HSC aging.

Therefore, the focus of this dissertation is (a) to determine whether the chromatin structure of genomic regulatory regions is altered with human HSC aging, (b) to examine if changes in cytosine modifications with age predispose for myeloid malignancies, and

(c) to elucidate the transcriptional consequences of age-associated epigenetic alterations. We hypothesized that aging is associated with changes in the epigenetic landscape of human HSCs, impacting enhancers and other regulatory elements, which likely contributes to aged HSC loss of function. To this end, we performed a comprehensive characterization of histone and cytosine modifications in young and aged HSC collected from healthy donors. We further correlated these observations with gene expression levels in HSC. Finally, we confirmed that *in vitro* downregulation of some of the top genes epigenetically downregulated with aging results in phenotypic changes in human hematopoietic stem and progenitor cells that partially recapitulate the aging phenotype.

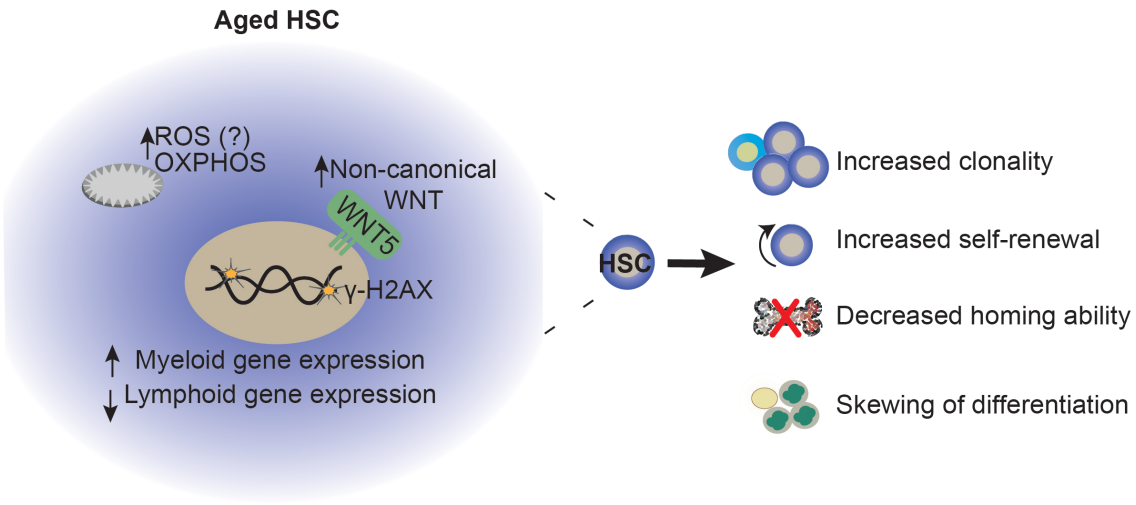


Figure 1.1: Phenotypic and functional changes with murine HSC aging. Aged HSCs have altered WNT signaling, show signs of mitochondrial damage, are more metabolically active than young HSC, and have altered gene expression of myeloid and lymphoid genes. Increased γ-H2AX with age suggests that there is also greater DNA damage in aged HSC. While there is an increase in HSC self-renewal and frequency in the bone marrow with age, these HSC are actually impaired. Aged HSC have decreased regenerative abilities when transplanted, and become skewed towards the myeloid lineages.

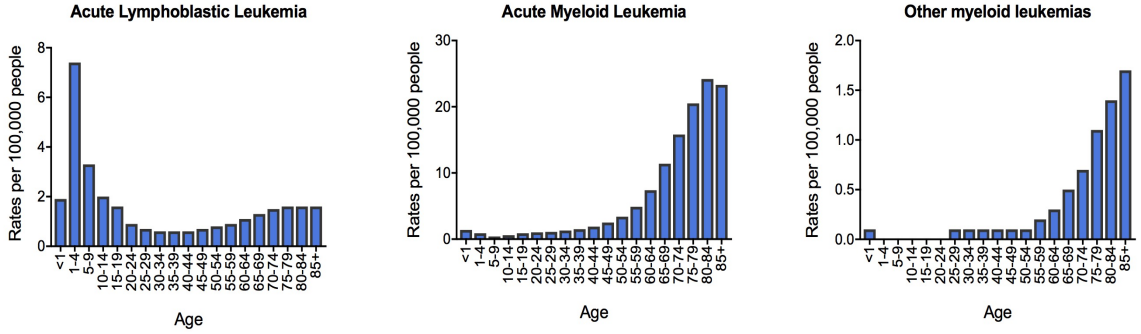


Figure 1.2: Myeloid leukemias are more frequent with age. Bar plots denoting the rates per 100,000 people of Acute lymphoblastic leukemia (ALL), Acute Myeloid Leukemia (AML) or other myeloid leukemias by age group. Data is from the United States Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute Wonder database, from 1999-2014.

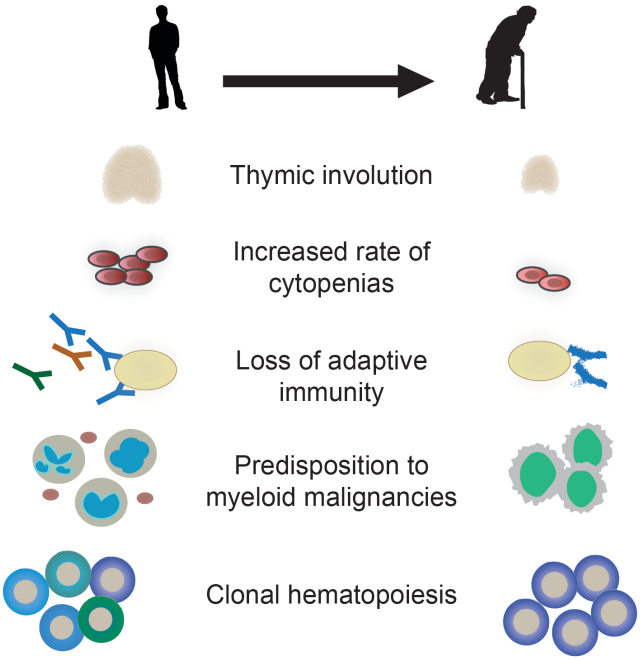


Figure 1.3: Hematopoiesis is impaired with age. With age, there is a loss of thymic mass and T-cell function, although most thymopoiesis takes place in the periphery. Older individuals have higher rates of cytopenias such as anemia, decreased ability to respond to infections, and an increased predisposition to develop myeloid malignances such as MDS and AML. Another feature of aged hematopoiesis is an increase in clonality, with epigenetic modifiers frequently being mutated.

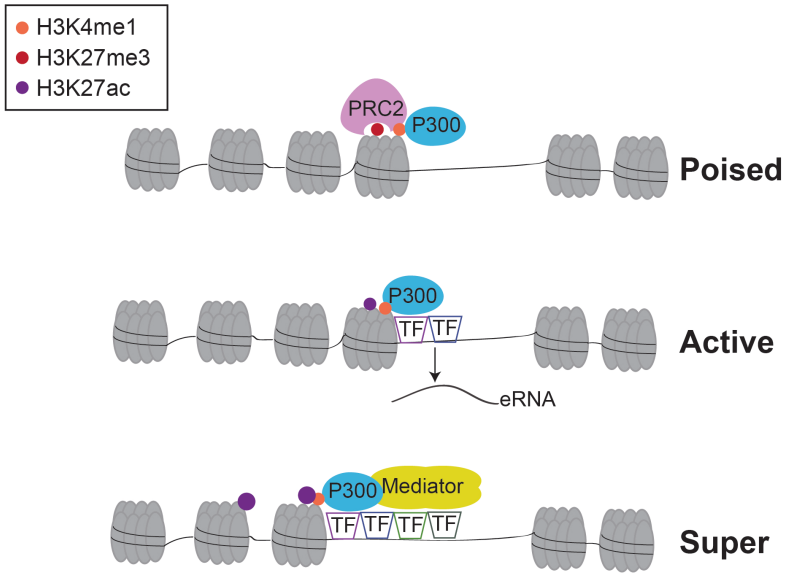


Figure 1.4: Unique combinations of histone marks and protein complexes define enhancer subtypes. Poised enhancers are primed for rapid activation of their target genes, which are silenced. They are defined by H3K4me1 and H3K27me3, and are bound by P300 as well as the repressive PRC2 complex. In contrast, active enhancers actively induce transcription of their target genes. These enhancers contain H3K27ac as well as many lineage specific transcription factors. Super enhancers are large enhancers with very high levels of H3K27ac and are bound by P300 in addition to the mediator complex.

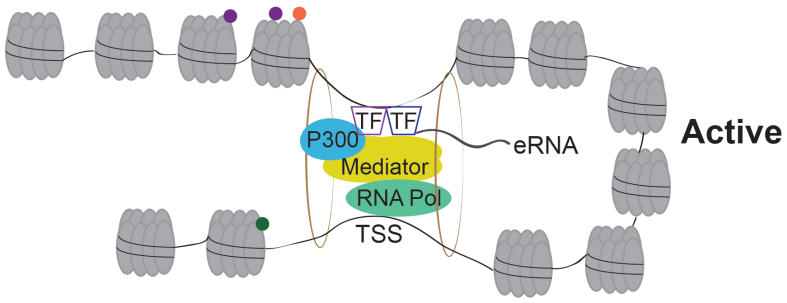


Figure 1.5: Active enhancers activate target genes through chromatin looping. Lineage specific transcription factors, P300, and Mediator facilitate the binding of an enhancer to its target promoter. Cohesion proteins loop are around the DNA, bringing the different regions together. At least in some instances, eRNA helps to recruit Mediator to the enhancer region.

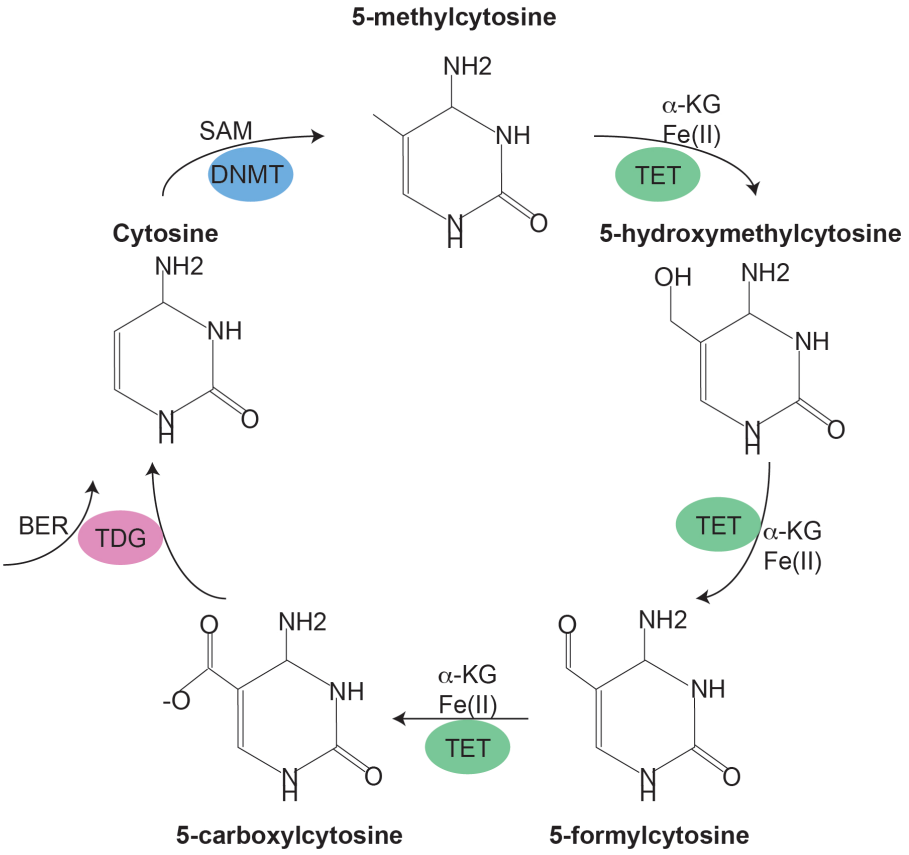


Figure 1.6: Cytosine modification pathway. Schematic of the DNA methylation and demethylation pathway. The DNMT family methylates cytosines, while the TET deoxygenases sequentially oxidize mC to 5-carboxylcytosine. Non-replication dependent DNA demethylation is achieved through the binding of TDG to 5-carboxylcytosine or 5-formylcytosine, and the subsequent removal of the cytosine base through base excision repair (BER).

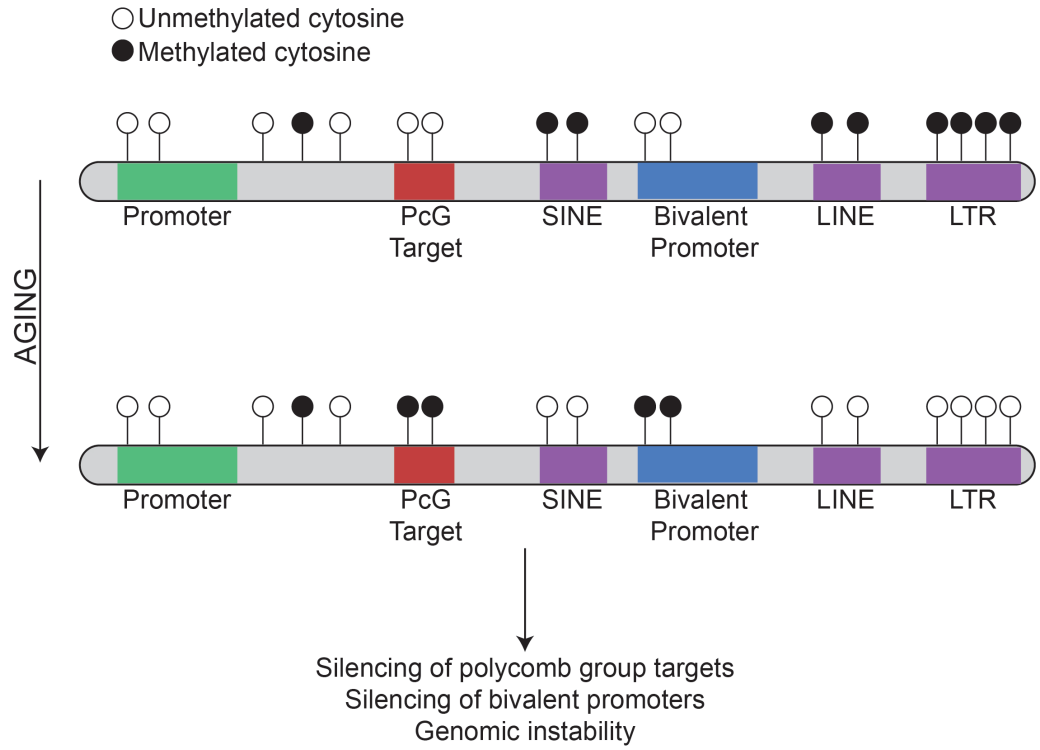


Figure 1.7: Focal methylation changes with age lead to altered gene expression and genomic instability. Whether total DNA methylation levels change with age is controversial. However, across species and tissues, focal DNA hypermethylation of polycomb group targets (PcG) and bivalent promoters has been observed. In contrast, repetitive elements such as SINE, LINE, and LTR are hypomethylated with age, possibly contributing to genomic instability.

CHAPTER 2

Materials and methods

Human bone marrow samples

All aged (65-75 yo) samples were derived from femurs from individuals undergoing hip replacement surgery at the University of Michigan hospital. Donors had no known history of hematological malignancy. Bone marrow mononuclear cells (BM MNC) were isolated using ficoll based density centrifugation and cryopreserved. Additional bone marrow mononuclear cells were purchased from Allcells, Hemacare, Stemcell Technologies, University of Pennsylvania Stem Cell and Xenograft Core, and Cincinnati Children's Hospital Processing Core for young (18-30 yo) and middle-aged (45-55 yo) donors. For scRNA-seq, bone marrow cells from young healthy donors were obtained at Cincinnati Children's Hospital Medical Center through informed consent under an approved institutional review board research protocol. Our cohort for the epigenomic and bulk transcriptome profiling was composed of a total of 29 young and 33 aged donors. Sex information was not available for all of the young donors, for the aged cohort, there were 9 males and 24 females. We did not detect any significant difference in HSCe frequency between sexes. We observed a slight influence of sex on gene expression and controlled for this in the RNA-seq analysis using DESeq.

FACS isolation of HSCe and HSPC

Miltenyi MACS magnetic bead purification was used to enrich BM MNC for CD34+ cells. The CD34+ fraction was then stained with CD2-PE-Cy5 (eBioscience, clone RPA-2.10), CD3-PE-Cy5 (eBioscience, clone UCHT1), CD4-TC PE-Cy5 (Invitrogen, clone S3.5), CD7-TC PE-Cy5 (Invitrogen, clone CD7-6B7), CD8-TC PE-Cy5 (Invitrogen, clone 3B5), CD10-PE-Cy5 (eBioscience, clone eBioCB-CALLA), CD11b-PE-Cy5 (eBioscience, clone ICRF44), CD14-TC PE-Cy5 (Life Technologies, clone TuK4), CD19-PE-Cy5 (eBioscience clone HIB19), CD20-PE-Cy5 (eBioscience, clone 2H7), CD56-TC (Invitrogen, clone MEM-188), Glycophorin A-PE-Cy5 (BD, clone GA-R2) and Thy1-biotin (eBioscience, clone 5E10, followed by CD34-APC (BD, clone 8G12), CD38-PE-Cy7 (eBioscience, clone HIT2), CD123-PE (BD, clone 9F5), streptavidin-APC-Cy7 (Life Technologies), CD45RA-FITC (Invitrogen, clone MEM-56) and DAPI. HSCe (DAPI-, Lin-, CD34+, CD38-) and HSPC (DAPI-, Lin-, CD34+,CD38+) were cell sorted using either a BD FACSAria I with a 70 μm nozzle or a BD FACS SORP Aria-Ilu.

ChIP-seq

HSCe were sorted into 1mL IMDM 20% FBS. For H3K4me1, H3K4me3, and H3K27me3, 14,000-20,000 HSCe were used per immunoprecipitation; for H3K27ac, 15,000-35,000 cells were used per immunoprecipitation, and for total H3, 10,000-15,000 cells were used per immunoprecipitation. ChIP-seq was then performed using the True MicroChIP (Diagenode, #C01010130) kit and antibodies that had been validated for specificity and reactivity using the MODified Histone Peptide Array (Active Motif, #13001). The manufacturer's protocol was followed using the following modifications. After quenching

with glycine and washing with PBS, samples were suspended in 100 μL undiluted Lysis buffer with 1x Diagenode protease inhibitor cocktail and 5 mM sodium butyrate per 10,000 cells. Samples used for H3 were sonicated in 1.5 mL TPX tubes in a Bioruptor Pico for 6 cycles of 30 seconds on and 30 seconds off. All other samples were sonicated in a Bioruptor XL for 55 cycles of 30 seconds on and 30 seconds off. Chromatin was immunoprecipitated for 12 hr at 4° C using 1μg H3K27me3 (Millipore 07-449, lot #21494165), 1μg H3K4me3 (Abcam ab8580, lot #GR164207-1), 0.5 μg H3K4me1 (Diagenode C15410194, lot #A1862D), 0.5 μg H3K27ac (Abcam ab4729, lot #GR155970-2), or 0.5 μg H3 (Abcam ab10799, lot #GR275925-1). After reverse crosslinking, DNA was purified using the minElute PCR Purification kit (Qiagen, #28004) and eluted in 16 μL of Tris-HCl ph 8.0. Enrichment was verified using QPCR with the primers GAGAGTCCTGGTCTTTGTCA and ACAGTGCCTAGGAAGGGTAT for H3K4me1 and H3, AGGGAGGGAATTAATCTGAG and ACAGTGCCTAGGAAGGGTAT for H3K4me3 and H3, TACTTGGTTTCTGCATCCTT and TCACTAAAGAAACCGTTCGT for H3K27me3 and H3, and GAGCAGAGGTGGGAGTTAG and TCAGACCCTTTCCTCACC for H3K27ac. The remaining DNA was then used for library preparation with the V1 MicroPlex Library Preparation kit (Diagenode, #C05010011). For the PCR amplification, a total of 16 amplification cycles was used. Libraries were purified using a 1:1 Ampure bead cleanup and eluted in 16 uL of Tris-HCl ph 8.0. Fold enrichment over input was then verified using QPCR. Multiplexed libraries were sequenced either on an Illumina NextSeq 500 or a HiSeq-2500 sequencer.

ChIP-seq alignment

FastQC was used to evaluate library quality [376].. Libraries with >90% sequence duplication rate were discarded from analysis. Illumina adapters were trimmed using Cutadapt (version 1.12)[377]. Reads were aligned to hg19 using Bowtie (version 2.2.6)[378]. All samples had at least 28,000,000 aligned reads. Only reads mapping to unique locations were retained for downstream analysis.

ChIP-seq peak calling

Peak calling for individual IPs was performed using the callpeak function from the Model Based Analysis of ChIP-seq 2 (MACS2 v.2.0.10.20131216) program[379]. For visual representation and differential peak calling, aligned reads for each mark for each age group were pooled using samtools v1.3.1[380]. Peaks were re-called using the pooled IP and the corresponding pooled Input with the nomodel option. For H3K4me1, H3K4me3, and H3K27ac, a q-value cutoff of 0.0001 with the narrow peak option was used. Broad peak calling with a q-value cutoff (--broad-cutoff) of 0.1 was used for H3K27me3. The IP and Input bedgraph files produced by macs2 peak calling were used for differential peak calling with the macs2 bdgdiff function, which adjusts for sequencing depth. To determine the fold change for significant ($\log_{10}$ likelihood ratio>3) differential peaks that were reduced with age, the fold-enrichment of (Pooled IP/Pooled Input) was calculated for each histone modification and age group and normalized to reads per million. The deepTools2 multiBigwigSummary function was then used to determine the number of counts for each differential peak, for each pooled sample. The fold change of Aged/Young

was then calculated with gtools[381]. Significant peaks for each age group and differential peaks (FDR<0.05, Log10 likelihood ratio>3, respectively) were annotated to the nearest TSS using the R package ChIP-enrich (v1.8.0)[382]. For the permutation analysis, random pools composed of young and aged samples were generated using R for H3K4me1, H3K4me3, H3K27ac, and H3K27me3, with n=5, 6, 4, and 6 samples per pool, respectively. Peak calling and differential peak calling were performed as described above (n=100 pair-wise comparisons per modification, respectively). The FDR was calculated for each histone modification using the significance analysis of microarrays (SAM) method[383]. Briefly, for each modification, the median number of differential peaks across all permutations was divided by the total number of peaks identified in young and aged HSCe by macs2 bdgdiff.

Enhanced reduced representation bisulfite sequencing

FACS isolated HSCe (n=5-7 per age group) were sorted into Qiagen RLT+. DNA was extracted using the Allprep DNA/RNA Micro Kit (Qiagen, #80204) and 2-10 ng of DNA per sample was used for library preparation. Enhanced reduced-representation bisulfite sequencing (ERRBS) was performed as previously described, with the following modifications to accommodate the low input amount of DNA[384]. Prior to adapter ligation, the methylated adapters were diluted to 150 nM, and for the gel size selection, fragments of 150-450 bp were excised. Libraries were sequenced on an Illumina HiSeq 2500.

Differentially methylated regions analysis

Reads were aligned against a bisulfite-converted human genome (hg19) using Bowtie and Bismark (versions 2.2.6 and 0.14.5, respectively)[385]. After filtering and normalizing by coverage, a methylDiff object containing the methylation differences and locations for cytosines that were present in at least 3 samples per age group (meth.min=3) were identified using MethylKit (version 0.9.4)[386] and R statistical software (version 3.2.1)[387]. Differentially methylated cytosines (DMC) were then inputted into the R package eDMR (v.0.6.3.1)[388], which was used to identify significant DMR (regions with at least 3 CpG, 2 of which are DMC, with absolute mean methylation difference ≥ 20, and DMR q-value ≤ 0.05). DMR were annotated to hg19 using the following parameters: DMRs overlapping a gene were annotated to that gene, intergenic DMRs were annotated to all neighboring genes within 50 kb, and if no gene was present within a 50-kb window, DMRs were annotated to the nearest transcription start site[389]. Functional annotation was performed using DAVID version 6.7[390,391]. For the heatmap, the percent methylation for each DMR was calculated for each sample. The row-scaled values were than plotted with heatmap.2, using the complete clustering method with the Euclidian distance, and using lightgrey for NA values.

Comparison of aging DMR to hematopoietic DMR

To detect if aging DMR were associated with hematopoietic identity genes, the percent methylation for each hematopoietic identity DMR defined by Farlik et al.[271] was calculated for each young and aged HSCe sample as described above. The percent methylation for

the remaining cell types was obtained from Farlik et al.'s published results. Principal component analysis and correspondence analysis were performed with mean centered and NA excluded data using made4[392] and plotted with ggplot2[393]. To assess if aging DMR were associated with regions that are differentially methylated between myeloid and lymphoid cells, the same analysis was performed as described above, using the myeloid vs. lymphoid DMR described by Farlik et al.

k-means clustering with AML blasts

The percent methylation for each age-associated hyper or hypo DMR was calculated as described above for each young and HSCe donor and AML patient[394]. NA values were removed prior to clustering. The optimal number of k-means clusters was calculated using the gap statistic method in factoextra[395]. K-means clustering was performed using R, and the heatmap was plotted using ComplexHeatmap[396], splitting between each cluster, and using lightgrey for NA values. Boxplots were plotted using ggplot2[393].

hmeDIP-seq

HSPC DNA (250-1000ng) was isolated as described above for each sample (n=5-7 per age group) and used for Hydroxymethyl-DNA immunoprecipitation with sequencing (hmeDIP-seq)[291]. DNA was sheared to 200bp using the Covaris and then purified using the Qiagen MinElute PCR purification kit. The eluate was end repaired using T4 DNA Polymerase (NEB Cat # M0203L), Klenow DNA Polymerase (NEB, #M0210L), and T4 PNK (NEB Cat # M0201L) for 30 minutes at 20° C. Samples were cleaned using the PCR Purification kit (Qiagen,

#28104) and A-Tailing was performed using Klenow exo 3-5' minus (NEB, #M0212L) at 37° C for 30 minutes. Custom made Illumina adapters were then ligated to DNA at 16° C overnight using T4 Ligase (NEB, #M0202M). After ligation, samples were cleaned using the MinElute PCR purification kit and 20% of each sample was reserved as Input. Remaining DNA was denatured at 99° C for 10 minutes, and then incubated with IP buffer (10 mM Sodium Phosphate pH 7.0, 140 mM NaCL, 0.05% Triton X-100) and 4μg/mL 5hmC antibody (Active Motif #39791, lot #15413007) for 12 hr at 4° C. To recover immunoprecipitated DNA, 20 μL of Dynabeads Protein G (Thermofisher, #1003D) were added to each IP and incubated at 4° C for 2 hr with rocking. Both IP and Input libraries were then digested with Proteinase K for 2 hr at 55° C with shaking. Magnetic separation was then used to isolate DNA, and the supernatant was purified using Ampure cleanup. From each IP and Input, 5 μL of DNA was reserved to test enrichment of 5hmC with QPCR for two targets. The primer sets ACCACATGAAAGGCCAGAAC and TGGTTCAAGAGGTGCTTTGC and TGTCGACATCAGACATGGTGAG and TTTGGGAAACAGGCTTCGAG were used. Remaining DNA was PCR amplified for 12 cycles (98°C for 20sec, 55° C for 30sec, 72° C for 30sec) using Kapa HiFi HotStart ReadyMix (Kapa Biosystems, #KK2601). The PCR product was cleaned using Ampure beads and eluted in Tris-HCl pH 8.0. Fold enrichment was verified using QPCR using the above primers. Libraries were duplexed and sequenced using a HiSeq-2500.

hmeDIP-seq peak calling

Illumina adapters were trimmed using Cutadapt (version 1.12)[377]. Reads were aligned to hg19 using Bowtie (version 2.2.6)[378]. Uniquely mapping reads were extracted using samtools and used for peak calling. Peak calling and differential peak calling were performed using the Peak Prioritization Pipeline (PePr, v1.1.7)[397] specifying narrow peaks and discarding duplicate reads. Aligned files for each 5hmC IP and its corresponding input were inputted separately into PePr in order to main inter-sample variability. Significant differentially hydroxymethylated regions (Benjamini-Hochberg FDR < 0.05 and a fold change > 1) were annotated as described above for DMRs. For comparison to hypo-DHMR in *TET2* mutated AML, raw hme-SEAL data[317] for *TET2* mutated AML and wild-type *TET2* AML (AML-ETO) was downloaded from GEO and processed as described above for hmeDIP-seq. Differential peak calling between *TET2* mutant and wild-type was performed as described above. Overlap between *TET2* hypo-DHMR and aging DHMR was done using bedtools intersect. Significance was calculated using Fisher's exact test.

RNA-seq

FACS isolated HSCe were sorted into Qiagen RLT+ (n=7-10 biological replicates). RNA was immediately isolated using the Allprep DNA/RNA Micro Kit (Qiagen, 80204) using the manufacturer's protocol to extract total RNA. Samples with a RIN>8.5 were used for library preparation. Ribosomal RNA was removed using RiboGone (Clontech, #634846). Stranded libraries were prepared by the University of Michigan Sequencing Core using

the SMARTer Stranded RNA-seq Kit (Clontech, #634836). Libraries were sequenced on the HiSeq-2500 with 50 bp paired-end sequencing.

RNA-seq alignment

Using Cutadapt (version 1.12), all reads were trimmed to 48 basepairs and adapters were removed[377]. Reads were aligned to the hg19 gencode v19 reference genome using the STAR aligner (version 2.5.2b), specifying the following parameters: outFilterType= BySJout, outFilterMultimapNmax=20, alignSJoverhangMin=8, alignSJDBoverhangMin=1, outFilterMismatchNmax=999, alignIntronMin=20, alignIntronMax=1000000, alignMatesGapMax=1000000, alignEndsType=EndToEnd[398].

Differential gene expression analysis

Gene counts were calculated using QoRTs (v1.0.7)[399]. QoRTs was run in second stranded mode using the hg19 gencode annotation file without entries for ribosomal RNA. Differential gene expression analysis was performed using DESeq2 v1.10.1[400]. A multifactor design was used in order to control for sex of the donor as well as any batch effect during library preparation. Dispersions were calculated using samples from both age groups and then contrasts were established for pair-wise comparisons. Significant genes were defined as having a fold change >1.5 and p-adjusted <0.05. Regularized log-counts (rld) were generated with DESeq2 and then row z-scores were calculated and used to plot heatmaps using the R package ComplexHeatmap (v1.15.1)[396] with average clustering and correlation distances.

GSEA

Gene set enrichment analysis (GSEA) was performed using a list of genes pre-ranked by the Wald statistic (stat column from DESeq2 output)[401]. A weighted enrichment score was used and gene set size was limited to 15-500 genes. To test enrichment for the Crews et al. aging signature, the published list of genes upregulated in aged HSCe (FPKM > 1, $p < 0.05$, L2FC > 1) was used as a gene set in GSEA[402].

Alternative splicing analysis

Replicate multivariate analysis of transcript splicing (rMATS) was used to examine alternative splicing[403]. STAR aligned reads were used as input with the transcripts from the hg19 gencode v19 genome as reference. An initial splicing cutoff of 0.01% was used, and data was then filtered to extract significant alternatively spliced events (FDR <0.05 and abs(Inclusion Rate Difference) >0.1). For comparison to skipped exon events that occur with inhibition of the spliceosome in CD34+ cells[404], R was used to identify the exons that were both significantly altered with age and with sudemycin D6 treatment.

Cell isolation for scRNA-seq

Mononuclear cells were isolated from 3 female donors aged 28-37 years old and a 71-year-old male donor with Ficoll-Paque (GE Healthcare), then enriched using CD34 Microbeads on a Automacs Pro separator (Miltenyi, San Diego, CA). CD34+ cells were stained with lineage antibodies conjugated to PerCP-Cy™5.5: CD2 (Becton, Dickinson and Company, clone RPA-2.10), CD3 (Becton, Dickinson and Company, clone UCHT1), CD4

(Becton, Dickinson and Company, clone RPA-T4), CD7 (Becton, Dickinson and Company, clone M-T701), CD8 (Becton, Dickinson and Company, clone RPA-T8), CD10 (Becton, Dickinson and Company, clone HI10a), CD11b (Biolegend, clone ICRF44), CD14 (Becton, Dickinson and Company, clone M5E2), CD19 (eBioscience, clone HIB19), CD20 (eBioscience, clone 2H7), CD24 (Biolegend, clone ML5), CD56 (Becton, Dickinson and Company, clone B159), CD66b (Becton, Dickinson and Company, clone G10F5), Glycophorin A (Biolegend, clone HIR2) and CD90-biotin (eBioscience, clone 5E10). To isolate HSCe, lineage stained cells were stained with: Streptavidin APC-Cy7 (Becton, Dickinson and Company), CD123-PE (eBioscience, clone 6H6), CD34-APC (eBioscience, clone 4H11), CD38-PE-Cy7 (eBioscience, clone HIT2), CD45RA-FITC (Invitrogen, clone MEM-56). Cell sorting was performed on a BD FACSAria II with a 100 μm nozzle.

sc-RNA-seq

To ensure maximum cell integrity, bone marrow was aspirated in the morning, cells were sorted at noon and loaded on the microfluidics chamber at 2PM. Single cell HSCe were prepared using the C1™ Single-Cell Auto Prep System (Fluidigm, San Francisco, CA), according to the manufacturer's instructions. In short, 13,000-20,000 flow-sorted cells were counted and resuspended at a concentration of 350 cells per μL PBS then loaded onto a primed C1 Single-Cell Auto Prep Integrated Fluidic Chip for mRNA-seq (5-10 μm). After the fluidic step, cell separation was visually scored, between 55-86 single cells were normally captured. Cells were lysed on chip, reverse transcription and PCR was performed using Clontech SMARTer® Ultra® Low RNA Kit for the Fluidigm® C1™ System using the

mRNA-seq: RT + Amp (1771x) according to the manual. After the completion of the mRNA-seq: RT + Amp (1771x) program, cDNAs were transferred to a 96 well plate and diluted with 3-6 μL C1™ DNA Dilution Reagent. cDNAs were quantified using Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies, Grand Island, NY) and Agilent High Sensitivity DNA Kit (Agilent Technologies (Santa Clara, CA). Libraries were prepared using Nextera XT DNA Library Preparation Kit (Illumina Inc, Santa Clara, CA) on cDNAs with an initial concentration>180 pg/μl that were then diluted to 100 pg/μL. In each single-cell library preparation, a total of 125pg cDNA was tagmented at 55 °C for 20 minutes. Libraries were pooled and purified on AMPure® bead-based magnetic separation before a final quality control using Qubit® dsDNA HS Assay Kit (Life Technologies, Grand Island, NY) and Agilent High Sensitivity DNA Kit. We required the majority of cDNA fragments to be between 375-425bp to qualify for sequencing. Single cell libraries were subjected to paired-end 75bp RNA-sequencing on a HiSeq 2500 (Illumina Inc., San Diego, CA). 96 scRNA-seq libraries were sequenced per HiSeq 2500 gel (~300 million bp/gel).

sc-RNA-seq analysis

sc-RNA-seq data (n=208 cells from 4 donors) was aligned to the human transcriptome (hg19) and quantified as transcripts per million (TPM) using RSEM[405]. On average, 6,000 transcripts were captured per cell. Associated TPM were supplied to the software AltAnalyze for dimensionality reduction and supervised analysis[406]. As part of the first step of AltAnalyze, iterative clustering and guide gene selection (ICGS) was performed to discard genes that had been covered in <10% of cells. Out of the 502 genes in the bulk

aging gene signature, 392 were covered in the sc-RNA-seq cohort and were used for singular vector decomposition (SVD). For the heatmap, $\log_2$ median normalized TPM were plotted for the genes that were differentially expressed at the bulk and single-cell level between young and aged donors (n=90 genes, empirical Bayes moderated t-test $p<0.05$). Each column represents 1 cell, and each row is for 1 gene, with Ward clustering with cosine similarity.

Enhancer analysis

To identify differential and unchanged poised enhancers ((H3K4me1>H3K4me3), H3K27me3+, H3K27ac-, and not within the promoter region), H3K4me1 peak calling of pooled samples for each age group was performed as described above. Bedtools intersect[407] was then used to discard any H3K4me1 tags from the macs2 treatment pileup where H3K4me1 fold enrichment/H3K4me3 fold enrichment <2. The resulting bedgraph files were used for differential peak calling with macs2 bdgdiff. Genomic annotation of the outputted peaks was performed using the R-package Genomation (v1.2.2)[408], and peaks within 3 kb of a TSS were removed from further analysis. Peaks overlapping H3K27me3, but not H3K27ac peaks were found with bedtools intersect. Active enhancers were defined as H3K27ac+, (H3K4me1 > H3K4me3), and not within the promoter region. To identify differential and unchanged active enhancers, differential peak calling of H3K27ac was performed as described above. Genomic annotation of the resulting peak files was then performed, and peaks within 3kb of a TSS were discarded. Bedtools intersect was used to find sites that overlapped with regions where H3K4me1 fold

enrichment/H3K4me3 fold enrichment <2. Active and poised enhancers were annotated to the nearest TSS using the R package ChIP-enrich[382].

RNA and ChIP-seq comparison

For the genes annotated to differential peaks, unchanged peaks, or active enhancers, the gene expression $\log_2$ fold-change (Old/Young) was extracted from the DESeq2 results object. Unchanged genes were defined as genes with common peaks, but no differential peaks between young and aged HSCe.

Western blot

Western blotting was performed on 5,000 HSCe per donor (n=3 per age group) and 10,000 HSCe per donor with 1-2 technical replicates for an additional 5 young and 4 aged donors. HSCe from a 53 yo donor were used to determine the linear range of the assay. Protein was extracted using a modified trichloroacetic acid (TCA; MP Biomedicals #196057) precipitation protocol[409]. Briefly, HSCe were sorted into 10% TCA and precipitated at 4° C. Extracts were spun at 13,000 RPM for 10 minutes at 4° C. Supernatant was discarded and samples were washed twice with acetone. Pellets were resuspended in TCA lysis buffer (9M Urea, 2% Triton x-100, 1% DTT) and lithium dodecyl sulphate (Life Technologies NuPage, #NP0007). For samples with 5,000 HSCe, 10 mM Tris-HCl pH 8.0 was also added. Samples were denatured at 70° C for 10 minutes. Protein from samples with 10,000 HSCe were loaded into a 4-12% Bis-Tris gel (Thermo, #NP0322BOX), 20 μL/well and then transferred to a polyvinylidene difluoride (PVDF) membrane. Blots were

imaged on a BioRad ChemiDoc. Densitometry was performed using the BioRad Image Lab software. For samples with 5,000 HSCe, and the standard curve, protein was loaded into a 15% SDS-PAGE gel, 13 μL/well and then transferred to a Immobilon FL polyvinylidene difluoride (PVDF) membrane (Millipore, #IPFL10100). Blot was imaged on a LiCor Odyssey CLx. Densitometry was performed using the LiCor Image Studio software. For all samples, intensity of H3 was normalized to that of a loading control, GAPDH. P-value was calculated using the parametric unpaired t-test was performed using Prism software. The following antibodies were used: Rabbit αGAPDH (Santa Cruz, #SC25778, lot K0615), Mouse αH3 (Abcam, #ab10799), αMouse IgG, HRP linked (Cell Signaling, #7076S, lot 32), and αRabbit IgG, HRP linked (Cell Signaling, #7074S, lot 26), αRabbit IR-Dye 800 CW (LiCor, #92632213 lot C61012-02), and αMouse IR-Dye 680 RD (LiCor, #925-68070, lot C70613-11). Power analysis was performed using the Cohen method using the pwr.t.test function within the 'pwr' R-package[410]. Effect size was calculated using the mean of the young donors (n=8) minus the mean of the aged donors (n=7) divided by the population standard deviation.

Transcription factor binding motif analysis

The Homer findMotifs function was used to detect significant (q-value < 0.05) enrichment of transcription factor binding motifs in 5hmC peaks gained with age[225]. For comparison of HSCe ChIP-seq to transcription factor binding sites in CD34+ cells, bedtools intersect was used to find the overlap between published transcription factor ChIP-seq peaks in CD34+ cells and H3K4me1, H3K4me3, H3K27me3, H3K27ac peaks, and active enhancers lost with HSCe aging[411]. Overlap between transcription factor peaks and peaks unchanged

with age was calculated to determine significance using Fisher's exact test. FDR < 0.002 was used as the significance cutoff.

Peak visualization

Peaks were visualized using the UCSC genome browser. The macs2 SPMR and bdgcmp functions were used to generate fold enrichment ChIP-seq tracks that are normalized by read count and to the IP's corresponding Input. Destranded RNA-seq tracks of the individual and pooled replicates were created using the STAR aligner parameters: --outWigType bedGraph --outWigNorm RPM –outWigStrand Unstranded.

Gene ontology analysis

Functional annotation of ChIP-seq peaks and enhancers was performed using the R package ChIP-enrich[382]. Regions were annotated to hg19 Gene Ontology Biological Processes data sets of less than 500 genes, using the method=chipenrich function. ChIP-seq peaks and enhancers were annotated to the nearest TSS. Results with an FDR<0.05 were considered significant.

Heatmaps and density plots of ChIP-seq and hmeDIP-seq

Heatmaps of differential ChIP-seq peaks, enhancers, and bivalent promoters were created using deepTools2 (v2.5.0.1)[412]. Individual and pooled ChIP-seq IP replicates were normalized to their corresponding inputs using the deepTools2 bamCompare function and the options –ratio log2 –outFileFormat bigwig –scaleFactorsMethod readCount[412].

The resulting bigwig files were then used in conjunction with deepTools2 computeMatrix and plotHeatmap to calculate and plot enrichment at regions of interest. Heatmaps are peak centered on the differential regions. Density plots of 5hmC signal were created using ngsPlot[413]. For ease of visualization, alignments from all replicates were pooled for each IP and Input and used for plotting. IP signal was normalized to Input using ngsPlot, and plotted at the coordinates of the peak centered differential peaks, all active enhancers identified in young HSCe, and all annotated gene bodies.

Boxplots of ChIP-seq data

For boxplots of histone modifications lost with age, pooled ChIP-seq IP replicates were $\log_2$ normalized as described above. Enrichment of each IP was then calculated for H3K4me1, H3K4me3, H3K27me3, and H3K27ac peaks lost with age using the deepTools2 computeMatrix function. For boxplots of genome wide H3 enrichment, individual replicates (n=2 young and n=2 aged) were $\log_2$ normalized to their corresponding inputs as described above. The deepTools2 multiBigwigSummary bins function was then used to calculate H3 enrichment over 150bp bins covering the whole genome. Boxplots were then plotted using base R.

Genomic annotation

The R package genomation was used to annotate ChIP and 5hmC peaks to promoters (+/- 1000 bp of TSS), introns, exons, and intergenic regions[408].

Bivalent promoter analysis

The bedtools intersect function was used to identify promoter regions (+/- 1000 bp of TSS) that were bound by both H3K4me3 and H3K27me3 for each age group. The same function was also used to compare the chromosomal coordinates of the bivalent promoters in each age group to identify bivalent promoters that change with age. Bivalent promoters were annotated to the nearest TSS using ChIP-enrich[408].

Targeted genomic sequencing

Genomic DNA was used to prepare libraries for hybrid capture per manufacturer protocol (Agilent). Libraries were quantified and pooled up to 24 samples per reaction in equimolar amounts totaling 500 mg of DNA. Agilent Custom SureSelect In Solution Hybrid Capture RNA baits were used to hybridize libraries, targeting 443 kbp of exonic DNA with 16890 probes as described in Lindsley et al[414]. Each capture reaction was washed, amplified, and sequenced on two lanes of an Illumina HiSeq 2000 100bp paired end run. Subsequent analysis of the target region was restricted to the regions corresponding to the following genes: *ANKRD26*, *CEBPA*, *ETV6*, *KIT*, *PRPF40B*, *SF3A1*, *TERC*, *ASXL1*, *CREBBP*, *EZH2*, *KRAS*, *PRPF8*, *SF3B1*, *TERT*, *ATRX*, *CSF1R*, *FLT3*, *LUC7L2*, *PTEN*, *SH2B3*, *TET2*, *BCOR*, *CSF3R*, *GATA2*, *MPL*, *PTPN11*, *SMC1A*, *TP53*, *BCORL1*, *CSNK1A1*, *GNAS*, *NF1*, *RAD21*, *SMC3*, *U2AF1*, *BRAF*, *CTCF*, *GNB1*, *NPM1*, *RAD51C*, *SRSF2*, *U2AF2*, *BRCC3*, *CUX1*, *IDH1*, *NRAS*, *RUNX1*, *STAG1*, *WT1*, *CALR*, *DDX41*, *IDH2*, *PHF6*, *SBDS*, *STAG2*, *ZRSR2*, *CBL*, *DNMT3A*, *JAK2*, *PIGA*, *SETBP1*, *STAT3*, *CBLB*, *EP300*, *JAK3*, *PPM1D*, *SF1*, and *STAT5B*.

Variant calling

Fastq files were aligned to the hg19 version of the human genome with The Burrows Wheeler Aligner (BWA v0.7.12) MEM module for paired end reads[415]. 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 (indels) using the reference variant databases[416]. Somatic variants were called using LoFreq v2.1.1 for all variants at >=1% variant frequency[417]. Additional variant filtering after variant calling were used with the following parameters: VF (variant frequency) <0.05, read depth at variant site <20, GQ and/or QUAL scores <30, IndelRepeatFilter >8. Variants with excessive strand bias and indels with VAF <10% adjacent to homopolymer repeats were excluded by manual curation. Filtered called variants were first annotated using ANNOVAR[418]. Variants predicted to alter splicing were assessed as described in Jian et al[419]. Variants located outside protein coding regions or splice sites of the genes listed above, and synonymous variants that were not predicted to alter splicing were filtered out. To remove common polymorphisms, variants with population frequencies of ≥1% in either 1000 genomes[420] or the Exome Aggregation Consortium (ExAC v.3.1) were similarly excluded unless they were also listed as confirmed somatic mutations in COSMIC[421]. Remaining variants were manually reviewed and considered likely somatic coding variants included in further analyses. Variant calling and interpretation were performed blinded to sample identifiers and associated phenotype information.

Lentiviral production and transduction of CD34+ cells

293FT cells were maintained according to supplier instructions (ThermoFisher Scientific). A custom pLKO.1 containing turboGFP vector was used to express shRNAs targeting LMNA (*5'-GATGATCCCTTGCTGACTTAC-3'*) and KLF6 (*5'-TTCAGCCTCAGAAATCAAATT-3'*) (Millipore Sigma). Lentiviruses were produced by co-transfection with packaging plasmids psPAX2 and pMD2.G using polyethylenimine transfection reagent (Polysciences). Lentivirus containing supernatant was collected 48-72 hours post-transfection, filtered through a 0.45-µm syringe filter, and concentrated using PEG-it virus precipitation solution (System Biosciences, #LV825A-1).

Primary human CD34+ cells were freshly isolated from mobilized peripheral blood obtained from University of Michigan Comprehensive Cancer Center Tissue Procurement Service. CD34+ cells were isolated using magnetic bead purification as described for bone marrow (Miltenyi Biotec). Isolated CD34+ cells were cultured in 20% FBS containing IMDM or SFEM II (StemCell Technologies, #09605) and pre-stimulated with recombinant human SCF (100 ng/ml), FLT3-L (10ng/ml), IL-6 (20ng/ml), and TPO (100ng/ml) (PreproTech). Lentiviral transduction of CD34+ cells was performed in the presence of 8 µg/ml polybrene (Millipore Sigma, TR-1003-G). Four days post-transduction, cells were sorted for CD34 and GFP double-positive cells on a FACS Aria Iiu (BD Biosciences). Empty vector was used as control comparison.

Colony-forming unity assay

Sorted, transduced CD34+ (n=4 and n=2 biological replicates for LMNA and KLF6, respectively) were seeded in methylcellulose, MethoCult H4435 (StemCell Technologies, #04435), in duplicate onto a 6-well SmartDish (StemCell Technologies, #27302) at varying densities of 1000 to 1,500 cells. Colonies were scored on a STEMvision after 14 days of incubation (StemCell Technologies).

Quantitative real-time PCR

RNA extraction from sorted, transduced CD34+ cells was performed using RNeasy Plus Micro kit according to manufacturer's instructions (Qiagen, #74034). cDNA was synthesized using Verso cDNA synthesis kit with random hexamer priming according to manufacturer's instructions (ThermoFisher Scientific, #AB1453). qPCR was performed in triplicate using SYBR Green on a QuantStudio-5 real-time PCR system (Applied Biosystems) for n=7 and n=2 biological replicates with *LMNA* or *KLF6* knockdown, respectively. Expression values were normalized to GAPDH. In total, 5 genes were evaluated by qPCR: *LMNA*, *KLF3*, *KLF6*, *BCL6*, and *EGR1.*

RNA-seq of LMNA knockdown

RNA from FACS isolated transduced CD34+ cells (n=4 donors) was extracted using the Qiagen RNeasy Plus Micro kit according to manufacturer's instructions (Qiagen, #74034) or Trizol. Stranded libraries were prepared by the University of Miami Sequencing Core using the Illuminia TruSeq Stranded Total RNA kit (Illumina, #20020596). Libraries were

sequenced on the NextSeq-500 with 75 bp paired-end sequencing. Data was aligned and processed as described above and used for GSEA with the lists of genes up- or down-regulated with HSCe aging as input gene sets.

Statistical analysis

Significance details can be found within the figure legends. For genome-wide sequencing assays, we corrected for multiple testing, and used $q<0.05$ as a cutoff to determine significance, for all other statistical tests, we used a p-value cutoff of $p<0.05$ to define significance. In the figures, * means $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p <0.0001$. For comparison of ChIP-seq data to gene expression data, R-statistical software (v3.2.1) was used to perform the Welch two-sided two-sample t-test. For the Western Blot analysis, a two-sided parametric unpaired t-test was used to calculate p-values with Prism software (v7.0). For the comparison of HSCe ChIP-seq peaks to transcription factor binding sites, p-values were calculated with Fisher's Exact test and corrected for multiple testing using the Bonferroni correction methods.

CHAPTER 3

Age-related decrease in histone activating marks targets key regulatory elements

Post-transcriptional modifications of histones regulate gene function. Some modifications, like H3K27me3, are associated with silenced gene promoters, whereas others, like H3K4me3, mark promoters that are actively transcribed[164,172]. Combinations of multiple histone modifications can further fine-tune gene regulation. For example, bivalent promoters contain both H3K4me3 and H3K27me3, and are associated with silenced developmental genes[189]. Presence of H3K4me1 and/or H3K27ac is a feature of enhancers, long-range gene regulatory elements that are frequently deregulated in leukemia[233-238,422]. Given that aged HSCs become skewed towards the myeloid lineage and have a decreased regenerative ability[23,25,30], mechanisms that regulate HSC loss of function with age are of particular importance. In murine HSC, deregulation of H3K4me3 and H3K27me3 may contribute to HSC decline with age[365]. However, whether these histone modifications, or those associated with enhancers are altered in human HSC aging has never been investigated. We hypothesized that age-acquired deregulation of histone modifications contributes to age-associated human HSC dysfunction. To this end, we performed chromatin immunoprecipitation followed by massively parallel sequencing

(ChIP-seq) of H3K4me1, H3K4me3, H3K27me3, and H3K27ac in young (18-30 yo) and aged (65-75 yo) HSCs, to explore how these marks change with age.

Profile of histone modifications in young HSCe

Analysis of H3K4me1, H3K4me3, H3K27me3, and H3K27ac in the HSC-enriched lineage-negative (Lin-) CD34+ and CD38- (herein HSCe) fraction from young donors (n=4-6 per histone modification) by ChIP-seq identified regions that were reproducibly epigenetically modified across multiple biological replicates (Figure 3.1A). As expected, each modification had a distinct genomic localization. Determination of the baseline distribution of H3K4me1 in young HSCe identified 123,344 peaks, localized primarily to intragenic and intronic regions, whereas for H3K27ac we found 30,833 peaks, most which were located at promoters and introns (Figure 3.1B). Functional annotation of the nearest transcription start sites (TSS) associated with H3K4me1 demonstrated enrichment for gene ontology categories in cell cycle, metabolic processes, RNA processing and kinase signaling, while H3K27ac was enriched at genes involved in chromatin structure, RNA processing, cell cycle and DNA repair (Figure 3.1A and C)**.** Analysis of the baseline distribution for H3K4me3 and H3K27me3 in young HSCe showed 26,908 and 146,116 peaks, respectively. H3K4me3 was found mainly associated with promoter regions while H3K27me3 was more enriched at intergenic regions (Figure 3.1B). Genes marked by the presence of H3K4me3 in HSCe were associated with RNA processing, protein degradation, translation and DNA damage, whereas regions with H3K27me3 were also associated with cell cycle and chromatin organization (Figure 3.1B-C)**.**

Enhancers in HSCe

Given the importance of enhancers in regulating gene activity, we next identified and characterized active enhancers (H3K4me1> H3K4me3, H3K27ac+, >3kb from TSS), poised enhancers (H3K4me1>H3K4me3, H3K27me3+, >3kb from TSS), as well as super-enhancers. We identified 13,050 active enhancers in HSCe. Not surprisingly, these enhancers were annotated to genes associated with hematopoiesis, and included hematopoietic transcription factors such as *GATA2/3*, *RUNX1/3*, *FLI1* as well as HSCe identity genes such as *CD34*. In contrast, poised enhancers (n=6,402) were associated with developmental genes, including several members of the HOX, SOX, and PAX transcription factor families (Figure 3.2A). Finally, we also sought to identify HSCe super-enhancers, large enhancer regions that drive expression of cell-type specific identity genes and are distinct from active and poised enhancers[210,211]. Using H3K27ac ChIP-seq signal and the ranking of super enhancer (ROSE) algorithm[210,423], we found 917 super-enhancers. Several of the super-enhancers with the highest H3K27ac enrichment are predicted to regulate genes involved in maintaining HSC stemness, self-renewal, or survival, including *MALAT1*, *MIR-181C*, *JARID2*, and *ETV6*[424-427] (Figure 3.2 B-C).

Bivalent Promoters in HSCe

Next, we sought to identify genes that are bivalently (H3K4me3+, H3K27me3+) regulated in HSCe. Bivalent promoters are associated with developmental genes, and loss of H3K4me3 or H3K27me3 at bivalent promoters can lead to silencing of tumor suppressor genes, or activation of oncogenes, respectively[428-430]. Of the 16,102 genes that

contained promoter (-/+ 1000 kb of TSS) H3K4me3 or H3K27me3 peaks in HSCe, 21.4% were bivalently marked (Figure 3.3 A-B). These bivalent genes were associated with developmental pathways and included several WNT pathway signaling genes such as *FZD2*, *FZD4*, *FZD8*, *WNT1*, *WNT3A*, and *WNT5A*, and *LRP5* (Figure 3.3 C-D). Genes with bivalent promoters were expressed at a level similar to genes marked by promoter H3K27me3 alone, consistent with previous reports that bivalent promoters mark silent developmental genes that are poised for rapid gene expression upon receiving differentiation signals (Figure 3.3E)[430,431].

Alteration of hematopoietic progenitor frequencies with age

As previous studies have reported variation in frequencies of human hematopoietic stem and progenitor cells in the bone marrow with aging, we sought to determine if such a phenotype was present in our cohort[37,432]. Using flow cytometry, we evaluated the frequencies of HSCe, granulocyte-monocyte progenitors (Lin-CD34+CD38+CD45.RA+CD123+; GMP), common-myeloid progenitors (Lin-CD34+CD38+CD45.RA-CD123+; CMP) and megakaryocyte-erythrocyte progenitors (Lin-CD34+CD38+CD45.RA-CD123-; MEP) in bone marrow from 41 young and 55 aged donors. As observed by others, we found an increase in HSCe frequency with age (p=1.59e-5), with no significant difference in HSCe frequency between males and females, as well as a decrease in the frequency of GMP with an increase in MEP amongst the Lin-CD34+ and CD38+ fraction ($p<0.001$)[37,432]. Of note, there were sex-dependent differences in the frequencies of MEP and CMP in young donors, and a decreased frequency of GMP in aged

males compared to females ($p<0.05$) (Figure 3.4A-D). This is especially interesting given the increased incidence of age-associated myeloid malignancies in males compared to females[433,434].

Few CHIP associated mutations occur in our cohort

Epigenetic modifiers such as *DNMT3A*, *TET2*, and *ASXL1* are frequently mutated in age related clonal hematopoiesis (ARCH)[100,101,103,104]. In order to understand what the incidence of clonal hematopoiesis was in our aging cohort and how this could impact our epigenomic studies, we performed targeted sequencing with 100X coverage of a panel of 128 genes for the donors used for genome-wide epigenetic or transcriptome profiling whenever sufficient material was available (Table 3.1). Only 1 of a total of 22 donors examined presented any mutation (variant allele frequency > 0.1), which corresponded to a *DNMT3A* mutation with variant allele frequency (VAF) of 0.12. Thus, due to the low VAF and the overall paucity of mutations, we concluded that it would be unlikely that any epigenetic or gene expression changes in the bulk HSCe population would be driven by the presence of CHIP-related mutations in this or other donors.

Reduction of activating histone modifications with age

Using a low-input ChIP-seq protocol, we sought to determine whether epigenetic profiles are reprogrammed with age. To this end, we assayed H3K4me1, H3K4me3, H3K27me3 and H3K27ac in young and aged HSCe. A direct comparison of these marks across multiple biological replicates for the two age groups (n=4-7, per mark per age

group; Table S1) revealed that aging is associated with widespread reductions in H3K4me1, H3K27ac and H3K4me3, affecting 20,786 (15% of young peaks), 15,651 (35%) peaks and 27,051 (45%) peaks, respectively (log_{10} likelihood ratio >3). Notably, relatively few regions had an increase in signal intensity of H3K4me1 (n=23 peaks), H3K4me3 (n=20 peaks), or H3K27ac (n=35 peaks) with age. By contrast, H3K27me3 displayed only 1,748 H3K27me3 peaks changed with aging, consisting of both gains (24% of differential peaks) and losses of H3K27me3 (76% of differential peaks). However, despite there being fewer H3K27me3 peaks affected with aging, these peaks displayed the greatest magnitude of change in signal intensity (Figure 3.5A-C).

In order to confirm the robustness of these observations and determine the false discovery rate (FDR) for our approach, we performed 100-fold permutation analysis of young and aged ChIP-seq data for each histone modification and determined the FDR to be 0.02, 0.14, 0.03, and 0.04 for H3K4me1, H3K4me3, H3K27me3, and H3K27ac, respectively. Importantly, the Western blot analysis of HSCe suggested that these differences are not due to a total reduction in histone 3 (H3) with age (Figure 3.6A-B). However, the limited sample size may have influenced this finding. Post-hoc power analysis (d= 0.31, n= 8 young and n= 7 aged) revealed that with a p = 0.05, there was a 9% chance of detecting an age-associated difference in H3 and the ratio of Young H3/Aged H3 at 95% confidence could be as high as 2.13 or as low as 0.35. To further test the possibility of age-related changes in total H3, we performed genome wide ChIP-seq of H3 in a separate cohort of donors. This independent analysis also did not show a reduction

of H3 either genome wide, or at sites with decreased H3K4me1, H3K4me3, H3K27me3, or H3K27ac signal with age (Figure 3.6C-D).

Genomic annotation and gene ontology analysis of these age-related epigenetic changes revealed significant enrichment (Wald test, B-H corrected p-value < 0.05) of H3K4me1 losses at promoter and intergenic regions linked to genes involved in hematopoiesis-related functions, while loss of H3K27ac, which was likewise enriched at promoter regions, was instead linked to genes associated with chromatin organization (n=303 out of 499 genes in pathway), protein metabolism, and mRNA processing and splicing (Figure 3.7 and Figure 3.8A-B). Changes in H3K4me3 and H3K27me3 were strongly enriched at promoter regions and associated with genes involved in multiple developmental processes, including regulation of cell development, cell fate determination and the WNT receptor signaling pathway (n=247 out of 282 genes in pathway). In addition, genes associated with H3K4me3 loss were also enriched for several RNA splicing and chromatin modification categories (Figure 3.7 and Figure 3.8A-B).

To determine the magnitude of these histone alterations with age and which genes were most significantly affected, we examined the fold change (FC) of Aged/Young histone modification signal for regions that had significant (log_{10} likelihood ratio >3) reductions of these marks with age (Figure S2A). As expected, not all significant peaks had a high fold change with age (Figure S2B). Pathway analysis of the top genes that were most changed with age (log_{10} likelihood ratio >3, FC <-2.0) showed that even with these more stringent gene lists, regions with reduced H3K4me3 or H3K27me3 were still associated with multiple developmental pathways, whereas peaks with decreased

H3K4me1 or H3K27ac were associated with immune cell activation and hematopoiesis (Figure S2C). Notably, amongst the most significant genes that lost H3K4me3 with age was lamin A (*LMNA*; FC= -2.8, $\log_{10}$ likelihood ratio= 25.3), which has been implicated in lineage differentiation in adult hematopoietic cells[435] and whose mutations have been linked to Hutchinson-Gilford progeria syndrome[436,437] (Figure 3.8B).

In order to identify hematopoietic transcription factors that may be involved in mediating these epigenetic changes, regions with age-related changes in histone modifications were compared to regions bound by transcription factors in human CD34+ cells[411]. Regions marked by loss of H3K4me1 were significantly enriched in transcription factors, with peaks that were lost overlapping EGR, FLI1, LMO2, LYL, RUNX1 or SCL binding sites in CD34+ cells (Bonferroni-adjusted p-value<2.5e-5 for all comparisons). Likewise, sites with reduced H3K27ac with aging were bound by EGR, FLI1, GATA2, LYL, RUNX1, or SCL in CD34+ cells (p-value<8.6e-7 for all comparisons). In contrast, regions with reduced H3K4me3 or H3K27me3 were not significantly enriched for transcription factor binding sites (Figure 3.9).

Global epigenetic alterations were compared to steady-state transcript levels in a separate cohort of donors of comparable ages. Genes with a decrease in H3K4me3, H3K4me1, and/or H3K27ac signal in our first cohort tended to be expressed at lower levels in aged HSCe from the second cohort, while the opposite was true for genes with a decrease in H3K27me3 signal. However, many genes showed little to no change in gene expression levels, indicating that for such genes the reduction in these activating histone

marks may be more reflective of an age-related change in poise status rather than a change in active transcription (Figure 3.10).

Age-related epigenomic changes target regulatory elements in the genome

In order to determine whether aging is specifically associated with epigenetic variation at gene regulatory elements, we focused the analysis on identifying specific changes at these regions. Enhancer deregulation through translocations, somatic-mutations or overamplification promotes oncogenesis, including age related myeloid leukemias[233-238,422]. Enhancers were defined as non-promoter regions marked by both (H3K4me1 > H3K4me3) and (H3K27me3 present) or, by both (H3K4me1 > H3K4me3) and (H3K27ac present) for poised and active enhancers, respectively. Notably, while only 725 out of 6,402 poised enhancers showed any age-related differences, over one third of active enhancers (n=4,519) showed significant loss of H3K27ac signal with aging (log_{10} likelihood ratio >3), with 3,700 and 2,142 of these having greater than a 1.5-fold or 2-fold change of H3K27ac signal with age, respectively (Figure 3.11). Amongst the genes associated with the 4,519 enhancers that had reduced H3K27ac were several transcription factors such as *PRDM16, RUNX3, ETV6, FLI1, GATA2, GFI1, HIF1A, IKZF1* and *KLF6* and, several epigenetic modifiers such as *BCOR, CBX7, DNMT3A, DOT1L* and *KMT2A* (Figure 3.12A). Functional annotation of all enhancers that experience an age-related loss (log_{10} likelihood ratio > 3, no FC cutoff) of H3K27ac revealed they were significantly (Wald test, B-H corrected p-value < 0.05) enriched in pathways associated with B- and T-cell

signaling and myeloid leukemia, with 25/62, 31/92, and 20/52 of the genes in the respective pathways displaying reduced H3K27ac with age (Figure 3.12B).

Additionally, active enhancers lost with age frequently overlapped with *EGR* (15%), *GATA2* (20%), *RUNX1* (32%), or *LYL* (10%) binding sites identified in CD34+ cells (p-value<1.8e-4 for all comparisons) (Figure 3.12C-D). At the RNA level, steady-state expression analysis showed that genes associated with these enhancers tended to be expressed at lower levels in aged HSCe when there was loss of H3K27ac and at higher levels in the rare instances when this mark was gained with aging (Figure 3.12E). Not all enhancers with a decrease in H3K27ac signal correlated with a decrease in transcript levels of their nearest gene. While this result may be a factor of the limitations inherent to annotating enhancers to the nearest gene, this lack of correlation may also be explained by data that shows that an enhancer's activity is best reflected by the active transcription of its associated eRNA than by the strength of the H3K27ac signal[220,221].

Given the marked difference in the localization of H3K4me3 and H3K27me3 marks with aging, as well as their association with developmental processes, we hypothesized that bivalent promoters may likewise be specifically affected during aging. Analysis of H3K4me3 and H3K27me3 profiles in young HSCe identified 3,967 bivalent promoters. Of these HSCe-specific bivalent promoters, 1,017 displayed significant loss of H3K4me3 during aging, including several *HOXC* cluster genes and *WNT* factors (Figure 3.13A-B). Pathway analysis of these bivalent promoters lost with aging showed that they were enriched in genes involved in WNT, Hedgehog, and Cadherin signaling, as well as with cancer-related pathways (Figure 3.13C). Age-related switches from bivalency to

repression due to loss of H3K4me3 were not associated with a corresponding change in expression, which is consistent with the fact that bivalent genes are not expressed to begin with (Figure 3.13B and D). By contrast, the few promoters that became aberrantly bivalent in aged HSCe due to gain in H3K27me3, did show a trend to lower expression levels (Figure 3.13D).

Summary

This chapter describes the first ever characterization of H3K4me1, H3K4me3, H3K27me3, and H3K27ac in human HSCe, and the widespread remodeling of these modifications with age. We found that each histone mark had a unique genomic localization and regulated distinct classes of genes. Analysis of bivalent promoters showed that like ESC and MuSC, HSC have bivalent promoters that are associated with mostly silenced developmental genes[189,367]. We also identified active enhancers that are likely critical for regulating hematopoiesis.

Comparative analysis of young and aged HSCe epigenomes revealed thousands of regions that had reduced signal of the activating histone modifications H3K4me1, H3K4me3, and H3K27ac with age. Notably, ChIP-seq of these marks in middle aged HSCe (45-55 yo, n=4-5 per modification) suggests that this reduction occurs gradually with age (Figure S3). Unlike what has been observed in invertebrate models and murine MuSC, we observed few regions with altered H3K27me3 with age[340,367]. However, this is in line with what has previously been observed in murine HSC[365]. A major finding in this study is that ~1/3 of active enhancers have reduced H3K27ac enrichment with age. These enhancers

putatively regulate such transcription factors as *RUNX3, FLI1, GATA2, GFI1, HIF1A,* and *KLF6,* and overall, were associated with B- and T-cell receptor signaling. This suggests that enhancer remodeling with age may contribute to impaired lymphoid differentiation and myeloid skewing. In addition, in aged HSCe there was loss of H3K4me3 at bivalent promoters associated with WNT and cadherin signaling. Thus, in losing H3K4me3, these promoters become more stably silenced, losing their potential for reactivation. Finally, analysis of somatic mutations in our cohort revealed that only 1/22 donors profiled possessed a mutation. Therefore, the epigenetic reprogramming we observe is mostly likely not due to CHIP.

Genes						
ANKRD26	CEBPA	ETV6	KIT	PRPF40B	SF3A1	TERC
ASXL1	CREBBP	EZH2	KRAS	PRPF8	SF3B1	TERT
ATRX	CSF1R	FLT3	LUC7L2	PTEN	SH2B3	TET2
BCOR	CSF3R	GATA2	MPL	PTPN11	SMC1A	TP53
BCORL1	CSNK1A1	GNAS	NF1	RAD21	SMC3	U2AF1
BRAF	CTCF	GNB1	NPM1	RAD51C	SRSF2	U2AF2
BRCC3	CUX1	IDH1	NRAS	RUNX1	STAG1	WT1
CALR	DDX41	IDH2	PHF6	SBDS	STAG2	ZRSR2
CBL	DNMT3A	JAK2	PIGA	SETBP1	STAT3	
CBLB	EP300	JAK3	PPM1D	SF1	STAT5B	

Table 3.1: Genes examined for somatic mutations.

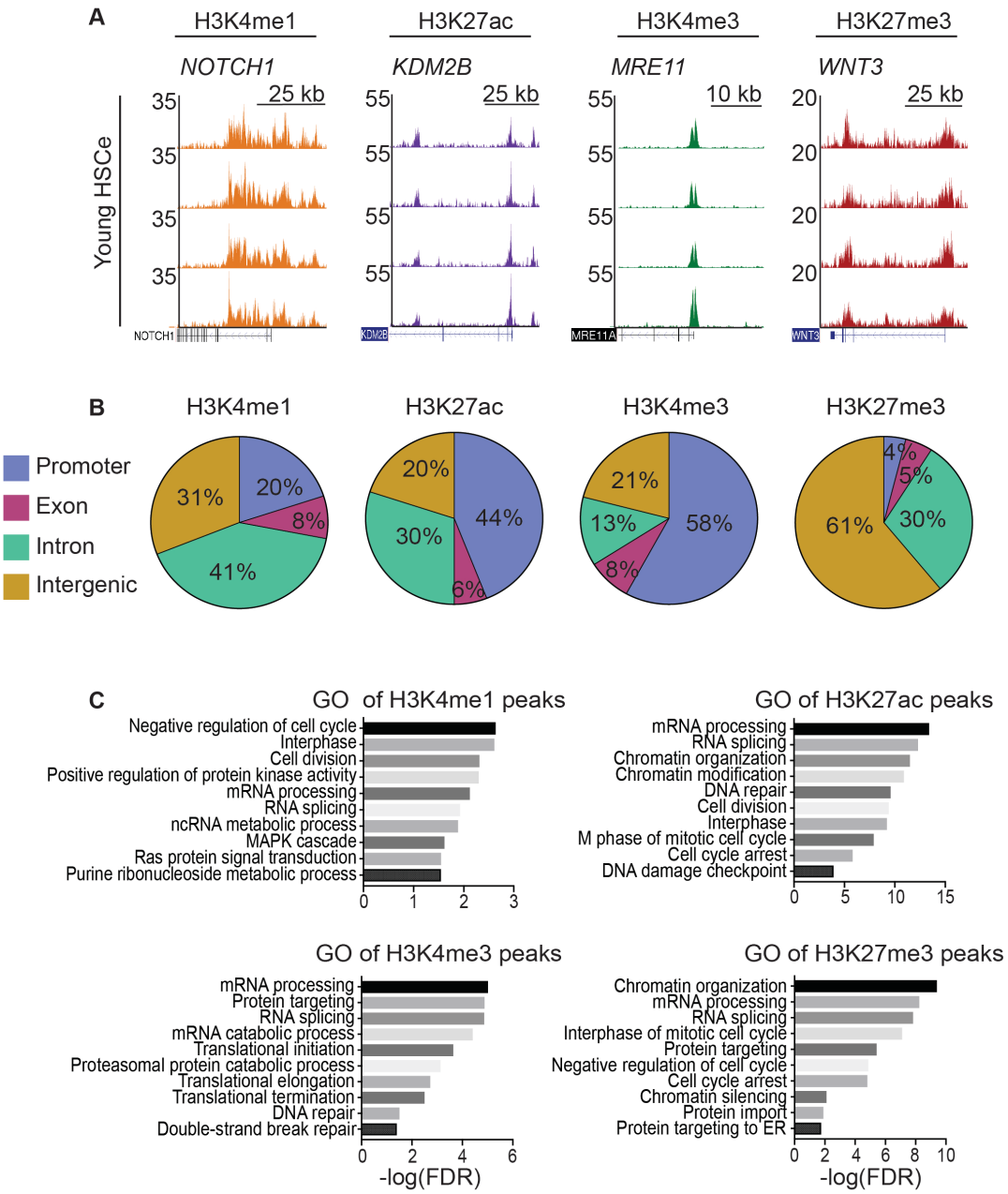


Figure 3.1: Baseline distribution of histone modifications in HSCe. (A) Representative examples of ChIP-seq peaks identified in young HSCe. UCSC tracks are normalized to Input and reads per million, each track is derived from one donor. **(B)** Genomic annotation of H3K4me1, H3K4me3, H3K27me3 and H3K27ac ChIP-seq peaks in young HSCe (q-value<1.0e-4 for H3K4me1, H3K4me3, and H3K27ac; q-value<0.01 for H3K27me3). Peaks were identified using pooled ChIP-seq replicates (n=4-6 per modification). **(C)** ChIP-enrich Gene Ontology Biological Processes functional annotation of genes annotated to H3K4me1, H3K4me3, H3K27me3, and H3K27ac peaks identified in young HSCe. Select significant (FDR<0.05) categories are shown for each modification.

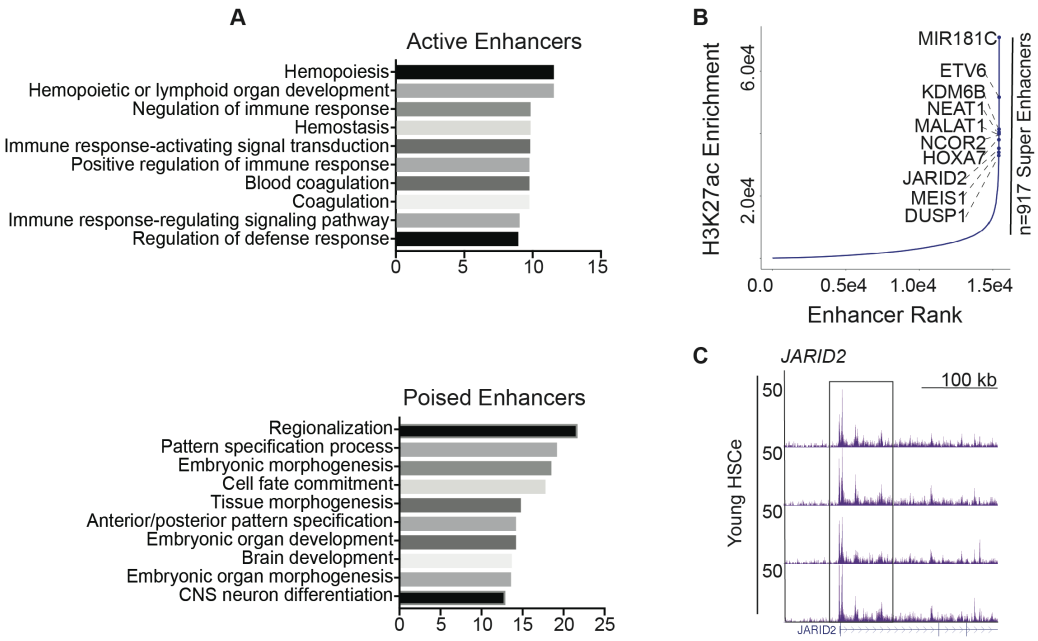


Figure 3.2: Enhancers in HSCe mark cell identity genes. (A) Top 10 significant gene-ontology biological processes enriched in genes annotated to active enhancers (top) and poised enhancers (bottom), ChIP-enrich. **(B)** Plot of H3K27ac peak rank versus H3K27ac enrichment as determined by ROSE. Points represent select super-enhancers within the top 30 most significant super-enhancers. **(C)** UCSC tracks depicting a super-enhancer enriched in HSCe. Each track is derived from 1 donor and is normalized to its corresponding input as well as number of reads.

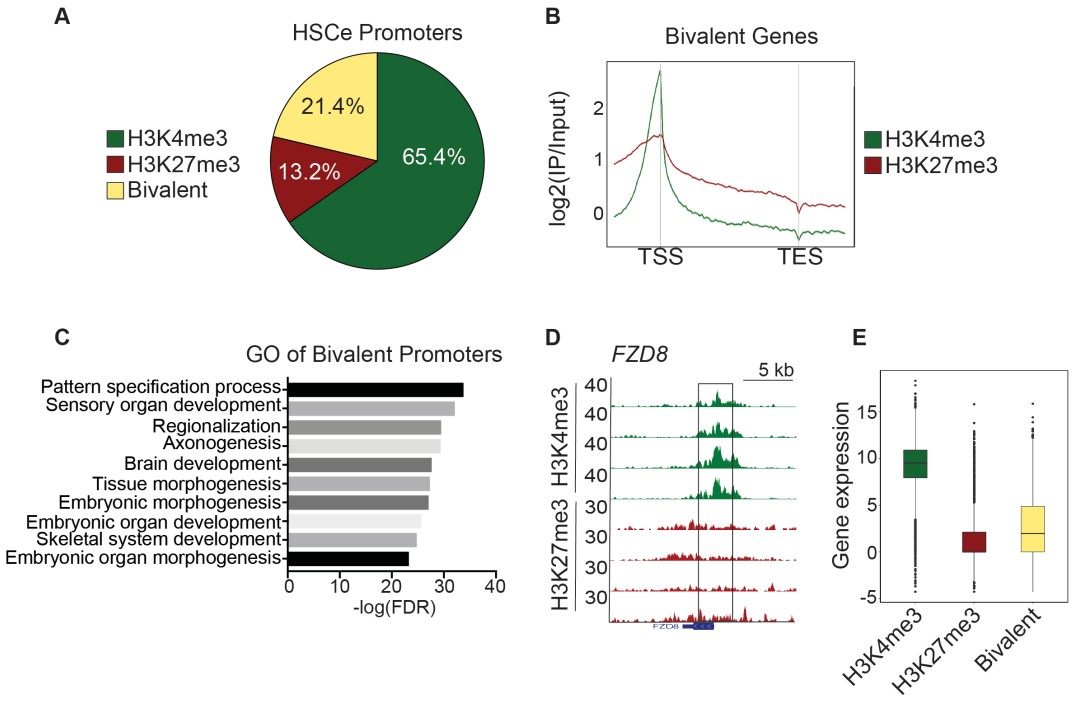


Figure 3.3: Bivalent promoters mark silenced developmental genes. (A) Pie chart of the percentage of genes with promoters (-/+ 1000 kb of TSS) that are bivalently marked or marked by only H3K4me3 or H3K27me3. **(B)** Density plot of H3K4me3 and H3K27me3 signal over gene bodies of genes with bivalent promoters. The $\log_2$(Pooled IP/Pooled Input) is shown for each histone modification. **(C)** Top 10 significant (FDR < 0.05) Gene Ontology Biological Processes associated with genes annotated to bivalent promoters. **(D)** UCSC tracks depicting a bivalent promoter. Each track is derived from 1 donor and is normalized to reads per million as well as its corresponding input. **(E)** Boxplots of the regularized log counts (DESeq2) for genes with promoters marked by H3K4me3 alone, H3K27me3 alone, or bivalently marked. Counts for each gene were averaged across all donors. Genes with an expected count of <1 have negative values.

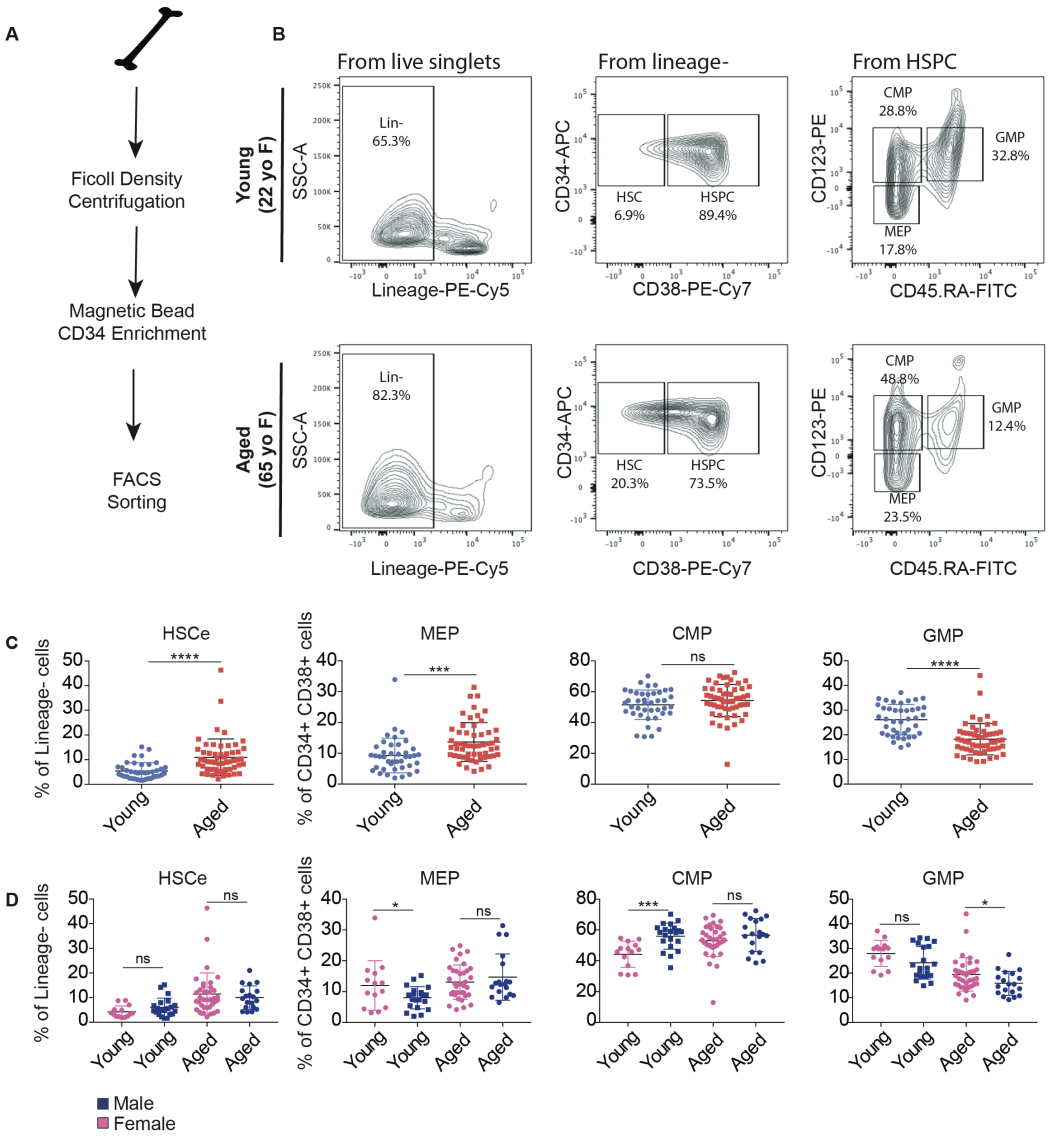


Figure 3.4: Altered frequencies of hematopoietic progenitors with age. (A) Cell isolation schema used to isolate HSCe (Lin-, CD34+, CD38-) and HSPC (Lin-, CD34+, CD38+) from human healthy bone marrow donors. **(B)** FACS gating strategy. Shown are representative plots for a young donor (top) and aged donor (bottom). **(C)** Plots of the frequencies of HSCe, MEP, CMP, and GMP in young and aged samples. Each dot is representative of 1 donor (n=41 young and n=55 aged). **(D)** Plots of the frequencies of HSCe, MEP, CMP, and GMP in young and aged samples, segregated by sex. p-values calculated using t-test in Prism, p>0.05=ns, p ≤ 0.05=*, p≤0.01=**, p≤0.001=***, p≤0.0001=****.

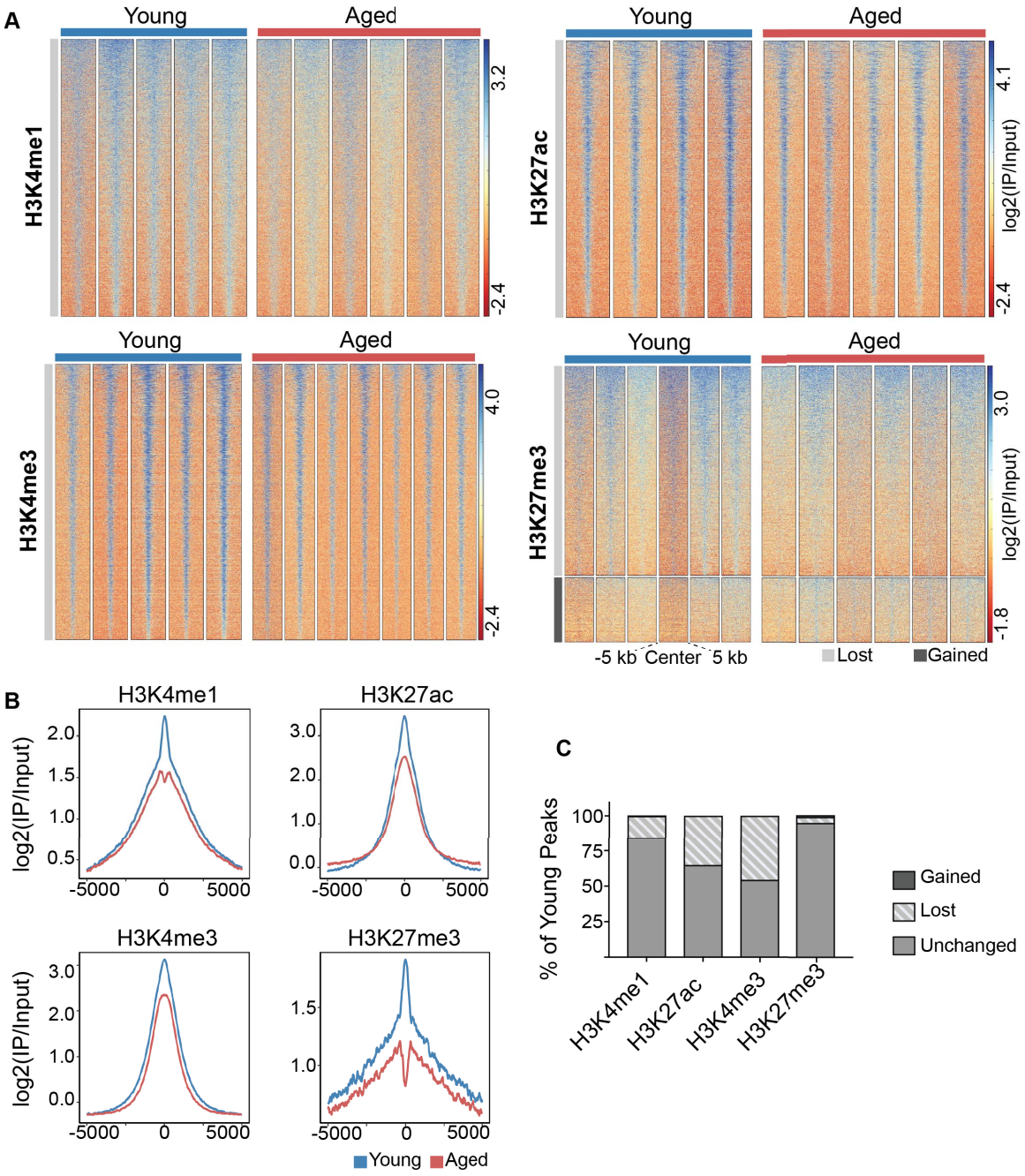


Figure 3.5: Loss of activating histone modifications with age. (A) Heatmap representation of regions with either loss or gain (log_{10} likelihood ratio > 3) of H3K4me1, H3K27ac, H3K4me3, or H3K27me3 signal in aged HSCe compared to young. The log_2(IP/Input) signal is plotted for each replicate, centered on the differential peak +/- 5 kb. Each column is representative of an individual replicate. (**B**) Metaplots of log_2(IP/Input) signal at regions that lose (log_{10} likelihood ratio > 3) H3K4me1, H3K27ac, H3K4me3, or H3K27me3 signal in aged HSCe compared to young. Represented are 10-kb regions centered around the differential peak areas. (**C**) Bar plots illustrating the percentage of peaks that are unchanged, gained, or lost in aged HSCe compared to young.

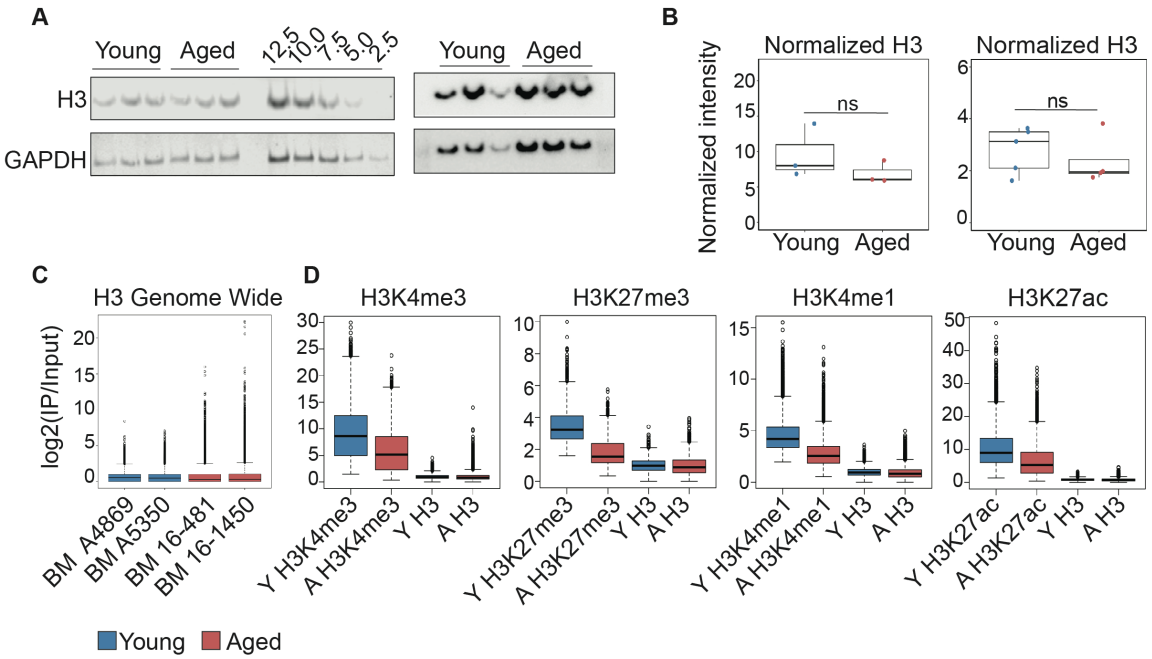


Figure 3.6: Western blot and ChIP-seq analysis of histone 3 (H3). (A) Representative Western blot analysis performed with 5,000 cells and imaged with the LiCor Odyssey (left) and 10,000 cells imaged with ECL (right) for each biological replicate of young and aged HSCe for H3 and GAPDH. HSCe from a 53 yo donor were also used to establish the linear range of the assay, with 2,500-12,500 HSCe loaded per lane (left). **(B)** Densitometry was performed using Image Studio (left) and Image Lab (right) and normalized intensity of H3 was calculated using GAPDH signal. Each point is representative of a biological replicate (left) or the average intensity of the 1-2 technical replicates (right). P-value was calculated using the parametric unpaired t-test was performed using Prism software. **(C)** Boxplots of genome wide H3 enrichment for 2 young (blue) and 2 aged (red) donors. The genome was binned into 150 bp windows and $\log_2$ enrichment of H3 was calculated for each bin. **(D)** Log_2 enrichment of H3 and H3K4me3 (leftmost), H3K27me3 (second from left), H3K4me1 (second from right) and H3K27ac (rightmost) for the respective histone modifications lost with age. Boxplots are of the $\log_2$ enrichment of the Pooled IP/Pooled Input.

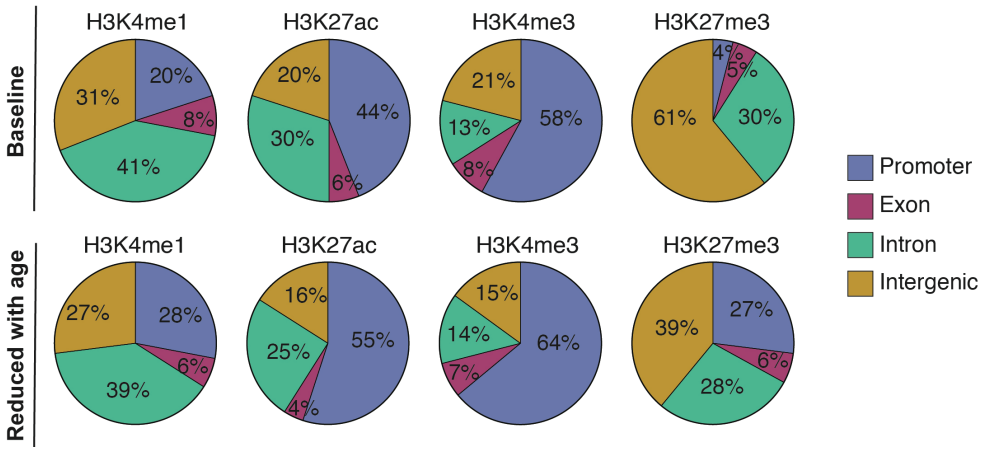


Figure 3.7: Genomic annotation of regions with age-associated histone modifications alterations. Genomic annotation of H3K4me1, H3K4me3, H3K27me3 and H3K27ac ChIP-Seq peaks at baseline (young) and at sites that lose enrichment in these marks in aged HSCe compared to young HSCe.

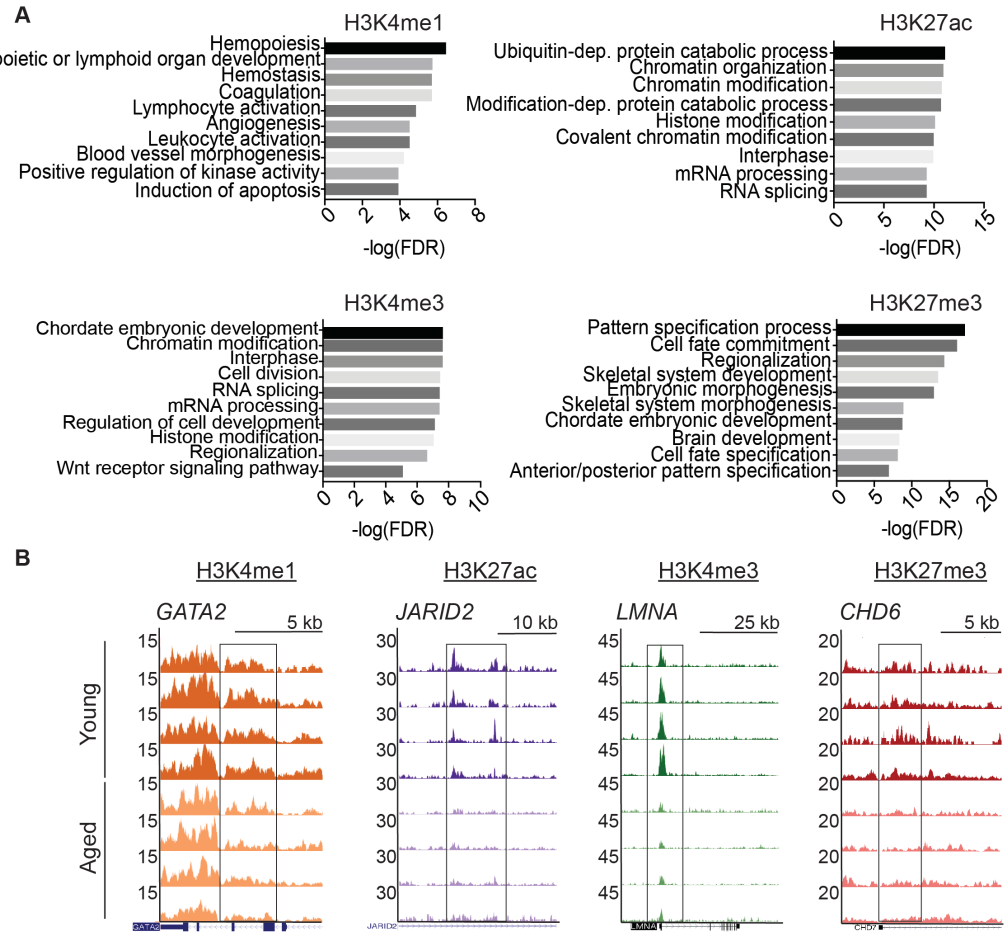


Figure 3.8: Age-associated loss of histone modifications occurs at genes involved in key HSC function. (A) ChIP-enrich Gene Ontology Biological Processes functional annotation of genes annotated to peaks that have reduced H3K4me1, H3K27ac, H3K4me3, or H3K27me3 signal in aged HSCe compared to young. Select significant (FDR<0.05) annotations are shown. **(B)** UCSC tracks for examples of genes with reduced H3K4me1, H3K27ac, H3K4me3, or H3K27me3 signal with age. Each track represents 1 donor (n=4 aged and n=4 young). IPs are normalized to reads per million and their corresponding Input.

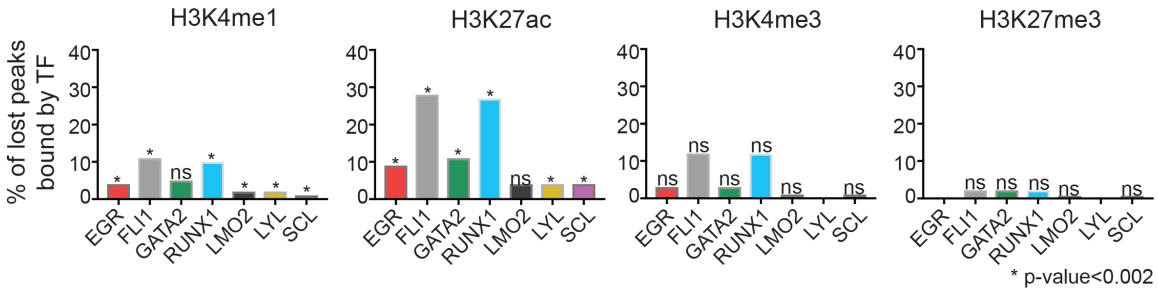


Figure 3.9 Regions with age-associated decrease in H3K4me1 or H3K27ac are enriched for hematopoietic TF binding. Bar plots of the percentage of ChIP-seq peaks lost with age that overlap transcription factor ChIP-seq peaks identified in CD34+ cells. P-values were calculated with Fisher's Exact test and corrected for multiple testing using the Bonferroni correction method. Results with Bonferroni-adjusted p-value<0.0002 were considered significant.

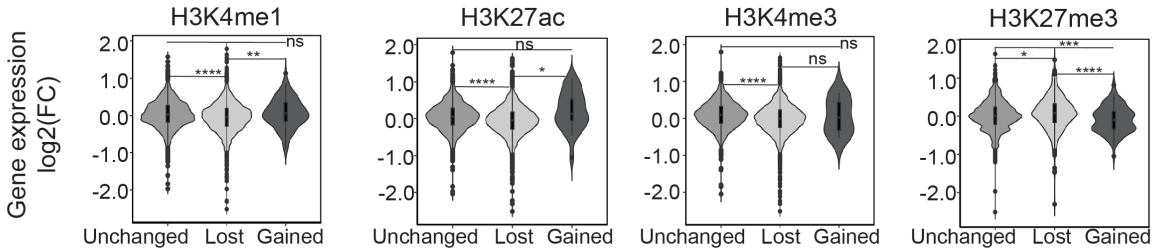


Figure 3.10: Alterations in histone modifications correlate with altered gene expression with age. Violin plots of the $\log_2$ fold-change (aged/young) in gene expression of genes annotated to regions where H3K4me1, H3K27ac, H3K4me3, or H3K27me3 is unchanged, lost, or gained in aged HSCe compared to young. p-values from Welch two-sample t-test, p>0.05=ns, p ≤ 0.05=*, p≤0.01=**, p≤0.001=***, p≤0.0001=****.

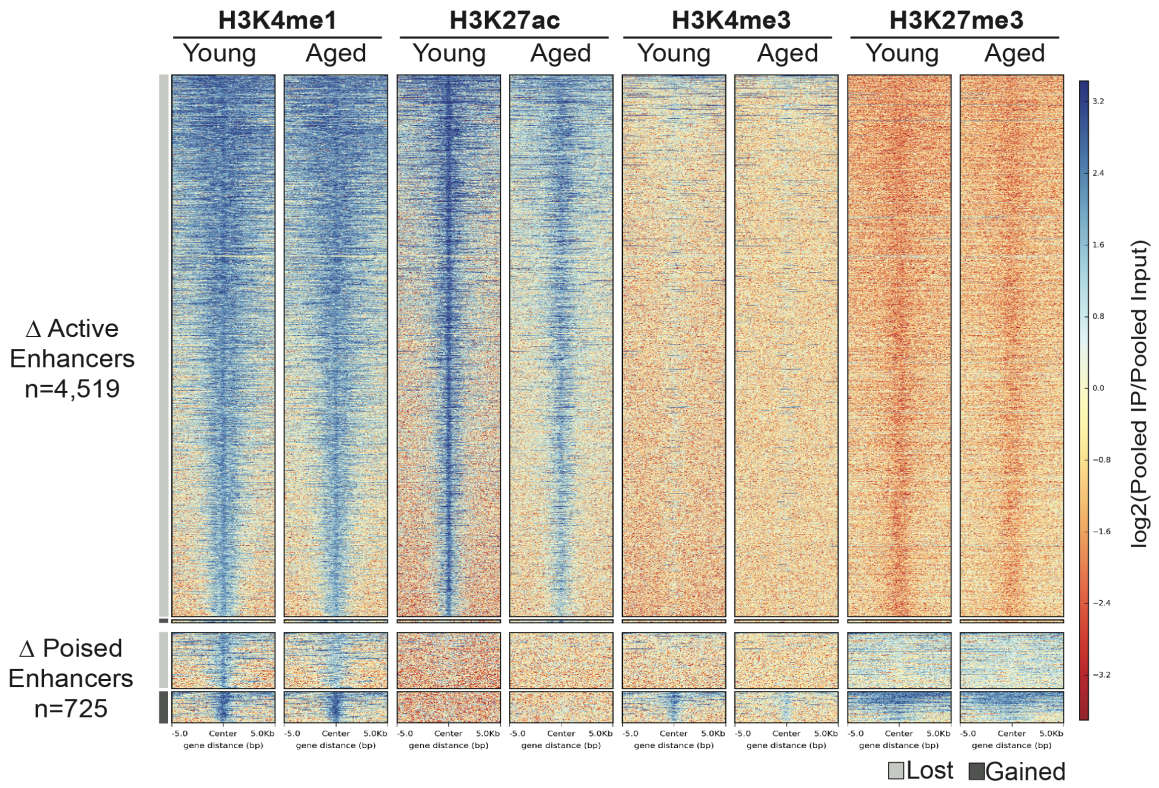


Figure 3.11: Reduced H3K4me1 and H3K27ac at enhancers with age. Heatmap of H3K4me1, H3K27ac, H3K4me3 and H3K27me3 signals at regions corresponding to active *(top)* and poised *(bottom)* enhancers that are enriched or depleted in aged HSCe compared to young. The $\log_2$(Pooled IP/Pooled Input) signal is plotted for each age group, centered on the differential peak +/- 5kb.

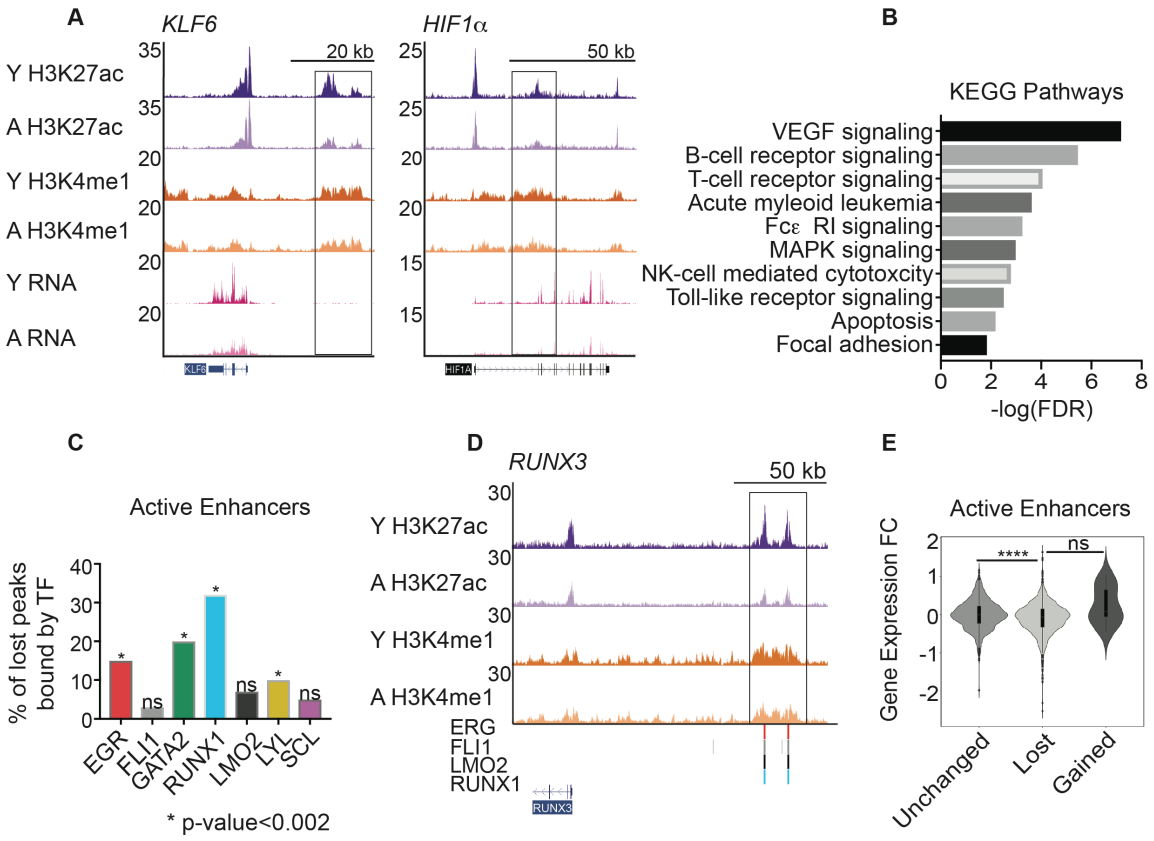


Figure 3.12: Age-associated loss of active enhancers occurs at hematopoietic and immune related genes. (A) UCSC genome browser tracks of the putative active enhancers of *KLF6*, and *HIF1α* with decreased H3K27ac (log_{10} likelihood ratio>3) in aged HSCe. Tracks are of pooled replicates for each age group, normalized to reads per million and to the corresponding Input for ChIP-seq. (**B**) Bar plot representation of select KEGG pathways (FDR <0.05) enriched in active enhancers that are lost in aged HSCe compared to young. (**E**) Violin plot of the log_2 fold-change (aged/young) in gene expression at active enhancers that are unchanged, lost, or gained in aged HSCe compared to young. p-values from Welch two-sample t-test.

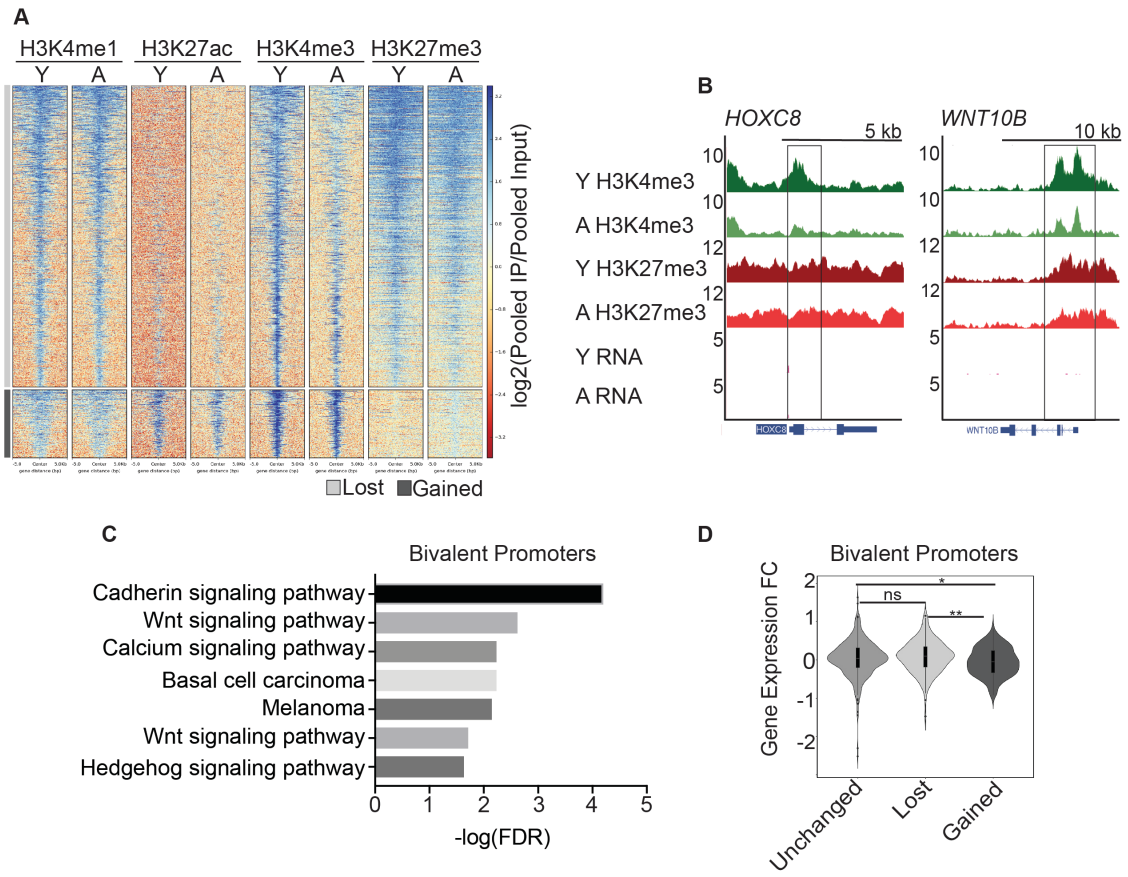


Figure 3.13: Switch from bivalency to repression with aging. (A) Heatmap of H3K4me1, H3K27ac, H3K4me3 and H3K27me3 signals at regions corresponding to bivalent promoters that are enriched (*top*) or depleted (*bottom*) in aged HSCe compared to young. The $\log_2$(Pooled IP/Pooled Input) signal is plotted for each age group, centered on the differential peak +/- 5kb. **(B)** UCSC genome browser tracks illustrating epigenetic changes at genes that lose bivalent promoters with age. All tracks are of pooled replicates for each age group (n=5-10), and normalized to reads per million. ChIP-seq tracks are also normalized to their corresponding Input. **(C)** Bar plot representation of select KEGG pathways (FDR <0.05) enriched in bivalent promoters that are lost in aged HSCe compared to young. **(D)** Violin plot of the $\log_2$ fold-change in gene expression at bivalent promoters that are unchanged, lost, or gained in aged HSCe compared to young. p-values from Welch two-sample t-test, p>0.05=ns, p ≤ 0.05=*, p≤0.01=**, p≤0.001=***, p≤0.0001=****.

CHAPTER 4

Remodeling of cytosine modifications in aged HSCe at CTCF and hematopoietic transcription factor binding sites

Both focal and global alterations in cytosine methylation (mC) have been observed in aged murine HSC, and global changes in 5-hydroxymethycytosine (5hmC) with age have also been detected in murine HSC and in other tissues[353,363,365,438-440]. In human peripheral blood, mutations in the DNA methyltransferase *DNMT3A* and the methylcytosine deoxygenase *TET2* have been observed with aging[100,101,103,104]. However, whether human HSC have altered mC or 5hmC with age and age-associated methylation changes of the same genes that are altered with murine HSC aging was unknown. Therefore, we examined how these cytosine modifications change with age in human hematopoietic stem and progenitor cells.

Cytosine modifications in young hematopoietic progenitors

We first sought to characterize mC in human HSC and determine if 5hmC profiles in hematopoietic progenitors are similar to those described in other cell types. To do this, we examined the landscape of two cytosine modifications, mC (n=7 donors) and 5hmC (n= 7 donors). With enhanced reduced representation bisulfite sequencing (ERRBS), we were able to measure methylcytosine at 3,307,233 CpGs. These CpG were mostly located

at promoter regions and were highly enriched for CpG islands (CpGi) (Figure 4.1A-B). In HSCe, 1,314,544 of the covered CpG were methylated (average methylation >80%). In comparison to all CpG covered by ERRBS, these mC were enriched at gene bodies and intergenic regions ($p<2.2e-16$, binomial test) (Figure 4.1A). We next assayed genome-wide 5hmC localization using hydroxymethylated DNA immunoprecipitation followed by massively parallel sequencing (hmeDIP-seq). Due to the high amounts of DNA required to assay 5hmC, we profiled this mark using hematopoietic stem and progenitor cells (HSPC; CD34+, CD38+, Lineage-), and identified 677,620 5hmC peaks. Both 5hmC peaks and mC were localized primarily at intergenic regions and gene bodies (Figure 4.1A). Even though there were comparable frequencies of mC and 5hmC at promoter regions, there was greater enrichment of mC at CpG islands and shores compared to 5hmC (Figure 4.1A-B). This depletion of 5hmC at CpG islands is similar to what has been observed in neuronal tissues[441]. Additionally, as has been observed in embryonic stem cells, there was enrichment of 5hmC upstream of transcription start sites, with even higher 5hmC signal at active enhancers, suggesting that 5hmC plays a role in HSPC gene regulation (Figure 4.1C-D)[291] .

Focal DNA methylation changes with age

Due to the methylation changes described in murine HSC and the finding that the DNA methyltransferase *DNMT3A* is mutated with normal human aging, we hypothesized that with age there is remodeling of cytosine modifications in human HSC[33,100,254,365]. To investigate this, we profiled mC in young and aged HSCe using ERRBS (n=5-7 per age

group)[384]. Comparison of cytosine methylation landscapes in young and aged HSCe revealed focal differences in DNA methylation in aged HSCe, identifying 529 differentially methylated regions (DMR) (beta binomial test, q-value < 0.05 and absolute difference ≥20%), encompassing 2,249 differentially methylated CpGs (DMC) (Figure 4.2A-B). These DMCs were depleted at promoters and CpG islands and enriched at intergenic regions (p<2.2e-16, binomial test) (Figure 4.2C-D). Few (<2%) of these intergenic age-associated DMCs localized to active enhancers. Yet given the CpG sparse nature of enhancers, this is expected. However, previous studies have shown that intergenic methylation of cytosines can occur within insulator regions, areas that impede promoter and enhancer interaction and are often bound by the transcriptional regulator CCCTF-Binding Factor (CTCF)[245]. Comparison of aging DMRs to CTCF ChIP-seq data from CD34+ cells shows that 28% of DMRs are within putative CTCF binding sites, suggesting that some of these age associated methylation changes may play a role in mediating long-range chromatin interactions (Figure 4.2E)[442]. Additionally, despite the more limited nature of these mC changes as compared to those of histone marks, DMRs were significantly enriched for pathways relevant to HSC biology such as WNT, Cadherin and cell adhesion pathways, which also displayed age-related changes in histone modifications (Figure 4.2F).

Aging human and murine HSC methylation alterations are distinct

As global and site-specific methylation alterations have been observed in aged mouse HSC, we next decided to investigate if our aging human HSCe methylation profile was similar to that of murine HSCs[365]. As the murine study used whole genome bisulfite

sequencing (WGBS), whereas we utilized a reduced representation approach, we limited the comparison to those genomic regions that were covered by ERRBS in our cohort. Out of the 7,505 aging murine DMRs described by Sun et al., 473 were covered in all samples in our study. Hierarchical clustering and correspondence analysis (COA) using these murine DMRs failed to segregate young and aged human HSCe (Figure 4.3A-B). While overlap of the murine and human DMRs by gene name did reveal a subset of genes (n=88) that are differentially methylated in both mouse and HSCe aging, only 51 of these DMRs became differentially methylated in the same direction in mice and humans (Figure 4.3C). Thus, the majority of differentially methylated genes are species dependent. However, given the increased heterogeneity in human samples versus pooled inbreed mouse samples, the cell types used (HSCe vs. long-term HSC), and the different assays utilized to profile DNA methylation, the disparity between age-associated methylation changes in human and mouse HSC is not surprising.

Methylation changes with age may predispose for AML

With age, there is an increased frequency of myeloid malignancies, such as acute myeloid leukemia (AML)[443]. Alterations of DNA methylation, both at promoter and enhancer elements are frequently observed in AML[394,444,445]. Therefore, we sought to investigate if mC changes with age predispose for the aberrant methylation observed in AML. Using published ERRBS methylation data from blasts from 119 AML patients (15-77 yo) and our young and aged HSCe samples, we performed k-means clustering with the regions we had previously identified as being hypermethylated (n=342) or

hypomethylated (n=187) with HSCe aging (Figure 4.4A-B)[394]. We sought to identify aging DMRs that were also differentially methylated in AML, regardless of patient age. We stipulated that if a DMR was only differentially methylated in aged AML patients, it could simply be a normal aging signature, like that of our aged HSCe donors. However, if the DMR is also differentially methylated in young and middle-aged patients it would indicate that the DMR is a universal feature of both AML and aging. Amongst our HSCe cohort and the 119 AML patients, we were able to identify clusters of aged-associated hypermethylated and hypomethylated DMRs that were also differentially methylated in AML, independent of patient age. These hypermethylated DMRs included genes such as *KLF6*, *RUNX1*, and *HOXC4/6*, and hypomethylated DMRs included *SKI* and *SOCS1* (Figure 4.4C-D).

Altered methylation of lineage specific regions with age

In murine models, aged HSCs have decreased lymphoid potential and hypermethylation of lymphoid-associated genes[23,438,446]. So, we sought to determine if regions that become differentially methylated with human HSCe aging are associated with methylation patterns found in lymphoid progenitors or other hematopoietic lineages. For this purpose, we utilized a "hematopoietic signature" generated by Farlik et al. from human hematopoietic progenitors (LT-HSC, MPP, CMP, MEP, GMP, CLP, MPP0, MPP1, MPP2, and MPP3), which is capable of classifying hematopoietic progenitors based on their methylation profile[447] (Figure 4.5A). Both young and aged HSCe displayed a methylation profile distinct from LT-HSC and intermediate of that of lymphoid and

myeloid progenitors (Figure 4.5B). Given that LT-HSCs (Lineage-, CD34+, CD38-, CD90+, CD45RA-) are a purer population than HSCe (Lineage-, CD34+, CD38-), it is reasonable that our HSCe would not be identical to LT-HSC and would have some features of more mature progenitors. Correspondence analysis using the hematopoietic signature partially segregated young and aged HSCe, suggesting that methylation of key hematopoietic lineage genes is altered with age (Figure 4.5C). Interestingly, when restricting the analysis to regions that are differentially methylated between erythromyeloid and lymphoid progenitors (n=87 covered in our cohort), principal component analysis showed no distinction between young and aged HSCe (Figure 4.5D). Furthermore, direct comparison of aging DMRs and erythromyeloid vs. lymphoid DMRs revealed only 2 regions differentially methylated in both conditions. Thus, it appears that with human HSCe aging, methylation of myeloid and lymphoid genes is not dramatically changed. However, aged HSCe do show a trend towards altered methylation when examining a broader hematopoietic signature derived from more cell types.

Abundant regional gains in 5hmC with age

Given that age-associated changes in 5hmC have been observed in murine HSC and other tissues[353,354,363-365,439], and that the methylcytosine deoxygenase *TET2* –which is important for HSC function[448]– is frequently mutated with age[100,103,104], we hypothesized that with age there are alterations in 5hmC in HSC. In order to determine this, we used hmeDIP-seq for genome-wide 5hmC profiling of young (n=7) and aged (n=5) HSPC. With aging, 5hmC signal over gene bodies showed an increase in average 5hmC

levels, while enhancer regions showed a decrease in this cytosine modification in the elderly cohort (Figure 4.6A). Supervised analysis identified 14,554 regions that gained 5hmC in aged HSPCs (Fold Change >1, FDR <0.05), which corresponds to 2% of 5hmC peaks changing with age (Figure 4.6B). Age-related hyper differentially hydroxymethylated regions (DHMRs) were enriched at intron and exons and did not overlap with regions that were differentially methylated with age (Figure 4.6C-D). Notably, transcription factor binding motif analysis revealed that hyper DHMRs were significantly enriched for binding sites for the GATA and KLF transcription factor families, indicating that changes in 5hmC at these genomic regions may be mediated by age-related changes in the levels or function of these hematopoietic transcription factors (Figure 4.6E). Notably, hmeDIP-seq of middle-aged HSPC (n=5, 45-55 yo) suggests that the hyper DHMR become progressively hyper-hydroxymethylated with age, although the increase seems to accelerate after middle age (Figure 4.6F).

As *TET2* is frequently mutated in age-associated myeloid malignancies[309-313] and patients with *TET2* mutation show widespread changes in 5hmC[317], we hypothesized that alterations in 5hmC with age may predispose for AML. To investigate this, we compared our age-associated hyper DHMR to DHMR that become hypo hydroxymethylated in *TET2* mutant AML (compared to *TET2* wild-type AML, n=157,926)[317]. Surprisingly, regions that become aberrantly hydroxymethylated in AMLs with *TET2* mutations and those which experience an increase in 5hmC with aging appear to be mutually exclusive (p<1.0e-4, Fisher's exact test). This suggests that age-associated DHMRs may not play a role in TET-mediated malignant transformation.

Summary

In this chapter, the contributions of the cytosine modifications mC and 5hmC to HSC aging were explored. We found that in young HSCe, mC is enriched at gene bodies and intergenic regions. Like ESC, young HSPC displayed enrichment of 5hmC at gene bodies and enhancers, suggesting that this modification is important for regulating gene expression in these cells[291]. With age, we observed few (n=529), but biologically relevant changes in mC. Regions that became differentially methylated with age were enriched at intergenic regions and overlapped with CTCF binding cites, indicating that they may act as insulators, preventing enhancer activation of promoters. Additionally, DMRs were associated with genes involved in WNT signaling and cell adhesion, genes that are critical for HSC function and that are aberrantly methylated in myeloid malignancies[296]. Importantly, analysis of the aging methylation profile in AML showed that some aging DMRs may predispose for leukemogenesis. While we observed poor homology between our DMRs and regions that are differentially methylated with murine aging, this is likely due to the differences in cell types profiled and the assays used.

In contrast to mC, we observed more widespread changes in 5hmC, identifying 14,554 hyper DHMR. While 5hmC was enriched in gene bodies with age, the mark was decreased at enhancers in aged HSPC, implying that reduction of 5hmC at these regions could be contributing to enhancer silencing in aged HSCe. Notably, hyper-DHMRs were enriched for binding sites for the *KLF* and *GATA* family transcription factors, which also displayed differentially regulation by histone modifications with age. In all, the data in this chapter establish that with human HSC aging, there is widespread epigenetic remodeling

of cytosine modifications at regions that are likely to be important for HSC function. However, future studies will be needed to determine if these DMRs and DHMRs contribute to loss of HSC function with age.

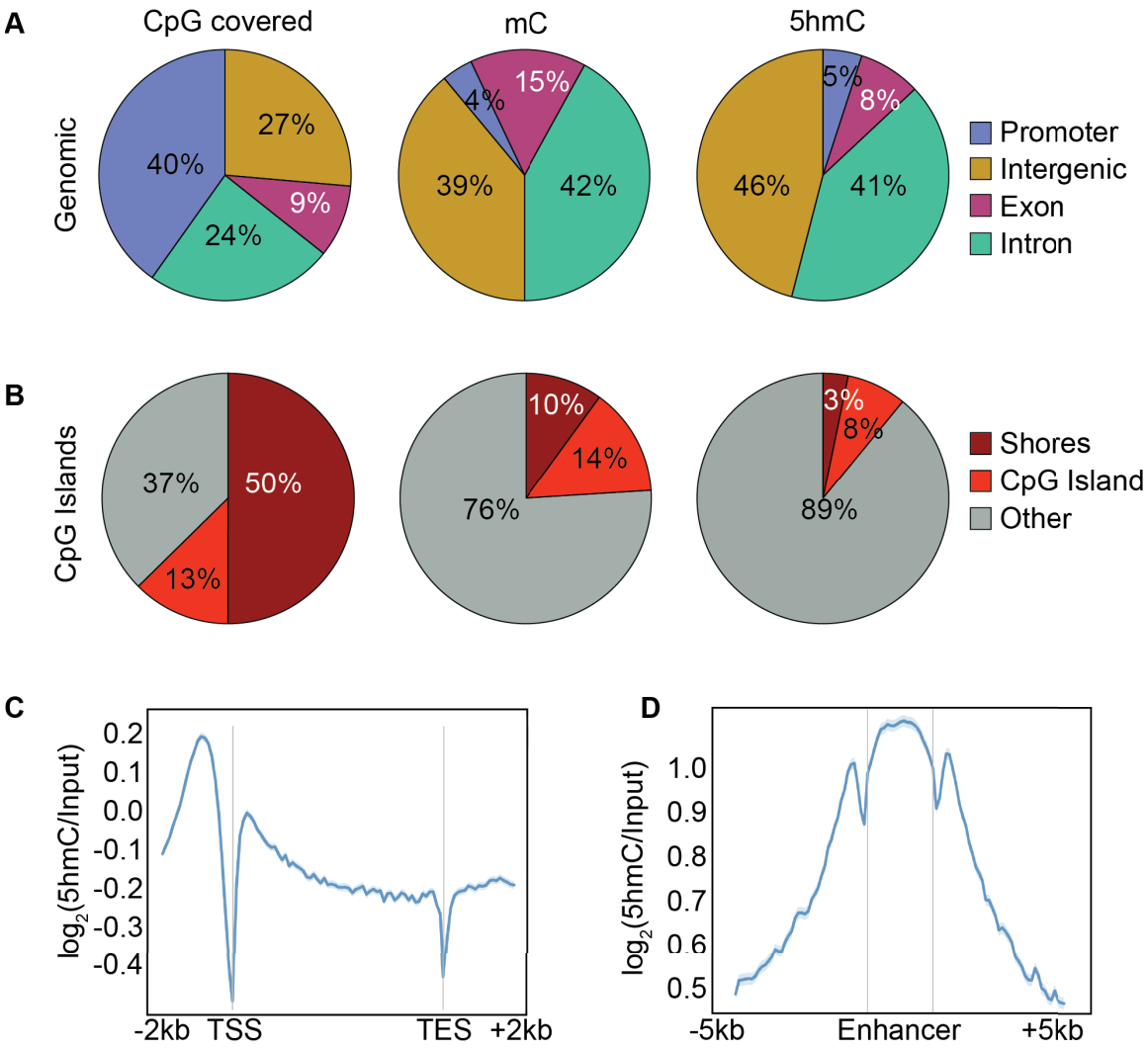


Figure 4.1: Cytosine modifications in hematopoietic progenitors. (A) Genomic annotation of all CpGs covered by ERRBS in at least 3 samples (left) and methylated cytosines (mC; average methylation of all samples >80%), and 5hmC peaks in HSPC (right). The promoter region was defined as +/- 1000 bp of TSS, Genomation. **(B)** Annotation of methylated cytosines identified in HSCe (left) and 5hmC peaks in HSPC (right) to CpG islands, shores (+/- 2000 bp of CpG), or other genomic regions, Genomation. **(C-D)** Density plot of the $\log_2$(pool5hmC/pooled Input) signal over all gene bodies (C) and HSCe active enhancers (D).

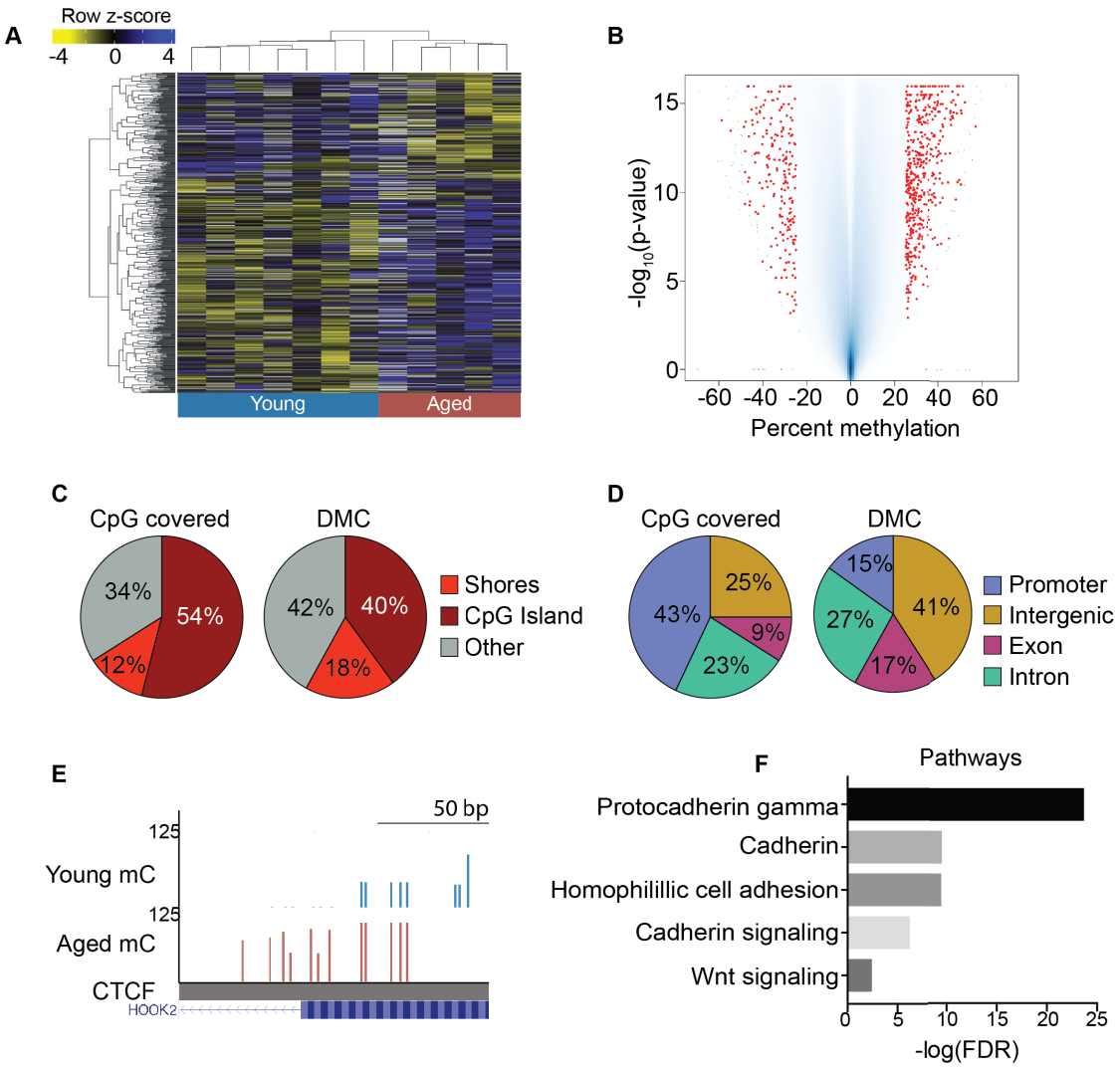


Figure 4.2: Differential methylation of WNT and Cadherin associated genes with age. (A) Row-mean-centered heatmap of the percent methylation for the 529 differentially methylated regions (DMRs) (FDR < 0.05, absolute methylation difference ≥ 20%) in aged HSCe versus young HSCe. Each row is a DMR and each column is one donor. **(B)** Volcano plot of the percent methylation versus $-\log_{10}$(p-value) of the cytosines covered by ERRBS in young and aged HSCe. Red dots denote DMC that are significantly (FDR < 0.01, absolute methylation difference ≥ 25%) differentially methylated in aged HSCe compared to young. **(C)** Genomic annotation to CpG islands (C) and genomic regions **(D)**. For each annotation type, the CpGs that were covered by ERRBS in at least 3 samples (CpG covered) and DMC (FDR < 0.01, absolute methylation difference ≥ 25%) are shown. **(E)** Example of a DMR that is hypermethylated with age and overlaps with a CTCF binding site. For methylation data, each bar is a CpG and the scale refers to percent methylation. The bottom grey bar denotes a CTCF ChIP-seq peak identified in CD34+ cells. **(F)** Bar plot representation of significant (FDR <0.05) pathways associated with genes that are differentially methylated in aged HSCe.

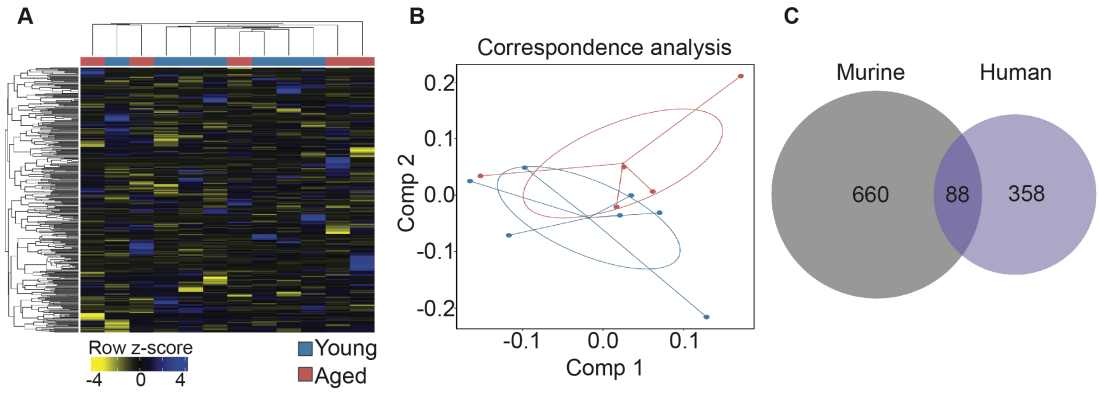


Figure 4.3: Little similarity between mC changes in murine and human HSC aging. (A) Heatmap of the row-mean-centered DNA methylation for young and aged human HSCe at the murine aging DMRs covered in all human HSCe samples (n=473 DMRs). Each row is a DMR, and each column in a donor. **(B)** Graphical representation of the correspondence analysis performed using the percent methylation for young and aged human HSCe at murine aging DMRs. Component 1 versus 2 is shown, with each point representing a donor. **(C)** Venn diagram of the overlap by gene name of genes that differentially methylated in murine or human HSC aging. Murine genes were restricted to only those corresponding to regions also present in the human genome.

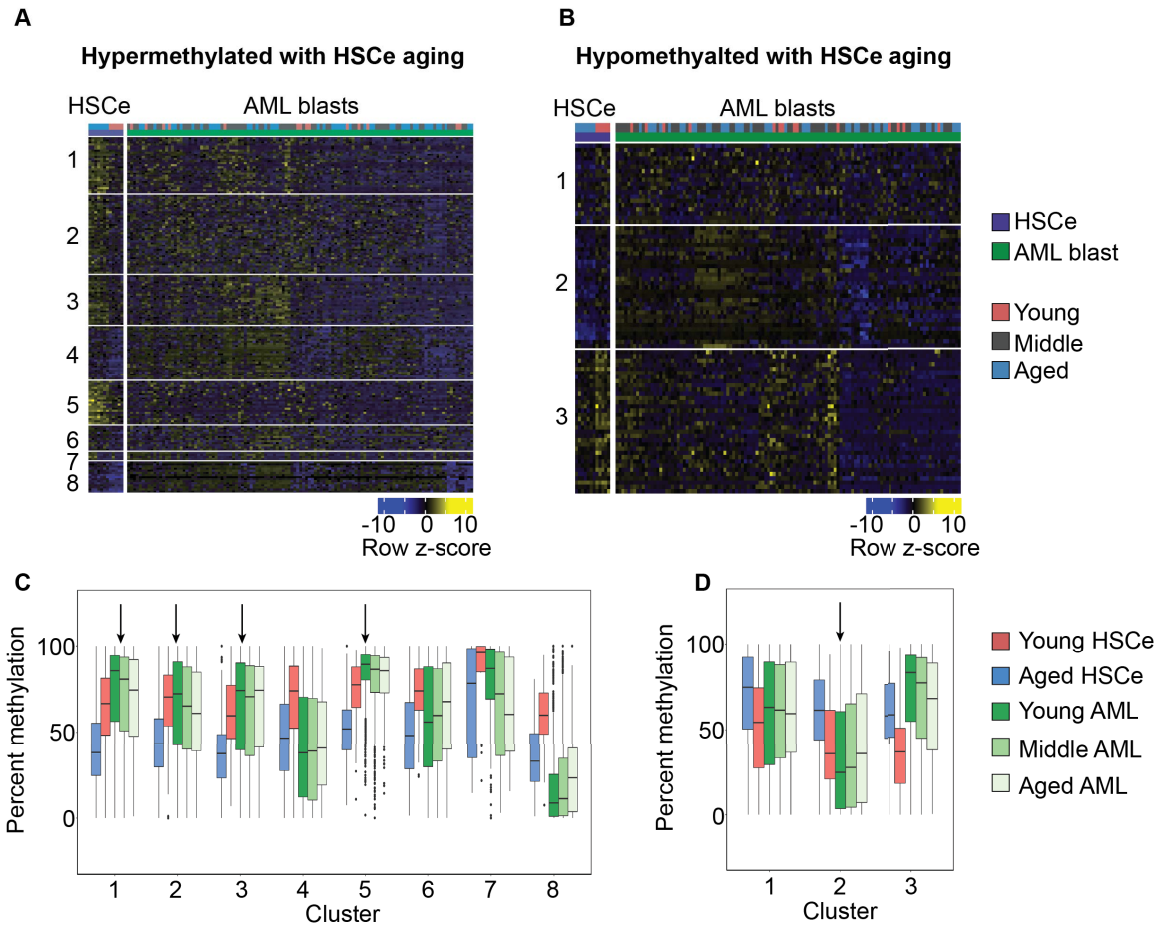


Figure 4.4: Aging methylation changes may predispose for AML. (A) Heatmaps of regions that are hypermethylated (A) or hypomethylated **(B)** in aged HSCe compared to young. Rows are ordered using k-means clustering (k=8 for hypermethylated, and k=3 for hypomethylated DMR). Bars above heatmap denote donor age and cell type. **(C)** Boxplots of the percent methylation for regions that are hypermethylated (C) or hypomethylated **(D)** in aged HSCe compared to young. Percent methylation was calculated for each DMR within the clusters generated by k-means clustering. Arrows denote clusters of DMRs that may be predisposing to AML.

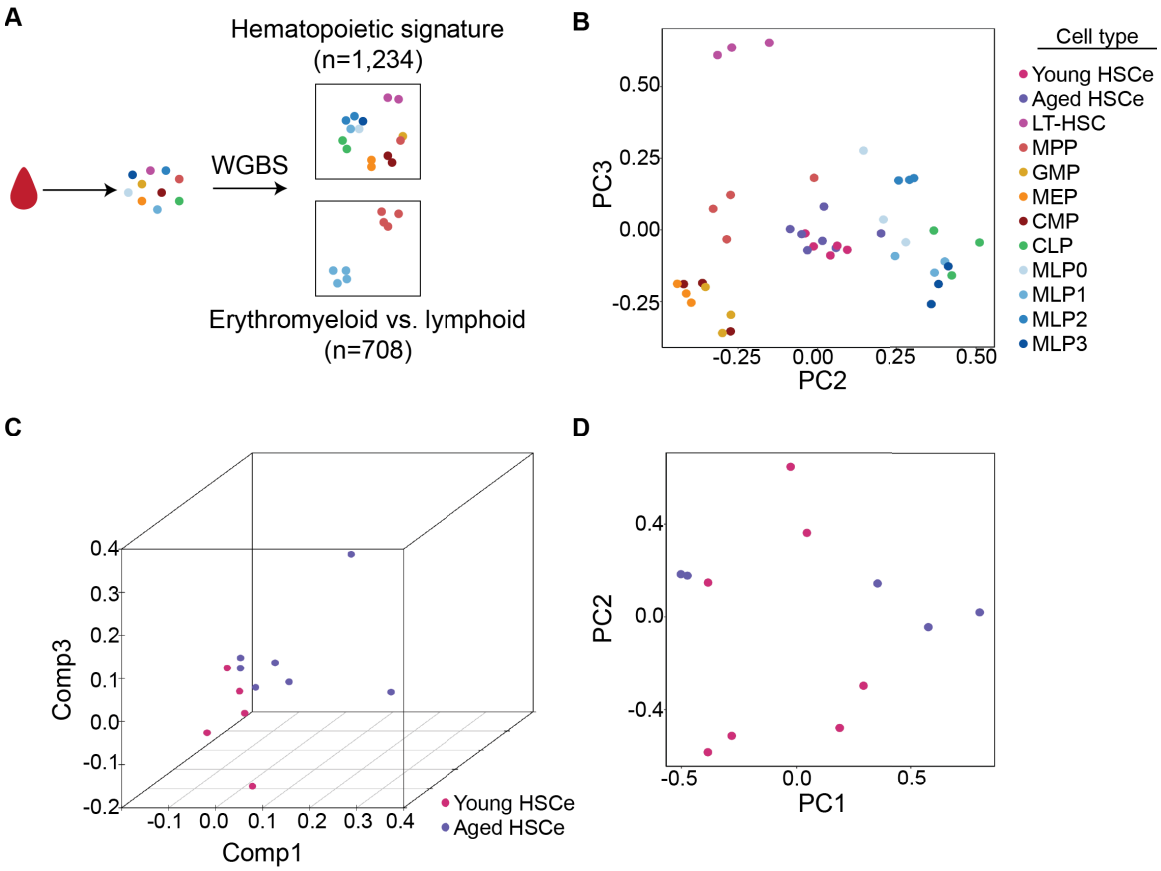


Figure 4.5: Aging DMRs are not associated with myeloid vs. lymphoid methylation differences. (A) Cartoon representation of the hematopoietic signature and erythromyeloid vs. lymphoid DMRs from Farlik et al. used for downstream analysis. **(B)** Principal component analysis (PCA) of young and aged HSCe and hematopoietic progenitors performed using the hematopoietic signature regions from Farlik et al. that were covered by ERRBS in all young and aged HSCe samples. **(C)** PCA as in (B), plotting only young and aged HSCe samples. **(D)** PCA of young and aged HSCe, using the erythromyeloid vs. lymphoid DMRs that were covered by ERRBS in all young and aged HSCe profiled.

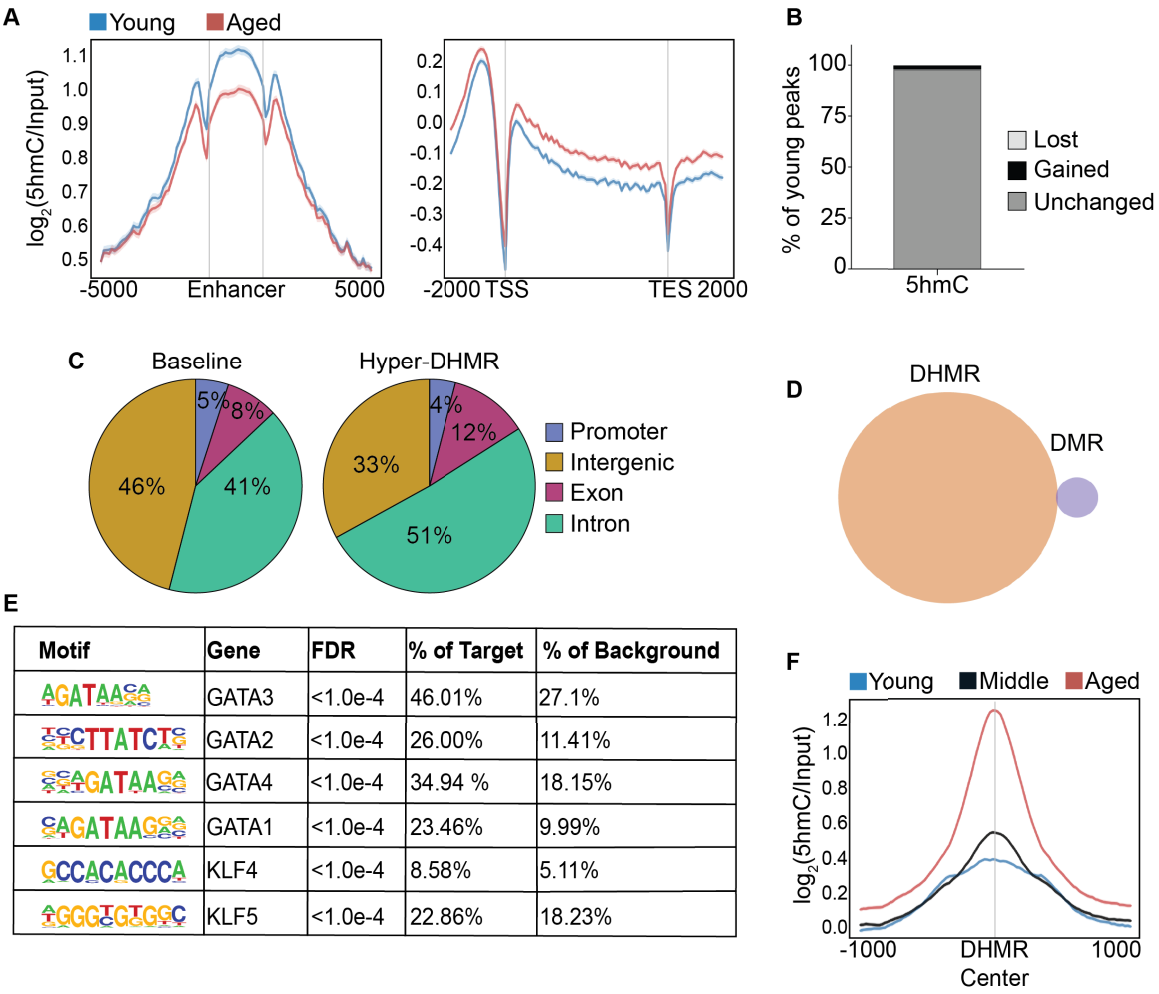


Figure 4.6: Widespread gains in 5hmC with age. (A) Metaplots of average 5hmC enrichment at active enhancers (left) and, gene bodies (right). The $\log_2$(Pooled 5hmC/Pooled Input) enrichment is shown for each age group. **(B)** Bar plot illustrating the percentage of young 5hmC peaks in HSPC with unchanged, gained, or lost 5hmC signal with age. **(C)** Genomic annotation of aging hyper-DHMR, and 5hmC peaks at baseline (young) **(D)** Venn diagram of the overlap of age-associated DMRs, and hyper-DHMR. **(E)** Transcription factor binding DNA motifs enriched in age-associated hyper-DHMRs, q-value < 0.0001. **(F)** Metaplot of differentially hyper-hydroxymethylated regions in aged HSCe compared to young, representing 2kb around regions harboring DHMRs. Signal is also shown for middle-aged donors (n=5). The $\log_2$(Pooled 5hmC/Pooled Input) enrichment is shown for each age group.

CHAPTER 5

Epigenetic remodeling with age is accompanied by altered expression of epigenetic modifiers and transcription factors

Crews et al. previously reported that changes in gene expression with human HSC aging are associated with pathways that are critical for HSC homeostasis, such as proteostasis and oxidative phosphorylation[402]. In aged murine HSCs, there is upregulation of myeloid differentiation genes, with concurrent downregulation of lymphoid associated genes, which may explain the myeloid skewing that is seen with age[438]. Given the extensive alterations of cytosine and histone modifications we observed in aged HSCe, we hypothesized that with age, there is altered gene expression of epigenetic modifiers. We also sought to determine if there were other transcriptional changes that may contribute to loss of HSC function with age. For this purpose, we utilized RNA-seq to explore how the HSCe transcriptional landscape changes with age.

Human HSCs experience age-related downregulation of epigenetic modifiers and hematopoietic transcription factors

In order to determine if expression of epigenetic modifiers changes with age, and to identify the core gene expression signature associated with HSCe aging, we performed a supervised analysis of RNA-seq profiles of young and aged HSCe (n=10 per age group). 502 genes were differentially expressed with age (adjusted p-value <0.05 and Fold change ≥ 1.5) (Figure 5.1A). Amongst these genes were genes previously reported as differentially expressed in aged murine HSCe, such as *CDC42*[72] and *JUN*[365]. In addition, we identified downregulation of certain splicing factors (*U2AF1* and *SREK1*), as well as age-related changes in key hematopoietic transcription factors, including upregulation of *GFI1b* and *EGR1,* and downregulation of *HIF1A*, *HSF1, CBFB, BCL6* and the *KLF* factors 3, 6, 7 and 10. Finally, we also detected age-related changes in the expression of several epigenetic modifiers and co-repressors (upregulation of *CITED2* and *HDAC11* with downregulation of *KDM3A, SETD6, SETD8* and *SETD1A*) (Figure 5.1B). Downregulation of *SETD1A*, a mono, di and tri H3K4 methyltransferase[165], may explain the age-related losses of H3K4me1 and H3K4me3 seen in this cohort (Figure 3.5), while downregulation of KLF transcription factors may result in loss of protection of their binding sites, rendering them more susceptible to age-related changes in cytosine modifications.

Gene set enrichment analysis (GSEA) revealed downregulation of several gene sets linked to DNA damage response and lymphoid signaling, with concomitant upregulation of inflammatory response and IFN-related gene sets in the aging HSCe fraction. In addition, enrichment for the previously reported aged human HSCe signature from Crews et al[402] was also confirmed in our cohort (Figure 5.2A-B). In order to confirm the HSC-specific nature of the aging signature identified, we compared it to a publicly

available dataset of tissue-specific aging profiles and found very little overlap with any of the reported signatures. This finding indicates that the observed expression changes are not a general signature of aging but rather they are highly specific for HSCe and clearly distinguishable not only from other tissues, but even from differentiated cells in the peripheral blood[370].

Downregulation of genes that are epigenetically deregulated with age partially recapitulates aging phenotype

Next, we sought to determine whether genes that are epigenetically deregulated with age contribute to the aging HSC phenotype. For this purpose, we focused on *LMNA* and *KLF6*, two of the most significantly downregulated genes in the aged HSCe signature. *KLF6* encodes for the Kruppel-like factor 6, a transcription factor involved in inflammation and myeloid differentiation[449], while *LMNA* encodes for the lamin A/C protein and is mutated in Hutchinson-Gilford Progeria Syndrome, a disease of premature aging, in which patients have elevated platelet levels, but no other overt alterations in myelopoieis[450]. In addition, *LMNA* has been shown to affect the differentiation of hematopoietic progenitor cells[435]. Finally, deregulation of *LMNA* has been shown to alter the interaction of chromatin with lamin associated domains[451,452] and hence, age-related changes in *LMNA* could explain the epigenetic alterations we have observed. Downregulation of both LMNA and KLF6 could be explained by epigenetic changes at their loci. The *LMNA* locus displayed marked reduction of promoter-associated H3K4me3 as well as reduction of H3K27ac at

three putative enhancer sites with age**.** Similarly, we observed loss of H3K27ac at three putative enhancers of *KLF6* in aged HSCe (Figure 5.3A).

To determine if these two epigenetically deregulated genes contribute to the aging HSC phenotype, we performed shRNA-mediated knockdown of *LMNA* and *KLF6* in human CD34+ stem and progenitor cells. We first tested the ability of *LMNA* and *KLF6* downregulation to induce expression changes comparable to those seen in aged HSCe in a panel of 5 genes from the observed age-related expression signature (Figure 5.1). *LMNA* knockdown in CD34+ cells resulted in reproducible downregulation of *BCL6*, *KLF3*, and *KLF6*, three of the genes with reduced expression in aged HSCe within 72h (n=7 biological replicates). However, expression of *EGR1*, which was increased with HSCe aging, did not significantly change with *LMNA* knockdown. Likewise, knockdown of *KLF6* (n=2 biological replicates) resulted in decreased expression of *LMNA* and *BCL6* but did not impact the expression levels of *KLF3* or *EGR1* (Figure 5.3B). Gene set enrichment analysis of RNA-seq performed on a subset of the samples in which LMNA had been knocked down (n=4), showed that genes that were downregulated with HSCe aging trended towards downregulation with LMNA knockdown, while no significant upregulation trend was observed for genes upregulated with age (Figure 5.3D). In addition to the effect on gene expression, both *LMNA* and *KLF6* knockdown resulted in an increase in the formation of granulocyte-monocyte colonies when plated on methylcellulose (with or without a resulting increase in total colony numbers), demonstrating that downregulation of these genes at least partially recapitulates the myeloid differentiation potential bias seen in aged HSCs (Figure 5.3C). However, this phenotype will need to be studied in more detail,

including using a stromal co-culture system in order to assess if lymphoid output is impaired with LMNA or KLF6 downregulation.

Altered mRNA splicing with HSCe aging

Since the splicing factors *U2AF1* and *SREK1* are differentially expressed with age, and mutations of the spliceosome components detected in MDS have been shown to cause altered splicing of hematopoietic regulators and myeloid malignancy[453,454], we next investigated whether any genes are alternatively spliced with HSCe aging. Using replicate multivariate analysis of transcript splicing (rMATS), we identified 479 alternative splicing events with aging, the majority of which were due to altered exon skipping (n=308) (absolute inclusion difference > 0.1, FDR <0.05) (Figure 5.4A-B). Interestingly, many of these skipped exon events (n=136) have also been observed in human hematopoietic progenitors upon treatment with the spliceosome inhibitor sudemycin D6, suggesting that these age associated splicing alterations may indeed be due to loss of spliceosome function with age ($p<2.2e-16$, Fisher's exact test)[404]. Amongst the genes that were differentially spliced with age were the epigenetic modifiers *KAT6A*, *JMJD1C*, *BRD2*, *CREBBP*, and *HDAC3*.

Notably, we also detected that exon 10 (Ensembl ID: ENSE00002501713) of the H3K4me1 and H3K4me2 methyltransferase *KMT2A* was less likely to be skipped in aged HSCe compared to young (Inclusion difference=-0.21, FDR<0.05). Additionally, in young HSCe there was an increased presence of an alternative 5' splice site (5ASS) in exon 54 (Ensembl ID: ENSE00003526887) of the H3K4me1 methyltransferase *KMT2C*[455] (also

known as *MLL3*) compared to aged (Inclusion difference=0.3, FDR<0.05). Interestingly, this exon contains the FYR C-terminal domain which in *KMT2A* (but not *KMT2C*) is important for dimerization with the FYR N-terminal domain[456]. Taken together, these findings suggest that epigenetic remodeling of histone modifications with age could be a manifestation of changes in RNA splicing of epigenetic modifiers. However, these splicing events will need to be validated in a new cohort and then assayed to determine what biological effect they may have.

Summary

In this chapter, we show that there is transcriptional deregulation of key hematopoietic transcription factors and epigenetic modifiers with HSCe aging. Specifically, we observed downregulation of *KDM3A, SETD6, SETD8* and *SETD1A* with age. A decrease in *SETD1A* with age may lead to the observed reductions in H3K4me1 and H3K4me3 in aged HSCe. Downregulation of *KDM3A* (also known as *JMJD1B*), a H3K9me1 and H3K9me2 demethylase suggests that there could also be a loss of silencing of constitutive heterochromatin with HSC aging, similar to what has been observed in aged MSC[368,457]. In addition, we observed that two of the genes that were most downregulated with age, *KLF6* and *LMNA*, also had significant alterations of histone modifications at their enhancers and promoters with age. Knockdown of these genes in CD34+ cells recapitulated certain features of the aging HSC phenotype, suggesting that epigenetic deregulation with age may in fact contribute to loss of HSC function. Additionally, given the importance of the nuclear lamina in chromatin organization, with heterochromatin

often being localized to lamina-associated domains (LADs)[452], it is possible that downregulation of *LMNA* with age contributes to epigenetic remodeling in aged HSCe. Like the genes mentioned above, we also found that the spliceosome components *U2AF1* and *SREK1* were downregulated with age, prompting us to investigate if changes in alternative splicing occur with HSCe aging. Analysis of splicing using rMATS revealed hundreds of transcripts that are mis-spliced with age, including many histone code writers and erasers. This result indicates that loss of function of epigenetic modifiers, due to altered splicing, may contribute to remodeling of histone modifications in aged HSCe.

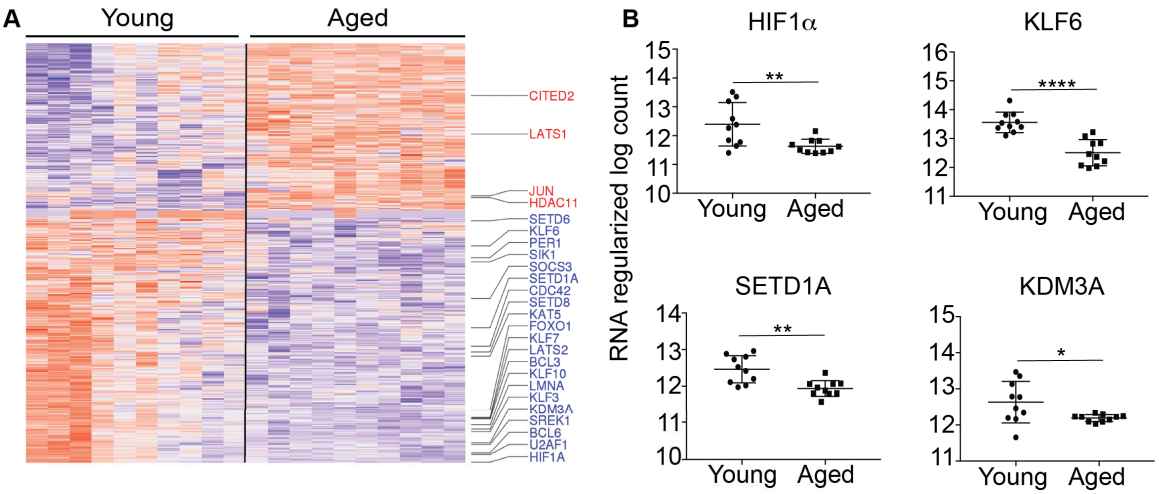


Figure 5.1: Altered gene expression of transcription factors and epigenetic modifiers with age. (A) Row-scaled heatmap of the regularized log counts from DESeq2 for the genes that are differentially expressed (FDR <0.05, fold-change>1.5) in aged HSCe compared to young. Each row is a differentially expressed gene, and each column is representative of 1 donor. **(B)** Plots of the regularized log counts (DESeq2) for select genes that are differentially expressed with age (FDR <0.05, fold-change>1.5). Each point is representative of 1 donor. p-values from Welch two-sample t-test, p>0.05=ns, p ≤ 0.05=*, p≤0.01=**, p≤0.001=***, p≤0.0001=****.

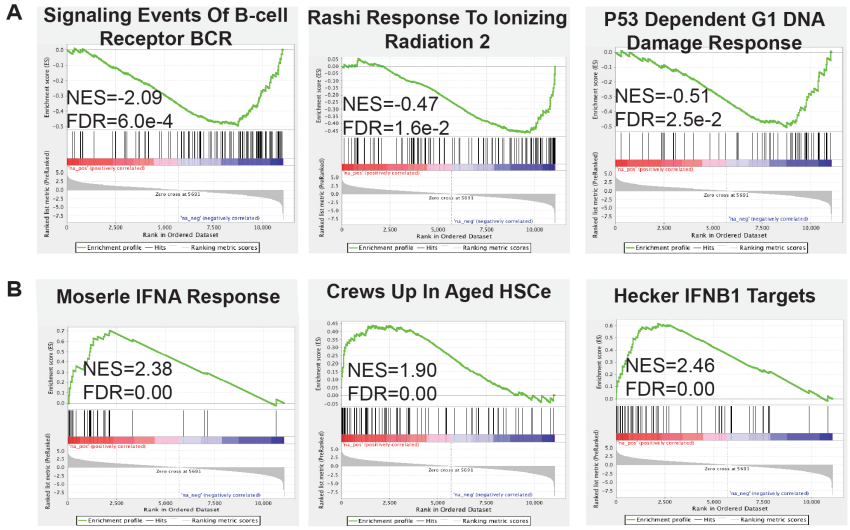


Figure 5.2: DNA damage and immune pathways are affected with HSCe aging (A) Gene Set Enrichment Analysis (GSEA) leading edge plots of gene sets associated with genes that are up- or **(B)** down- regulated with HSCe aging. GSEA was ran using a list pre-ranked by the Wald-statistic (DESeq2), with the weighted enrichment score.

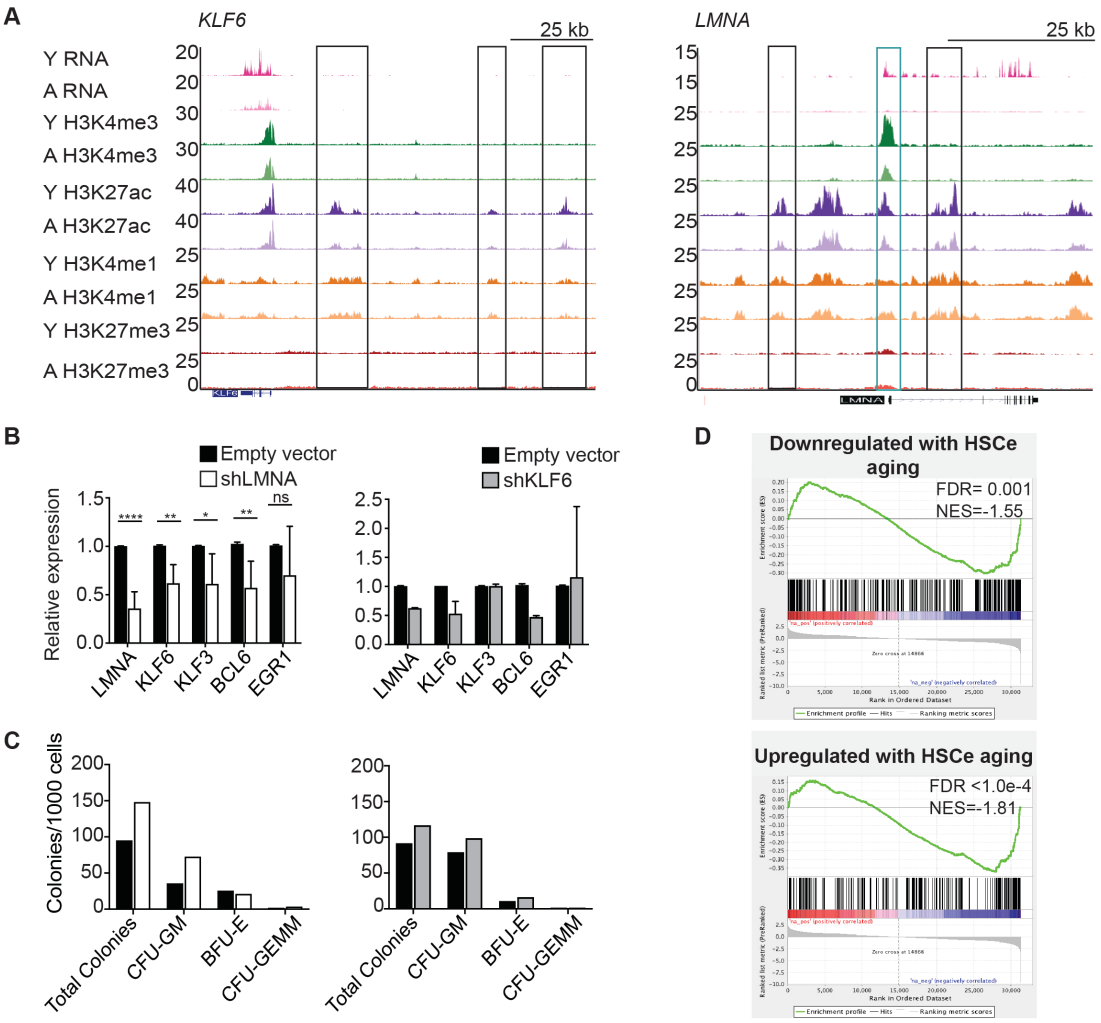


Figure 5.3: LMNA or KLF6 knockdown partially recapitulates aging phenotype. (A) UCSC tracks of pooled replicates of RNA-seq and ChIP-seq of several histone marks for young (Y) and aged (A) HSCe at the *KLF6* and *LMNA* loci. Boxes illustrate putative enhancers with age-associated decrease in H3K27ac (black boxes) and the *LMNA* promoter (teal box) with decreased H3K4me3. **(B)** qPCR quantification of HSCe aging signature genes with shRNA-mediated knockdown of *LMNA* (left) or *KLF6* (right) in CD34+ cells (n=7 and 2 independent biological replicates, respectively). Error bars depict the standard error of the mean. P-value calculated using paired t-test, p>0.05=ns, p ≤ 0.05=*, p≤0.01=**, p≤0.001=***, p≤0.0001=****. **(C)** Representative colony-forming unit (CFU) assays of CD34+ cells with knockdown of *LMNA* (left) or *KLF6* (right). Colony numbers per 1,000 CD34+ cells plated are plotted for total colony number, granulocyte-macrophage (CFU-GM), granulocyte-erythrocyte-macrophage-megakaryocyte (CFU-GEMM) and burst-forming unit erythroid (BFU-E). The average of 2 technical replicates is plotted for 1 representative donor for each target. **(D)** GSEA leading edge plots showing the enrichment of the gene sets for genes up- or down-regulated with HSCe aging in CD34+ cells with *LMNA* knockdown (n=4 independent biological replicates). GSEA was ran using a list pre-ranked by the Wald-statistic (DESeq2), with the weighted enrichment score.

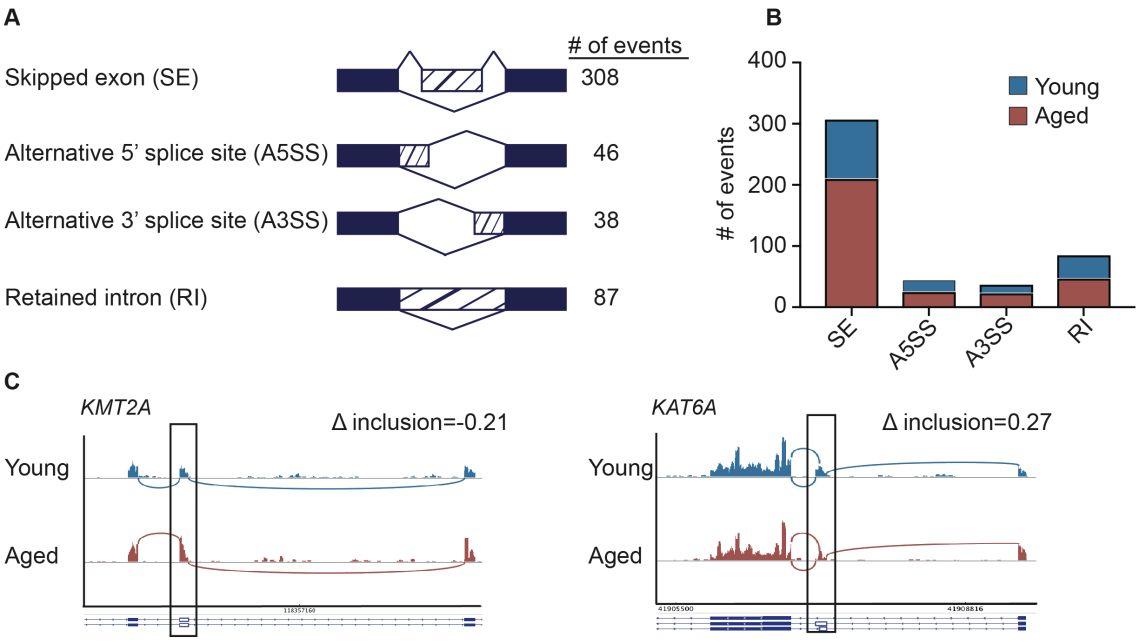


Figure 5.4: Alterations in alternative splicing with age. (A) Diagram of the types of alternative splicing events profiled, and the number of significant events found with HSCe aging (rMATS, absolute inclusion difference > 0.1, FDR <0.05). **(B)** Bar plots of the age-associated alternative splicing events. Events with greater inclusion in aged compared to young (-0.1 < inclusion difference) are colored in red. **(C)** Sashimi plots showing two age-associated skipped exon events. Pooled RNA tracks are shown for each age group.

CHAPTER 6

Discussion and future directions

A loss of HSC function with age contributes to a decline in adaptive immunity, and increased rates of idiopathic cytopenias and myeloid malignancies. While the contribution of the epigenome to aged HSC dysfunction has been explored in murine models, little has been described in human HSC. Given that the aging HSC phenotype varies among mouse strains, and that genomic elements such as enhancers are not always conserved among species, studying the epigenome in human cells is of particular importance[458,459]. Advancements in low-cell-number ChIP-seq methods have now made it possible to study the histone code in primary cells. In this dissertation, I present the first comprehensive epigenomic characterization of human HSCe with aging.

Confounding factors

While our findings suggest that the observed epigenetic differences between young and aged individuals are due to aging, there are several confounding factors that we were unable to capture and hence, address in this study. For instance, it is possible that some of the epigenetic alterations we attributed to aging may have been due to an undocumented increased rate of diabetes in the aged cohort. The frequency of diabetes

(Type I and II combined) increases from 4% in people aged 18-44 yo to 25% in individuals 65 years or older[460]. Patients with diabetes are more susceptible to infections and show impairments in innate immune function compared to those without the disorder[461]. In addition, epigenetic alterations of DNA methylation in pancreatic tissue have been observed in diabetic patients[462]. Thus, it is possible that some of the epigenetic changes we found, especially in genes associated with immunity, may be due to an increased prevalence of diabetes in the aged cohort. In addition, because there is a higher rate of non-steroidal anti-inflammatory drug (NSAID) usage in elderly individuals[463], some of the epigenetic alterations that we observed may be attributable to chronic usage of these substances by individuals in the aged group. As the name suggests, these compounds have known effects on the hematopoietic system. Short-term NSAID treatment causes an increase in mobilization of hematopoietic progenitors cells from the bone marrow niche to peripheral blood in mice, baboons, and humans[464], and exposure to common NSAIDs, including ibuprofen, aspirin, and acetaminophen, decreases antibody production in human peripheral blood *in vitro*[465]. Because NSAID treatment alters immune response and HSC frequency, it is possible that epigenetic alterations of genes involved in these processes in the aged cohort are not due to age alone, but rather, are due in part to long-term use of NSAIDs.

Another limitation of this study is that we were unable to delineate the effects of donor sex on epigenetic reprogramming with HSC aging. Given that lifespan is different between males and females[466], that sex-hormones influence HSC function[467-469], and that myeloid malignancies occur more frequently in males than females[433,434], it is likely that

with age, there are sex-associated differences in the epigenetic modifications we examined. Unfortunately, due to the short-supply of samples available for epigenetic profiling, we could not determine how HSC epigenetic aging differs between males and females. Additionally, because there was an unequal ratio of male/female donors between the two age groups (0.4 in aged group and 1.4 in young group, of donors with available information), it is possible that some of the epigenetic alterations we observed between the two age groups are a manifestation of the epigenetic differences between males and females. In order to address these concerns in the future studies, this study will continue to be extended with more donors, taking care to have equal numbers of males and females in each age group.

Owing to the limited commercial availability of aged primary bone marrow samples, young and aged samples were collected and processed in a different manner. Mononuclear cells from hip bone marrow aspirates were purchased for the young donors, whereas mononuclear cells from aged donors were isolated from bone marrow scraped from hips that had been surgically removed. Due to the more physical nature of their extraction and the longer time required to isolate them, it is possible that aged HSC were exposed to different stressors, such as longer exposure to normoxic conditions, than young HSC. Additionally, aged bone marrows were collected primarily in the morning, whereas we have no documentation when the young marrows were collected and they may well have been isolated at a different time during the day. Circadian rhythms have been shown to influence normal HSC function[470] and be altered in leukemic stem cells in myeloid malignancies[471]. At the epigenetic level, circadian oscillations of both histone and

cytosine modifications have been observed, and cytosine modifications that are influenced by circadian rhythms are also altered in murine aging[472]. In regards to this study, if the samples from the two age groups were collected at different time points, some of the observed epigenetic alterations may originate from differences in the circadian cycles. Another confounding factor relating to sample collection in this study is that all aged samples were collected from one site, whereas young samples were obtained from multiple geographic locations. Environmental factors such as exposure to chemicals and diet are known to influence the epigenome[473]. Therefore, some of the epigenetic alterations we observed in the aged cohort may be a reflection of the environmental differences in the locations that donors lived. The collection of new aged samples from the Miami area, which not only has different environmental exposures but also a more ethnically and socio-economically diverse population, will also help control for this as the study is continued in the future.

While this study was performed on a highly-purified group of cells, it is possible that shifts of sub-populations within the HSCe pool occur with age. Analysis of LT-HSC (Lin-, CD34+, CD38-, CD90+) and ST-HSC (Lin-, CD34+, CD38-, CD90+) revealed no significant differences in these populations with age (Figure S5). However, given the sizable range of LT-HSC frequencies within the HSCe compartment (3-32%, median=12% in young; 1-54%, median=12% in aged), there is notable intra-sample variability among the donors. Additionally, as cell surface markers to identify My-HSC, Ly-HSC, or platelet biased HSC are not currently available, we were unable to assess if these populations, or other currently unidentified sub-fractions of the HSCe compartment, differed between

young and aged donors. Thus, it is possible that the epigenetic alterations that we found between the young and aged group do not simply reflect changes within one cell type with age, but rather are indicative of a population shift with aging. In order to address this, single-cell epigenetic assays will be needed. While single-cell ChIP-seq or hmeDIP-seq is not currently possible, single-cell whole genome bisulfite sequencing could be used to examine the epigenetic heterogeneity of mC within the HSCe compartment.

Multiple levels of epigenetic deregulation with age converge on key hematopoietic genes and pathways

Despite the expected variability associated with human aging, we were able to capture reproducible epigenetic changes in multiple biological replicates. We show that aging results in genome-wide changes at the epigenomic level in human HSCe, affecting both cytosine and histone modifications. These age-associated epigenetic changes converge on similar genes and pathways, including hematopoietic transcription factors, epigenetic modifiers, and genes involved in cell adhesion and WNT signaling, indicating that different layers of the epigenome are implicated in changes of the same key, biologically relevant pathways in aged HSCs (Figure 6.1). This age-associated shift in epigenetic poise with concurrent altered gene expression was particularly evident at active enhancers, suggesting that like what has been observed for cancer[235,236,474,475], age-related dysfunction of human HSC may also be mediated through the deregulation of these long-range regulatory elements. As such, many of the putative enhancers that lose H3K27ac with age were associated with genes involved in lymphoid signaling, opening the

possibility that HSCe enhancer dysfunction may be one of the mechanisms through which lymphoid and immune function are compromised with aging.

Previous studies of DNA methylation in aging murine and human hematopoietic cells have yielded conflicting results, with murine studies reporting higher methylation levels in aged HSCs and human studies finding decreased methylation levels[365,438,476]. However, these studies were performed using cell populations of different purities, making any extrapolation from one species to the other difficult. When examining mC in the sorted HSCe fraction from healthy donors, we detected focal differential methylation with age, including both gains and losses of this cytosine modification, targeting key pathways in HSC biology such as WNT and cadherin signaling, and cell adhesion. Remarkably, we found that select aging DMRs are also differentially methylated in AML, irrespective of patient age. This finding suggests that changes in DNA methylation with age may indeed predispose for myeloid malignancy as they are seen universally in all age-ranges of AML. In contrast to mC, alterations in 5hmC with aging were more numerous, but still represented only a fraction of total 5hmC peaks. Comparison of regions that become hyper-hydroxymethylated with age to those that are affected by *TET2* mutation in AML[317], showed that age-associated DHMRs did not overlap with AML DHMRs and may therefore not play a role in TET-mediated malignant transformation. However, age-associated DHMR were very strongly enriched for GATA and KLF family transcription factor binding motifs, suggesting that these regions may play a key regulatory role in gene expression in aged HSPC.

Notably, preliminary analysis of middle-aged donors (45-55 yo) showed that many epigenetic changes occur gradually with age. In future studies, prioritization of the genes that become progressively altered with age may be beneficial in identifying genes that initiate aging phenotypes in HSC. Middle-age donors showed greater heterogeneity at the gene expression level, and less so when examining histone and cytosine modifications. Given the small number of donors (n=4-7), this variation was difficult to control for. Heterogeneity in the middle-aged group may be a true representation of the biological process of aging, or it is also possible that hormonal changes and sex-specific differences in people of this age group contributed to the observed variation. Estradiol has been shown to increase hematopoietic stem cell self-renewal and retention in the vascular niche[467,468], while androgens increase erythropoieisis[469]. Thus, differences in hormone levels amongst individuals in this age-group due to menopause, andropause, or hormone therapy replacement therapy, could contribute to variation of HSC phenotypes. Unfortunately, no such data was available for donors profiled in this study. In the future, epigenetic and transcriptome profiling should be performed on more middle-age donors, ideally with donor hormonal status taken into account.

Age-related epigenetic changes may be a result of HSC reprogramming

Given the reports of the development of clonal hematopoiesis associated with aging[100,101,103,104], it is possible that the observed changes in the HSCe epigenome and transcriptome with age were due to the expansion over time of a previously existing subpopulation within the young HSCe fraction, rather than true HSCe reprogramming. In

order to investigate this, we performed single-cell RNA-seq (scRNA-seq) of HSCe from young and aged donors and compared the expression profiles of individual young cells to the age-acquired HSCe expression profile identified by the bulk, steady-state RNA-seq in our 65-75 yo cohort. This preliminary analysis revealed that while 4 out of 208 young HSCe cells possessed elements of the aged HSCe gene signature, no single young HSCe cell displayed the complete aged HSCe expression profile identified in the bulk Lin-CD34+CD38- compartment, indicating that the observed HSCe epigenetic and transcriptional changes are likely the result from at least partial epigenetic reprogramming during aging rather than the result of age-related population shifts and selection occurring in this compartment (Figure S4). However, with only 1 aged and 3 young individuals examined, and few cells captured, more donors will need to be profiled with scRNA-seq before any conclusions can be drawn.

Alterations in gene expression and splicing may contribute to epigenetic reprogramming and HSC loss of function

In addition to widespread epigenetic changes with age, we also observed alterations in gene expression of transcription factors and epigenetic modifiers with age. In contrast to murine HSC, we did not find an increase in myeloid gene expression with age. However, we did observe that select lymphoid associated gene pathways were downregulated with human HSCe aging. A possible mechanism that may explain some of the observed epigenetic changes with age is the altered gene expression or splicing of epigenetic modifiers with age, such as the downregulation of the histone 4 lysine 4

methyltransferase *SETD1A*, which may explain the reduced number of H3K4me1 and H3K4me3 peaks observed in aged HSCe. Additionally, as we and others have observed alterations in RNA splicing with HSC aging[402], and the *U2AF1* and *SREK1* splicing factors become downregulated with age, it is possible that aberrant alternative splicing of epigenetic modifiers with age may be contributing to age-associated epigenetic reprogramming. However, whether differential expression or isoform usage of epigenetic modifiers leads to alterations in the epigenome or the reverse instead is true, is yet to be determined. Future investigation of the functional and epigenetic consequences of the splicing changes we observed in *KMT2A* and *KMT2C* will help address this.

The aged bone marrow niche may contribute to HSC epigenetic remodeling with age

While our study focused on HSCe intrinsic characteristics of aging, extrinsic signals from the bone marrow niche should not be ignored. It is possible that the changes observed in aged HSCe are the result of microenvironment cues from an aged bone marrow niche, which becomes more adipogenic, with less bone cell formation. In this sense, loss of HSC function with age could be a type of "maladaptation" to a deteriorating niche[477,478]. Bone marrow transplants in mice have shown that an aged niche can contribute to myeloid skewing of HSC through CC-chemokine ligand 5 (CCL5) signaling[479]. Additionally, exposure of aged HSC to thrombo-cleaved osteoprotein improves engraftment and myeloid-lymphoid balance, while co-infusion of young endothelial cells with aged HSC can improve aged-HSC repopulating ability[480,481]. However, the precise

mechanism by which the bone marrow niche may contribute to HSC epigenetic reprogramming with aging remains to be elucidated.

Future directions to determine how epigenetic deregulation may contribute to HSC loss of function with age

An important finding of our study is the involvement of multiple layers of epigenetic changes at the *LMNA* and *KLF6* loci, resulting in strong downregulation of these genes in aged HSCe. The potential importance of this observation was validated though shRNA knockdown of *LMNA* and *KLF6* in CD34+ cells, which partially recapitulated features of the aging HSC phenotype. However, in order to truly test our hypothesis that epigenetic deregulation causes loss of HSC function with age, the role of the enhancer region itself, not just the whole gene needs to be assessed. To this end, future directions will include using chromosome conformation capture–on-chip with sequencing (4C-seq)[482] to confirm the interaction of the putative enhancers with their target promoters. To determine if deregulation of these enhancers in vitro results in an aged HSCe phenotype, CRISPR-Cas9 will be used to excise the putative enhancers for *LMNA* and *KLF6* in peripheral blood CD34+ cells from young donors. Single-cell colony forming assays[483] and liquid culture differentiation experiments will be used to determine if disruption of enhancer elements affects HSC differentiation into myeloid, lymphoid, and erythroid lineages. Additionally, transplant of CD34+ cells with knockdown of the *LMNA* or *KLF6* enhancer into lethally irradiated NSG mice will be used to assess if these enhancers affect HSC differentiation and regenerative ability *in vivo*.

Given that nuclear lamins are vital to the nuclear architecture and chromatin organization of the cell, it is possible that age-related downregulation of *LMNA* contributes to the wider remodeling of the HSCe epigenome with aging. Within the cell, there are both lamina-associated domains (LADs), which are associated with heterochromatin and are marked by H3K9me3, and inter-LADs, which are associated with active gene expression and histone marks such as H3K4me1, H3K4me3, and H3K27ac[452]. Modification of the nuclear lamina or the chromatin interactions with the lamina can alter gene expression and specification. For example, HDAC3-mediated chromatin-lamina interactions influence lineage fate of cardiac stem cells, and repositioning of the Bcl11b enhancer–lamina interaction is important for T-cell fate determination[484,485]. Furthermore, disruption of *LMNA* in mesenchymal stem cells and hematopoietic progenitors alters differentiation, although whether this causes similar epigenetic remodeling as seen with aging, is unknown[435,486]. In the future, it would be interesting to perform ChIP-seq for H3K4me1, H3K4me3, H3K27ac, and H3K27me3 in young HSCe with *LMNA* knockdown, to determine if loss of LMNA recapitulates the epigenetic alterations we observed in aged HSCe.

Besides *KLF6* and *LMNA*, we also identified several other genes that are likely to be important for HSC function and are promising candidates for future follow-up. For example, there is reduced H3K27ac signal at a putative enhancer of runt-related transcription factor 3 (*RUNX3*) with age. Notably, the enhancer appears to encode an eRNA and expression of *RUNX3* is significantly decreased in aged HSC compared to young (Figure S6). Strikingly, knockout of Runx3 in aged mice results in the development of a

myeloproliferative disorder[487]. However, it is not currently known if this putative enhancer regulates *RUNX3* or contributes to leukemogenesis. CRISPR-Cas9 of the RUNX3 enhancer in murine progenitors could be used to address this.

Another avenue that would be interesting to pursue, is the relationship between epigenetic deregulation with HSC aging, and the epigenetic factors involved in regulating HSC lineage potential. As HSC become progressively myeloid biased with cell replication[4], it appears that HSCs possess an epigenetic memory that is important for determining lineage fate. Since the frequency of myeloid and platelet biased HSCs increases with age[9,10,18], it is possible that the epigenetic alterations that occur with age are in part the same as those that occur during the transformation of a Ly-HSC to a more myeloid skewed HSC. However, there are not currently human cell-surface markers available to isolate these different populations for experiments such as ChIP-seq. Although, forced proliferation of young HSC may recapitulate certain features of the Ly-HSC to My-HSC transition.

Conclusion

In summary, this comprehensive study of the epigenomic and transcriptomic landscapes of healthy, human HSCe through aging, captured dynamic age-related changes in multiple layers of the epigenome in these cells, all of which converged in a core set of key developmental and hematopoiesis-related genes and pathways. These changes, which targeted regulatory elements such as enhancers and bivalent promoters, capture a shift in the epigenetic poise of aged HSCe likely to impact their functional

capabilities. In sum, this dissertation details the first ever multifaceted epigenetic profiling of human HSC, and establishes epigenetic deregulation as a feature of human stem cell aging.

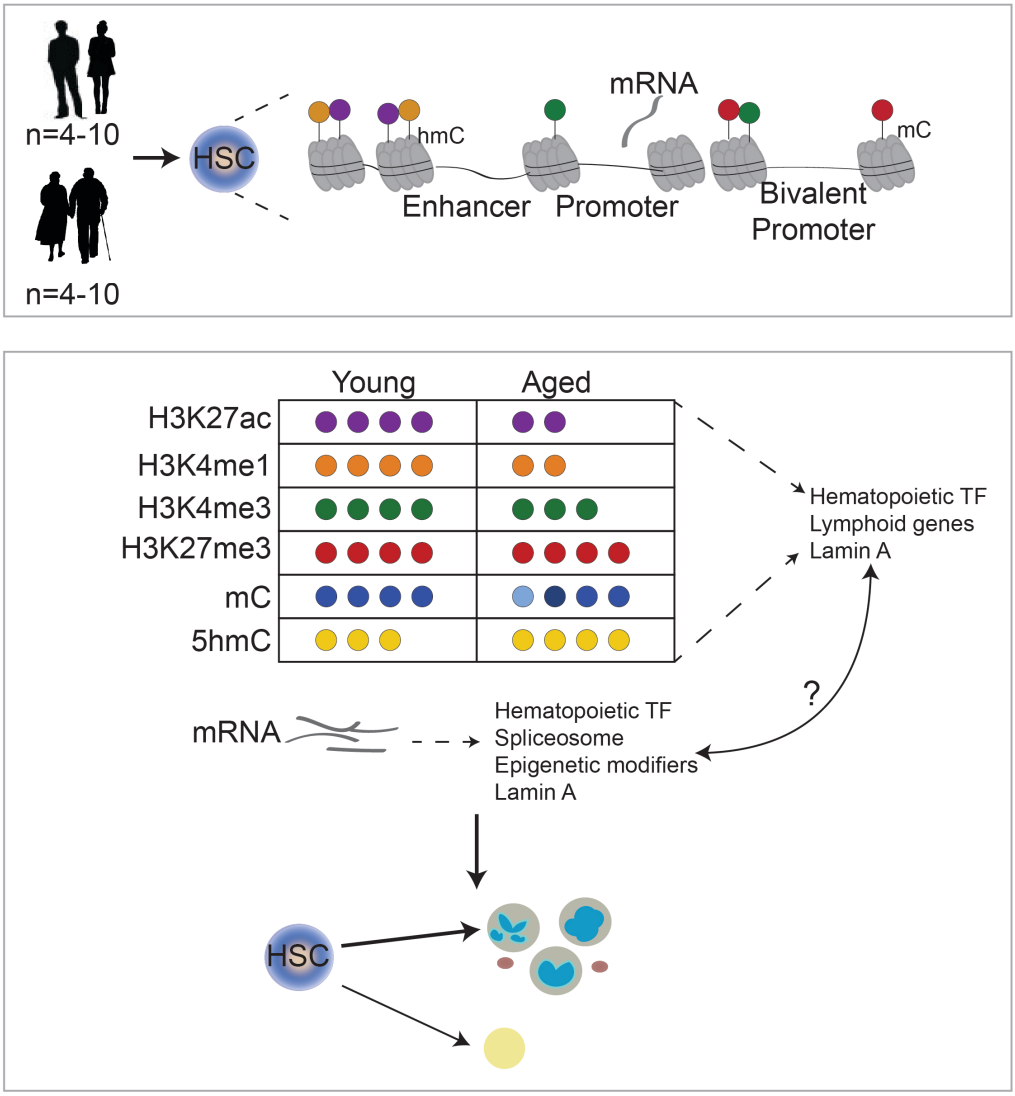
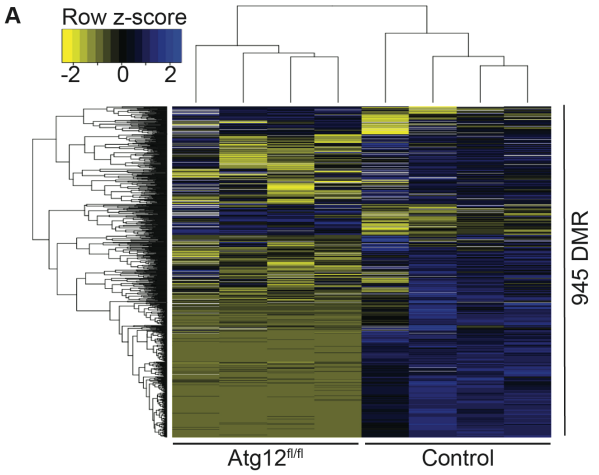


Figure 6.1: Epigenetic deregulation of HSC may contribute to loss of function with age. Schematic showing that with age multiple levels of the epigenome converge on similar pathways and genes, which may contribute to HSC myeloid bias and loss of function with age.

APPENDIX

Category	Young	Old
Total no. of donors	29	33
Male, no. (%)	Information not available for all specimens	9 (27%)
Female, no. (%)	Information not available for all specimens	24 (73%)
Donors used for HSCe RNAseq	10	10
Donors used for ERRBS	7	5
Donors used for hmeDIP-seq	7	5
Donors used for H3K4me1 ChIP-seq	5	6
Donors used for H3K4me3 ChIP-seq	5	7
Donors used for H3K27me3 ChIP-seq	6	6
Donors used for H3K27ac ChIP-seq	4	5

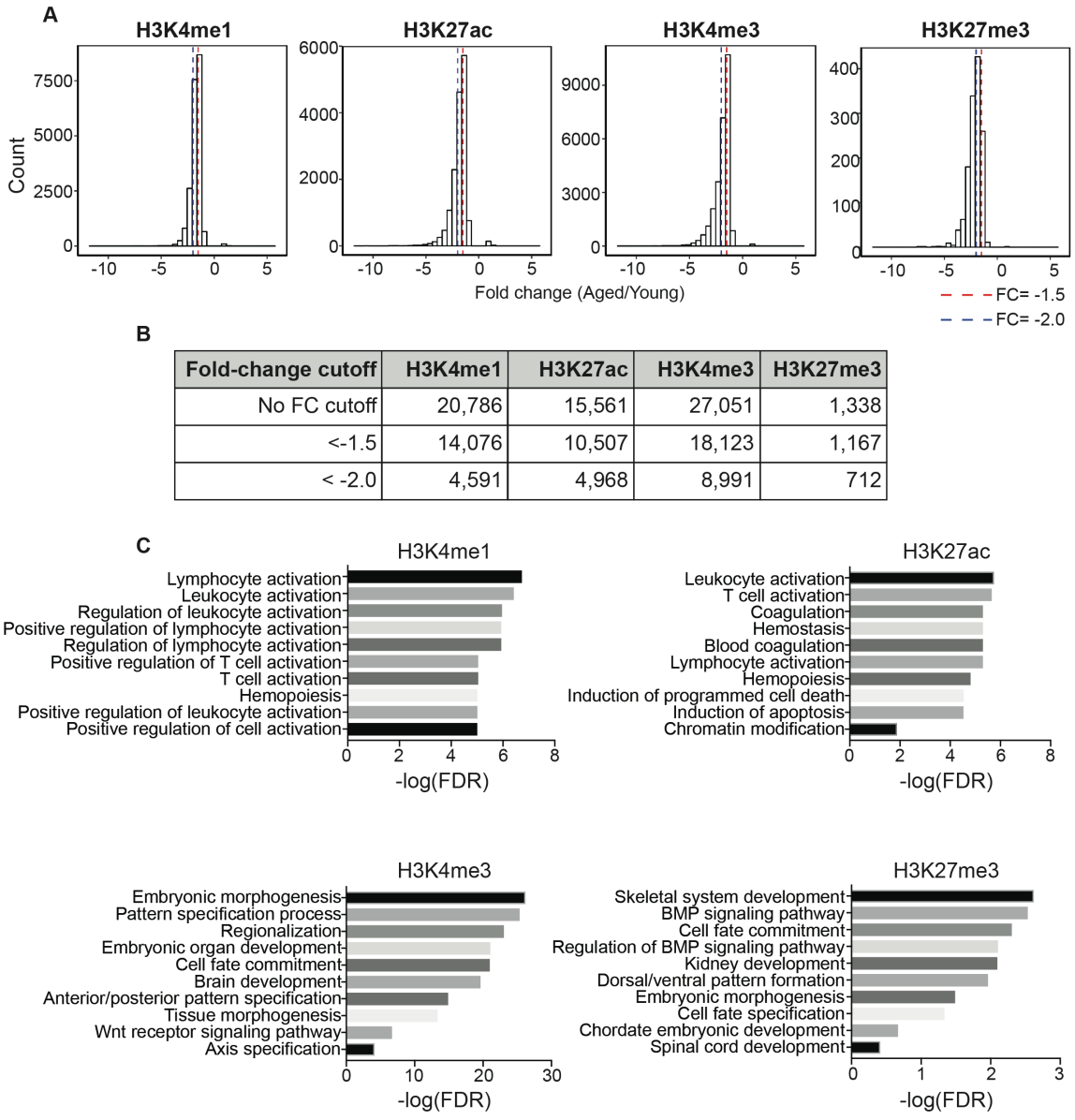
Table S1: Table of donors used for epigenetic and transcriptome profiling.



B

Hypomethylated DMRs		
Term	**Count**	**FDR**
Phosphoprotein	379	1.2E-05
Intracellular signaling cascade	76	1.0E-03
Negative regulation of amine transport	6	2.1E-03
Negative regulation of organic acid transport	6	2.1E-03
Negative regulation of L-glutamate transport	6	2.1E-03
Negative regulation of amino acid transport	6	2.1E-03
Protocadherin gamma	8	2.0E-03
Regulation of organic acid transport	7	2.9E-03
Regulation of amino acid transport	7	2.9E-03
Regulation of L-glutamate transport	6	6.9E-03
Protein of unknown function DUF634	7	6.4E-03
Regulation of anion transport	6	1.8E-02
Sequence-specific DNA binding	50	2.5E-02
Plasma membrane	173	2.7E-02
Tumor necrosis factor receptor binding	8	9.4E-02

Figure S1: Loss of autophagy results in epigenetic deregulation of HSC. (A) Row-mean-centered heatmap of the percent methylation for the 945 differentially methylated regions (DMRs) (FDR < 0.05, absolute methylation difference ≥ 20%) between Atg12$^{fl/fl}$ and control mice. Each row is a DMR and each column is one donor. **(B)** DAVID functional annotation of hypomethylated DMRs between control and Atg12$^{fl/fl}$ HSCs. Top 10 significant (FDR<0.05) categories are shown. No significant categories were identified for hypermethylated DMRs. Modified from Ho et al., *Nature*, 2017.



Fold-change cutoff	H3K4me1	H3K27ac	H3K4me3	H3K27me3
No FC cutoff	20,786	15,561	27,051	1,338
<-1.5	14,076	10,507	18,123	1,167
< -2.0	4,591	4,968	8,991	712

Figure S2: Fold change of differential ChIP-seq peaks. (A) Histograms depicting the fold change (FC) of aged/young histone modification signal for peaks with reduced (log-likelihood ratio <3) H3K4me1, H3K4me3, H3K27me3, and H3K27ac. Red and blue dashed lines are drawn at FC= -1.5 and FC= -2.0 respectively. **(B)** Table of the number of significant (log-likelihood ratio <3) ChIP-seq peaks detected with no FC cutoff, FC < -1.5, or FC <-2.0. **(C)** ChIP-enrich Gene Ontology Biological Processes functional annotation of genes annotated to peaks that have reduced H3K4me1, H3K27ac, H3K4me3, or H3K27me3 signal and FC < -2.0 in aged HSCe compared to young. Select significant (FDR<0.05) annotations are shown.

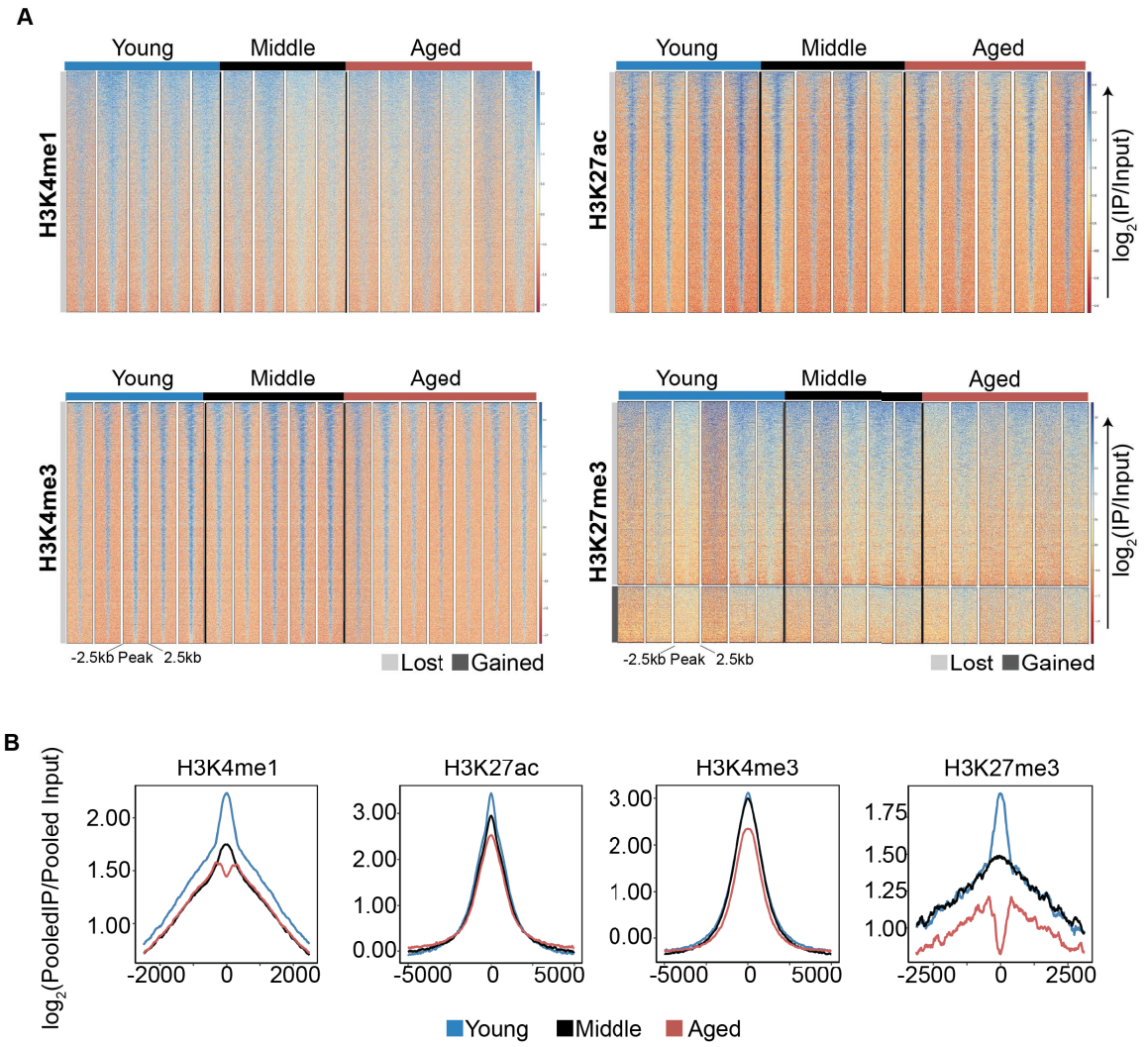


Figure S3: Progressive alterations of histone modifications with age. (A) Heatmap representation of regions with either loss or gain (log_{10} likelihood ratio > 3) of H3K4me1, H3K27ac, H3K4me3, or H3K27me3 signal in aged HSCe compared to young. The log_2(IP/Input) signal is plotted for each replicate, centered on the differential peak +/- 2.5 kb. Each column is representative of an individual replicate. **(B)** Metaplots of log_2(IP/Input) signal at regions that lose (log_{10} likelihood ratio > 3) H3K4me1, H3K27ac, H3K4me3, or H3K27me3 signal in aged HSCe compared to young. Represented are 5-10 kb regions centered around the differential peak areas.

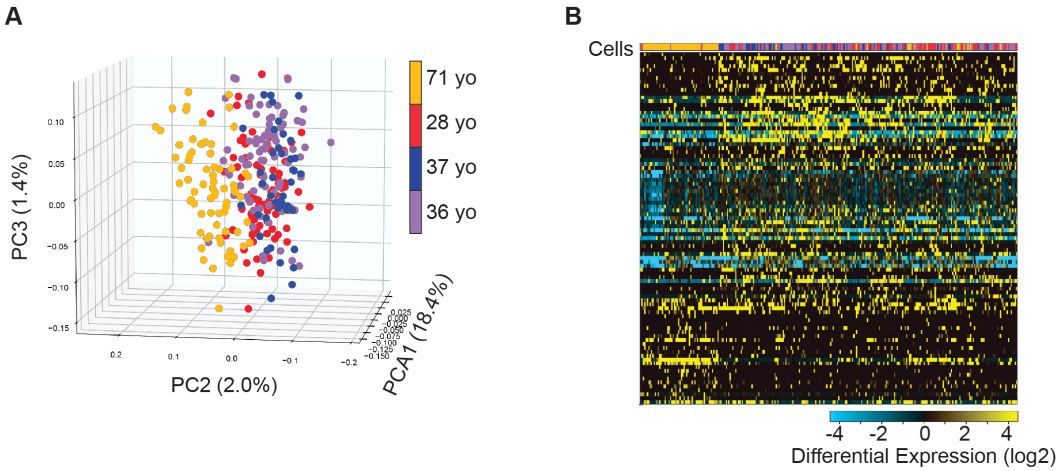


Figure S4: Epigenetic remodeling may be due to reprograming. (A) Single vector deconvolution of single-cell-RNAseq counts for the genes differentially expressed in the bulk aged HSCe compartment that were covered at the single-cell level (n=392 genes). Each point represents a single cell and each color represents a different donor of different ages. **(B)** Heatmap of single-cell-RNAseq differential expression for the 90 genes differentially expressed both in the bulk (FDR <0.05, fold-change>1.5) aged HSCe compartment and at the single-cell level (empirical Bayes moderated t-test p<0.05). Each column represents a single HSCe from a young (n=3) or aged (n=1) donor.

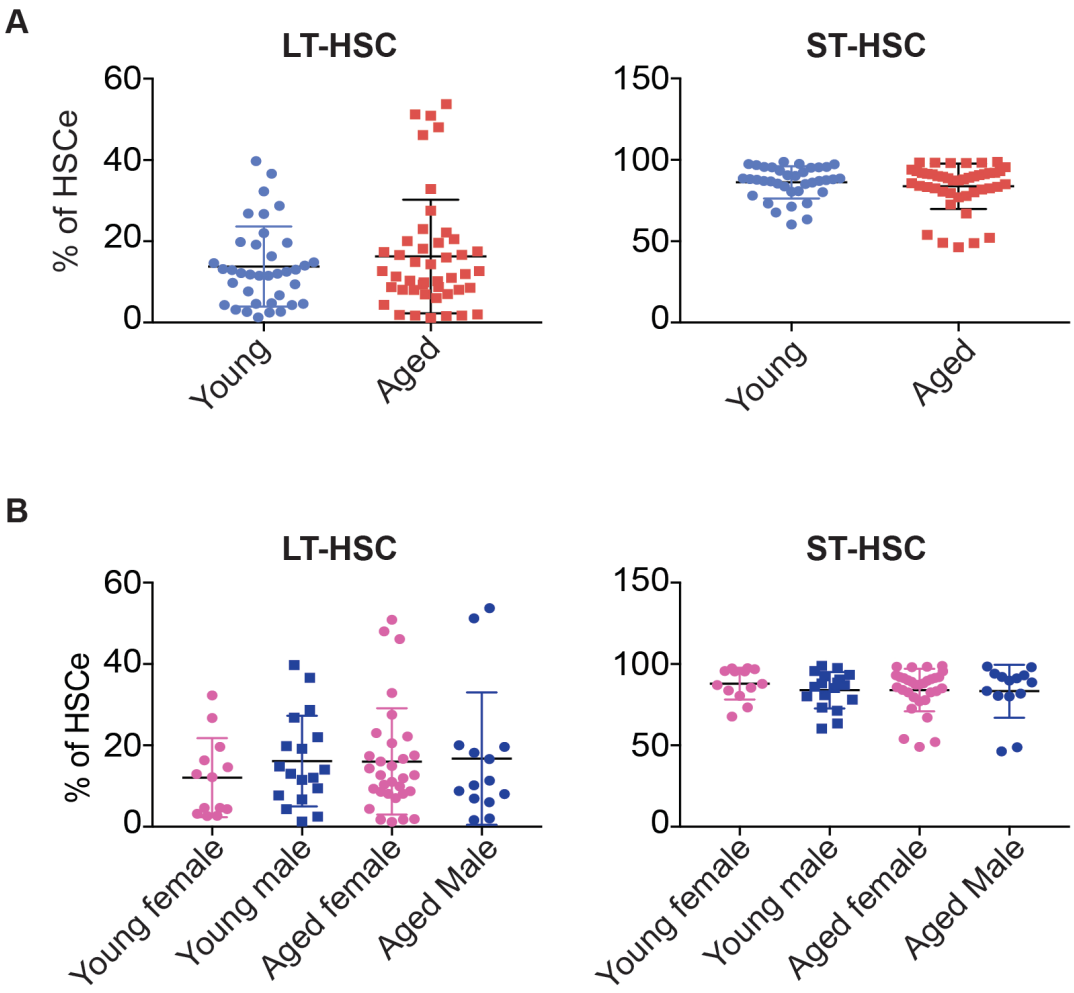


Figure S5: No significant change in LT-HSC frequency with age. (A) Plots of the frequencies of LT-HSC and ST-HSC in young and aged samples. Each dot is representative of 1 donor (n=37 young and n=45 aged). **(B)** Plots of the frequencies of LT-HSC and ST-HSC in young and aged samples, segregated by sex. No significant differences were detected using a student's t-test.

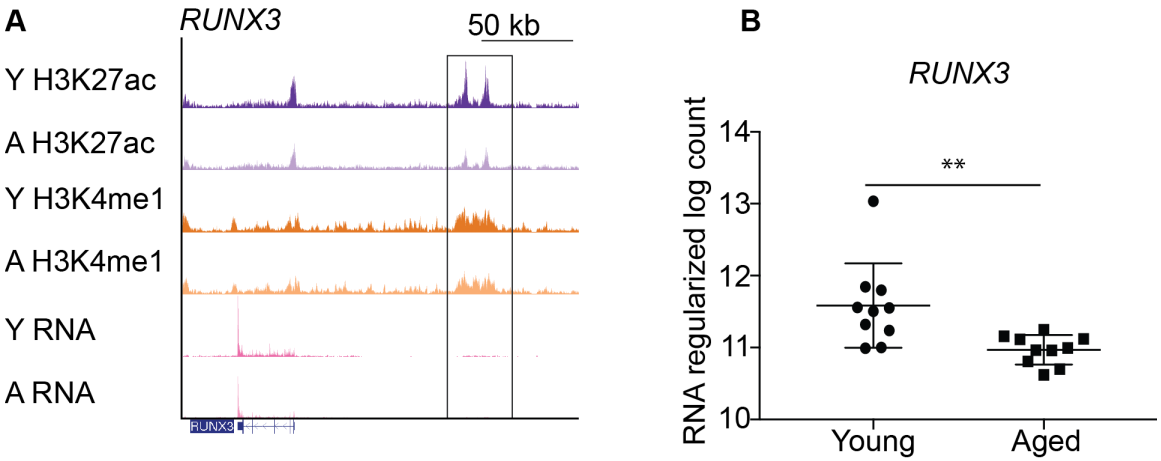


Figure S6: RUNX3 is epigenetically deregulated with age. **(A)** UCSC tracks at RUNX3. The putative enhancer region is highlighted. Pooled ChIP-seq tracks were normalized to their corresponding inputs and reads per million. RNA-seq tracks were also normalized to reads per million. **(B)** Regularized log counts of RUNX3 gene expression generated by DESeq2. Each point represents one donor. p-value from Welch two-sample t-test, p≤0.01=**

BIBLIOGRAPHY

1 Wang, J. C., Doedens, M. & Dick, J. E. Primitive human hematopoietic cells are enriched in cord blood compared with adult bone marrow or mobilized peripheral blood as measured by the quantitative in vivo SCID-repopulating cell assay. *Blood* **89**, 3919-3924 (1997).

2 Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109-1121, doi:10.1016/j.cell.2005.05.026 (2005).

3 Nombela-Arrieta, C. *et al.* Quantitative Imaging of Hematopoietic Stem and Progenitor Cell localization and hypoxic status in the Bone Marrow microenvironment. *Nature cell biology* **15**, 533-543, doi:10.1038/ncb2730 (2013).

4 Bernitz, J. M., Kim, H. S., MacArthur, B., Sieburg, H. & Moore, K. Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. *Cell* **167**, 1296-1309.e1210, doi:10.1016/j.cell.2016.10.022 (2016).

5 Dixon, R. & Rosendaal, M. Contrasts between the response of the mouse haemopoietic system to 5-fluorouracil and irradiation. *Blood cells* **7**, 575-587 (1981).

6 Venezia, T. A. *et al.* Molecular Signatures of Proliferation and Quiescence in Hematopoietic Stem Cells. *PLoS biology* **2**, e301, doi:10.1371/journal.pbio.0020301 (2004).

7 Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118-1129, doi:10.1016/j.cell.2008.10.048 (2008).

8 Busch, K. *et al.* Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature* **518**, 542-546, doi:10.1038/nature14242 (2015).

9 Beerman, I. *et al.* Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal

expansion. *Proc Natl Acad Sci U S A* **107**, 5465-5470, doi:10.1073/pnas.1000834107 (2010).

10 Challen, G. A., Boles, N. C., Chambers, S. M. & Goodell, M. A. Distinct hematopoietic stem cell subtypes are differentially regulated by TGF-beta1. *Cell Stem Cell* **6**, 265-278, doi:10.1016/j.stem.2010.02.002 (2010).

11 Dykstra, B. *et al.* Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell Stem Cell* **1**, 218-229, doi:10.1016/j.stem.2007.05.015 (2007).

12 Muller-Sieburg, C. E., Cho, R. H., Karlsson, L., Huang, J. F. & Sieburg, H. B. Myeloid-biased hematopoietic stem cells have extensive self-renewal capacity but generate diminished lymphoid progeny with impaired IL-7 responsiveness. *Blood* **103**, 4111-4118, doi:10.1182/blood-2003-10-3448 (2004).

13 Sieburg, H. B. *et al.* The hematopoietic stem compartment consists of a limited number of discrete stem cell subsets. *Blood* **107**, 2311-2316, doi:10.1182/blood-2005-07-2970 (2006).

14 Grover, A. *et al.* Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. *Nature communications* **7**, 11075, doi:10.1038/ncomms11075 (2016).

15 Sanjuan-Pla, A. *et al.* Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* **502**, 232-236, doi:10.1038/nature12495 (2013).

16 Wilson, N. K. *et al.* Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell* **16**, 712-724, doi:10.1016/j.stem.2015.04.004 (2015).

17 Carrelha, J. *et al.* Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature* **554**, 106-111, doi:10.1038/nature25455 (2018).

18 Pinho, S. *et al.* Lineage-Biased Hematopoietic Stem Cells Are Regulated by Distinct Niches. *Developmental cell*, doi:10.1016/j.devcel.2018.01.016 (2018).

19 Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194-1217, doi:10.1016/j.cell.2013.05.039 (2013).

20 Kowal, P., Goodkind, D. and He, W. An Aging World: 2015, International Population Reports. (U.S. Government Printing Office, Washington DC, 2016).

21 Kenyon, C. J. The genetics of ageing. *Nature* **464**, 504-512, doi:10.1038/nature08980 (2010).

22 Mair, W. & Dillin, A. Aging and Survival: The Genetics of Life Span Extension by Dietary Restriction. *Annual review of biochemistry* **77**, 727-754, doi:10.1146/annurev.biochem.77.061206.171059 (2008).

23 Morrison, S. J., Wandycz, A. M., Akashi, K., Globerson, A. & Weissman, I. L. The aging of hematopoietic stem cells. *Nat Med* **2**, 1011-1016 (1996).

24 Ertl, R. P., Chen, J., Astle, C. M., Duffy, T. M. & Harrison, D. E. *Effects of dietary restriction on hematopoietic stem-cell aging are genetically regulated*. Vol. 111 (2008).

25 de Haan, G., Nijhof, W. & Van Zant, G. Mouse strain-dependent changes in frequency and proliferation of hematopoietic stem cells during aging: correlation between lifespan and cycling activity. *Blood* **89**, 1543-1550 (1997).

26 de Haan, G. & Van Zant, G. Dynamic changes in mouse hematopoietic stem cell numbers during aging. *Blood* **93**, 3294-3301 (1999).

27 Geiger, H., True, J. M., de Haan, G. & Van Zant, G. Age- and stage-specific regulation patterns in the hematopoietic stem cell hierarchy. *Blood* **98**, 2966-2972 (2001).

28 Geiger, H., True, J. M., de Haan, G. & Van Zant, G. Age- and stage-specific regulation patterns in the hematopoietic stem cell hierarchy. *Blood* **98**, 2966-2972, doi:10.1182/blood.V98.10.2966 (2001).

29 Liang, Y., Van Zant, G. & Szilvassy, S. J. Effects of aging on the homing and engraftment of murine hematopoietic stem and progenitor cells. *Blood* **106**, 1479-1487, doi:10.1182/blood-2004-11-4282 (2005).

30 Rossi, D. J. *et al.* Cell intrinsic alterations underlie hematopoietic stem cell aging. *Proc Natl Acad Sci U S A* **102**, 9194-9199, doi:10.1073/pnas.0503280102 (2005).

31 Chambers, S. M. *et al.* Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation. *PLoS biology* **5**, e201, doi:10.1371/journal.pbio.0050201 (2007).

32 Sudo, K., Ema, H., Morita, Y. & Nakauchi, H. Age-associated characteristics of murine hematopoietic stem cells. *The Journal of experimental medicine* **192**, 1273-1280 (2000).

33 Beerman, I. *et al.* Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging. *Cell Stem Cell* **12**, 413-425, doi:10.1016/j.stem.2013.01.017 (2013).

34 Kollman, C. *et al.* Donor characteristics as risk factors in recipients after transplantation of bone marrow from unrelated donors: the effect of donor age. *Blood* **98**, 2043-2051 (2001).

35 Pang, W. W. *et al.* Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proc Natl Acad Sci U S A* **108**, 20012-20017, doi:10.1073/pnas.1116110108 (2011).

36 Kuranda, K. *et al.* Age-related changes in human hematopoietic stem/progenitor cells. *Aging cell* **10**, 542-546, doi:10.1111/j.1474-9726.2011.00675.x (2011).

37 Rundberg Nilsson, A., Soneji, S., Adolfsson, S., Bryder, D. & Pronk, C. J. Human and Murine Hematopoietic Stem Cell Aging Is Associated with Functional Impairments and Intrinsic Megakaryocytic/Erythroid Bias. *PLoS One* **11**, e0158369, doi:10.1371/journal.pone.0158369 (2016).

38 Mohrin, M. *et al.* Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis. *Cell Stem Cell* **7**, 174-185, doi:10.1016/j.stem.2010.06.014 (2010).

39 Reese, J. S., Liu, L. & Gerson, S. L. Repopulating defect of mismatch repair-deficient hematopoietic stem cells. *Blood* **102**, 1626-1633, doi:10.1182/blood-2002-10-3035 (2003).

40 Rossi, D. J. *et al.* Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. *Nature* **447**, 725-729, doi:10.1038/nature05862 (2007).

41 Blasco, M. A. *et al.* Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell* **91**, 25-34 (1997).

42 Nussenzweig, A. *et al.* Requirement for Ku80 in growth and immunoglobulin V(D)J recombination. *Nature* **382**, 551-555, doi:10.1038/382551a0 (1996).

43 de Boer, J. *et al.* A mouse model for the basal transcription/DNA repair syndrome trichothiodystrophy. *Mol Cell* **1**, 981-990 (1998).

44 Beerman, I., Seita, J., Inlay, M. A., Weissman, I. L. & Rossi, D. J. Quiescent hematopoietic stem cells accumulate DNA damage during aging that is repaired upon entry into cell cycle. *Cell Stem Cell* **15**, 37-50, doi:10.1016/j.stem.2014.04.016 (2014).

45 Flach, J. *et al.* Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. *Nature* **512**, 198-202, doi:10.1038/nature13619 (2014).

46 Moehrle, B. M. *et al.* Stem Cell-Specific Mechanisms Ensure Genomic Fidelity within HSCs and upon Aging of HSCs. *Cell reports* **13**, 2412-2424, doi:10.1016/j.celrep.2015.11.030 (2015).

47 Rübe, C. E. *et al.* Accumulation of DNA Damage in Hematopoietic Stem and Progenitor Cells during Human Aging. *PLOS ONE* **6**, e17487, doi:10.1371/journal.pone.0017487 (2011).

48 Collins, A. R. The comet assay for DNA damage and repair: principles, applications, and limitations. *Molecular biotechnology* **26**, 249-261, doi:10.1385/mb:26:3:249 (2004).

49 Wang, J. *et al.* Per2 induction limits lymphoid-biased haematopoietic stem cells and lymphopoiesis in the context of DNA damage and ageing. *Nat Cell Biol* **18**, 480-490, doi:10.1038/ncb3342 http://www.nature.com/ncb/journal/v18/n5/abs/ncb3342.html - supplementary-information (2016).

50 Harman, D. Aging: a theory based on free radical and radiation chemistry. *Journal of gerontology* **11**, 298-300 (1956).

51 Ceradini, D. J. *et al.* Progenitor cell trafficking is regulated by hypoxic gradients through HIF-1 induction of SDF-1. *Nat Med* **10**, 858-864, doi:10.1038/nm1075 (2004).

52 Parmar, K., Mauch, P., Vergilio, J.-A., Sackstein, R. & Down, J. D. Distribution of hematopoietic stem cells in the bone marrow according to regional hypoxia. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 5431-5436, doi:10.1073/pnas.0701152104 (2007).

53 Ito, K. *et al.* Reactive oxygen species act through p38 MAPK to limit the lifespan of hematopoietic stem cells. *Nat Med* **12**, 446-451, doi:10.1038/nm1388 (2006).

54 Tothova, Z. *et al.* FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. *Cell* **128**, 325-339, doi:10.1016/j.cell.2007.01.003 (2007).

55 Ito, K. *et al.* Regulation of oxidative stress by ATM is required for self-renewal of haematopoietic stem cells. *Nature* **431**, 997-1002, doi:10.1038/nature02989 (2004).

56 Jang, Y. Y. & Sharkis, S. J. A low level of reactive oxygen species selects for primitive hematopoietic stem cells that may reside in the low-oxygenic niche. *Blood* **110**, 3056-3063, doi:10.1182/blood-2007-05-087759 (2007).

57 Ho, T. T. *et al.* Autophagy maintains the metabolism and function of young and old stem cells. *Nature* **543**, 205-210, doi:10.1038/nature21388 (2017).

58 Simsek, T. *et al.* The distinct metabolic profile of hematopoietic stem cells reflects their location in a hypoxic niche. *Cell Stem Cell* **7**, 380-390, doi:10.1016/j.stem.2010.07.011 (2010).

59 Lagadinou, E. D. *et al.* BCL-2 inhibition targets oxidative phosphorylation and selectively eradicates quiescent human leukemia stem cells. *Cell Stem Cell* **12**, 329-341, doi:10.1016/j.stem.2012.12.013 (2013).

60 Klimmeck, D. *et al.* Proteomic cornerstones of hematopoietic stem cell differentiation: distinct signatures of multipotent progenitors and myeloid committed cells. *Molecular & cellular proteomics : MCP* **11**, 286-302, doi:10.1074/mcp.M111.016790 (2012).

61 Zhang, J., Nuebel, E., Daley, G. Q., Koehler, C. M. & Teitell, M. A. Metabolic regulation in pluripotent stem cells during reprogramming and self-renewal. *Cell Stem Cell* **11**, 589-595, doi:10.1016/j.stem.2012.10.005 (2012).

62 Yu, W. M. *et al.* Metabolic regulation by the mitochondrial phosphatase PTPMT1 is required for hematopoietic stem cell differentiation. *Cell Stem Cell* **12**, 62-74, doi:10.1016/j.stem.2012.11.022 (2013).

63 Chen, C., Liu, Y., Liu, Y. & Zheng, P. mTOR regulation and therapeutic rejuvenation of aging hematopoietic stem cells. *Science signaling* **2**, ra75, doi:10.1126/scisignal.2000559 (2009).

64 Chen, C. *et al.* TSC-mTOR maintains quiescence and function of hematopoietic stem cells by repressing mitochondrial biogenesis and reactive oxygen species. *The Journal of experimental medicine* **205**, 2397-2408, doi:10.1084/jem.20081297 (2008).

65 Magee, J. A. *et al.* Temporal changes in PTEN and mTORC2 regulation of hematopoietic stem cell self-renewal and leukemia suppression. *Cell Stem Cell* **11**, 415-428, doi:10.1016/j.stem.2012.05.026 (2012).

66 Miharada, K. *et al.* Cripto regulates hematopoietic stem cells as a hypoxic-niche-related factor through cell surface receptor GRP78. *Cell Stem Cell* **9**, 330-344, doi:10.1016/j.stem.2011.07.016 (2011).

67 Luchsinger, L. L., de Almeida, M. J., Corrigan, D. J., Mumau, M. & Snoeck, H. W. Mitofusin 2 maintains haematopoietic stem cells with extensive lymphoid potential. *Nature* **529**, 528-531, doi:10.1038/nature16500 (2016).

68 Reya, T. *et al.* A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* **423**, 409-414, doi:10.1038/nature01593 (2003).

69 Staal, F. J. & Sen, J. M. The canonical Wnt signaling pathway plays an important role in lymphopoiesis and hematopoiesis. *European journal of immunology* **38**, 1788-1794, doi:10.1002/eji.200738118 (2008).

70 Malhotra, S. & Kincade, P. W. Wnt-related molecules and signaling pathway equilibrium in hematopoiesis. *Cell Stem Cell* **4**, 27-36, doi:10.1016/j.stem.2008.12.004 (2009).

71 Florian, M. C. *et al.* A canonical to non-canonical Wnt signalling switch in haematopoietic stem-cell ageing. *Nature* **503**, 392-396, doi:10.1038/nature12631 (2013).

72 Florian, M. C. *et al.* Cdc42 activity regulates hematopoietic stem cell aging and rejuvenation. *Cell Stem Cell* **10**, 520-530, doi:10.1016/j.stem.2012.04.007 (2012).

73 Dang, W. *et al.* Histone H4 lysine-16 acetylation regulates cellular lifespan. *Nature* **459**, 802-807, doi:10.1038/nature08085 (2009).

74 Yoshikawa, T. T. & Schmader, K. Herpes Zoster in Older Adults. *Clinical Infectious Diseases* **32**, 1481-1486, doi:10.1086/320169 (2001).

75 Louria, D. B., Sen, P., Sherer, C. B. & Farrer, W. E. Infections in older patients: a systematic clinical approach. *Geriatrics* **48**, 28-34 (1993).

76 Armstrong, G. L., Conn, L. A. & Pinner, R. W. Trends in infectious disease mortality in the united states during the 20th century. *Jama* **281**, 61-66, doi:10.1001/jama.281.1.61 (1999).

77 Parsons, H. K. & Dockrell, D. H. The burden of invasive pneumococcal disease and the potential for reduction by immunisation. *International journal of antimicrobial agents* **19**, 85-93 (2002).

78 Thompson, W. W. *et al.* Mortality associated with influenza and respiratory syncytial virus in the United States. *Jama* **289**, 179-186 (2003).

79 Chen, W. H. *et al.* Vaccination in the elderly: an immunological perspective. *Trends in immunology* **30**, 351-359, doi:10.1016/j.it.2009.05.002 (2009).

80 Steinmann, G. G., Klaus, B. & Muller-Hermelink, H. K. The involution of the ageing human thymic epithelium is independent of puberty. A morphometric study. *Scandinavian journal of immunology* **22**, 563-575 (1985).

81 Haynes, B. F., Sempowski, G. D., Wells, A. F. & Hale, L. P. The human thymus during aging. *Immunologic research* **22**, 253-261, doi:10.1385/IR:22:2-3:253 (2000).

82 Jamieson, B. D. *et al.* Generation of functional thymocytes in the human adult. *Immunity* **10**, 569-575 (1999).

83 Poulin, J.-F. *et al.* Direct Evidence for Thymic Function in Adult Humans. *The Journal of Experimental Medicine* **190**, 479-486 (1999).

84 Bains, I., Antia, R., Callard, R. & Yates, A. J. Quantifying the development of the peripheral naive CD4+ T-cell pool in humans. *Blood* **113**, 5480-5487, doi:10.1182/blood-2008-10-184184 (2009).

85 den Braber, I. *et al.* Maintenance of peripheral naive T cells is sustained by thymus output in mice but not humans. *Immunity* **36**, 288-297, doi:10.1016/j.immuni.2012.02.006 (2012).

86 Sauce, D. *et al.* Lymphopenia-driven homeostatic regulation of naive T cells in elderly and thymectomized young adults. *Journal of immunology (Baltimore, Md. : 1950)* **189**, 5541-5548, doi:10.4049/jimmunol.1201235 (2012).

87 Qi, Q. *et al.* Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci U S A* **111**, 13139-13144, doi:10.1073/pnas.1409155111 (2014).

88 Batliwalla, F., Monteiro, J., Serrano, D. & Gregersen, P. K. Oligoclonality of CD8+ T cells in health and disease: aging, infection, or immune regulation? *Human immunology* **48**, 68-76 (1996).

89 Henson, S. M., Riddell, N. E. & Akbar, A. N. Properties of end-stage human T cells defined by CD45RA re-expression. *Current opinion in immunology* **24**, 476-481, doi:10.1016/j.coi.2012.04.001 (2012).

90 Frasca, D. *et al.* Aging down-regulates the transcription factor E2A, activation-induced cytidine deaminase, and Ig class switch in human B cells. *Journal of immunology (Baltimore, Md. : 1950)* **180**, 5283-5290 (2008).

91 Gibson, K. L. *et al.* B-cell diversity decreases in old age and is correlated with poor health status. *Aging cell* **8**, 18-25, doi:10.1111/j.1474-9726.2008.00443.x (2009).

92 Wang, C. *et al.* Effects of aging, cytomegalovirus infection, and EBV infection on human B cell repertoires. *Journal of immunology (Baltimore, Md. : 1950)* **192**, 603-611, doi:10.4049/jimmunol.1301384 (2014).

93 Guralnik, J. M., Eisenstaedt, R. S., Ferrucci, L., Klein, H. G. & Woodman, R. C. Prevalence of anemia in persons 65 years and older in the United States: evidence for a high rate of unexplained anemia. *Blood* **104**, 2263-2268, doi:10.1182/blood-2004-05-1812 (2004).

94 Kikuchi, M., Inagaki, T. & Shinagawa, N. Five-year survival of older people with anemia: variation with hemoglobin concentration. *Journal of the American Geriatrics Society* **49**, 1226-1228 (2001).

95 Izaks, G. J., Westendorp, R. G. & Knook, D. L. The definition of anemia in older persons. *Jama* **281**, 1714-1717 (1999).

96 Nilsson-Ehle, H., Jagenburg, R., Landahl, S., Svanborg, A. & Westin, J. Haematological abnormalities and reference intervals in the elderly. A cross-sectional comparative study of three urban Swedish population samples aged 70, 75 and 81 years. *Acta medica Scandinavica* **224**, 595-604 (1988).

97 Valent, P. *et al.* Definitions and standards in the diagnosis and treatment of the myelodysplastic syndromes: Consensus statements and report from a working conference. *Leuk Res* **31**, 727-736, doi:10.1016/j.leukres.2006.11.009 (2007).

98 Malcovati, L. *et al.* Clinical significance of somatic mutation in unexplained blood cytopenia. *Blood* **129**, 3371-3378, doi:10.1182/blood-2017-01-763425 (2017).

99 Cargo, C. A. *et al.* Targeted sequencing identifies patients with preclinical MDS at high risk of disease progression. *Blood* **126**, 2362-2365, doi:10.1182/blood-2015-08-663237 (2015).

100 Genovese, G. *et al.* Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine* **371**, 2477-2487, doi:doi:10.1056/NEJMoa1409405 (2014).

101 Jaiswal, S. *et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine* **371**, 2488-2498, doi:doi:10.1056/NEJMoa1408617 (2014).

102 Jaiswal, S. *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *The New England journal of medicine* **377**, 111-121, doi:10.1056/NEJMoa1701719 (2017).

103 Busque, L. *et al.* Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat Genet* **44**, 1179-1181, doi:10.1038/ng.2413 (2012).

104 Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med*, doi:10.1038/nm.3733 (2014).

105 Shlush, L. I. Age-related clonal hematopoiesis. *Blood* **131**, 496-504, doi:10.1182/blood-2017-07-746453 (2018).

106 Zink, F. *et al.* Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* **130**, 742-752, doi:10.1182/blood-2017-02-769869 (2017).

107 Yamamoto, J. F. & Goodman, M. T. Patterns of leukemia incidence in the United States by subtype and demographic characteristics, 1997-2002. *Cancer causes & control : CCC* **19**, 379-390, doi:10.1007/s10552-007-9097-2 (2008).

108 Institute, N. C. *SEER Cancer Stat Facts: Acute Myeloid Leukemia*, <http://seer.cancer.gov/statfacts/html/amyl.html> (

109 Dohner, H. *et al.* Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* **129**, 424-447, doi:10.1182/blood-2016-08-733196 (2017).

110 Dick, J. E. Acute myeloid leukemia stem cells. *Annals of the New York Academy of Sciences* **1044**, 1-5, doi:10.1196/annals.1349.001 (2005).

111 Rollison, D. E. *et al.* Epidemiology of myelodysplastic syndromes and chronic myeloproliferative disorders in the United States, 2001-2004, using data from the NAACCR and SEER programs. *Blood* **112**, 45-52, doi:10.1182/blood-2008-01-134858 (2008).

112 Malcovati, L. *et al.* Diagnosis and treatment of primary myelodysplastic syndromes in adults: recommendations from the European LeukemiaNet. *Blood* **122**, 2943 (2013).

113 Papaemmanuil, E. *et al.* Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**, 3616-3627, doi:10.1182/blood-2013-08-518886 (2013).

114 Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev* **23**, 781-783, doi:10.1101/gad.1787609 (2009).

115 Waddington, C. H. *The Epigenotpye*, 18-20 (1942).

116 Bloom, K. & Joglekar, A. Towards building a chromosome segregation machine. *Nature* **463**, 446-456, doi:10.1038/nature08912 (2010).

117 Oudet, P., Gross-Bellard, M. & Chambon, P. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* **4**, 281-300 (1975).

118 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260, doi:10.1038/38444 (1997).

119 Kornberg, R. D. & Thonmas, J. O. Chromatin Structure: Oligomers of the Histones. *Science* **184**, 865-868, doi:10.1126/science.184.4139.865 (1974).

120 Allan, J., Hartman, P. G., Crane-Robinson, C. & Aviles, F. X. The structure of histone H1 and its location in chromatin. *Nature* **288**, 675-679 (1980).

121 Sullivan, S. A., Aravind, L., Makalowska, I., Baxevanis, A. D. & Landsman, D. The Histone Database: a comprehensive WWW resource for histones and histone fold-containing proteins. *Nucleic acids research* **28**, 320-322, doi:10.1093/nar/28.1.320 (2000).

122 Buschbeck, M. & Hake, S. B. Variants of core histones and their roles in cell fate decisions, development and cancer. *Nature reviews. Molecular cell biology* **18**, 299-314, doi:10.1038/nrm.2016.166 (2017).

123 Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948-951, doi:10.1038/nature06947 (2008).

124 Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* **38**, 603-613, doi:10.1016/j.molcel.2010.03.016 (2010).

125 Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A. & Feinberg, A. P. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat Genet* **41**, 246-250, doi:10.1038/ng.297 (2009).

126 Cremer, T. *et al.* Chromosome territories--a functional nuclear landscape. *Current opinion in cell biology* **18**, 307-316, doi:10.1016/j.ceb.2006.04.007 (2006).

127 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).

128 Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol Cell* **48**, 471-484, doi:10.1016/j.molcel.2012.08.031 (2012).

129 Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385, doi:10.1038/nature11049 (2012).

130 Lin, H., Su, X. & He, B. Protein lysine acylation and cysteine succination by intermediates of energy metabolism. *ACS chemical biology* **7**, 947-960, doi:10.1021/cb3001793 (2012).

131 Chen, Y. *et al.* Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Molecular & cellular proteomics : MCP* **6**, 812-819, doi:10.1074/mcp.M700021-MCP200 (2007).

132 Zhang, K., Chen, Y., Zhang, Z. & Zhao, Y. Identification and verification of lysine propionylation and butyrylation in yeast core histones using PTMap software. *Journal of proteome research* **8**, 900-906, doi:10.1021/pr8005155 (2009).

133 Liu, B. *et al.* Identification and characterization of propionylation at histone H3 lysine 23 in mammalian cells. *J Biol Chem* **284**, 32288-32295, doi:10.1074/jbc.M109.045856 (2009).

134 Tan, M. *et al.* Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146**, 1016-1028, doi:10.1016/j.cell.2011.08.008 (2011).

135 Xie, Z. *et al.* Lysine succinylation and lysine malonylation in histones. *Molecular & cellular proteomics : MCP* **11**, 100-107, doi:10.1074/mcp.M111.015875 (2012).

136 Allfrey, V. G., Faulkner, R. & Mirsky, A. E. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proc Natl Acad Sci U S A* **51**, 786-794 (1964).

137 Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705, doi:10.1016/j.cell.2007.02.005 (2007).

138 Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074-1080, doi:10.1126/science.1063127 (2001).

139 Hodawadekar, S. C. & Marmorstein, R. Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. *Oncogene* **26**, 5528-5540, doi:10.1038/sj.onc.1210619 (2007).

140 Gregoretti, I. V., Lee, Y. M. & Goodson, H. V. Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. *Journal of molecular biology* **338**, 17-31, doi:10.1016/j.jmb.2004.02.006 (2004).

141 You, A., Tong, J. K., Grozinger, C. M. & Schreiber, S. L. CoREST is an integral component of the CoREST- human histone deacetylase complex. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 1454-1458 (2001).

142 Nagy, Z. & Tora, L. Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene* **26**, 5341-5357, doi:10.1038/sj.onc.1210604 (2007).

143 Kurdistani, S. K., Tavazoie, S. & Grunstein, M. Mapping global histone acetylation patterns to gene expression. *Cell* **117**, 721-733, doi:10.1016/j.cell.2004.05.023 (2004).

144 Schubeler, D. *et al.* The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev* **18**, 1263-1271, doi:10.1101/gad.1198204 (2004).

145 Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169-181, doi:10.1016/j.cell.2005.01.001 (2005).

146 Roh, T. Y., Cuddapah, S. & Zhao, K. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev* **19**, 542-552, doi:10.1101/gad.1272505 (2005).

147 Gates, L. A. *et al.* Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. *J Biol Chem* **292**, 14456-14472, doi:10.1074/jbc.M117.802074 (2017).

148 Morris, S. A. *et al.* Identification of histone H3 lysine 36 acetylation as a highly conserved histone modification. *J Biol Chem* **282**, 7632-7640, doi:10.1074/jbc.M607909200 (2007).

149 Li, Q. *et al.* Acetylation of Histone H3 Lysine 56 Regulates Replication-Coupled Nucleosome Assembly. *Cell* **134**, 244-255, doi:10.1016/j.cell.2008.06.018 (2008).

150 Li, N. *et al.* ZMYND8 Reads the Dual Histone Mark H3K4me1-H3K14ac to Antagonize the Expression of Metastasis-Linked Genes. *Mol Cell* **63**, 470-484, doi:10.1016/j.molcel.2016.06.035 (2016).

151 Jurkowska, R. Z. *et al.* H3K14ac is linked to methylation of H3K9 by the triple Tudor domain of SETDB1. *Nature communications* **8**, 2057, doi:10.1038/s41467-017-02259-9 (2017).

152 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

153 Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593-599, doi:10.1038/35020506 (2000).

154 Dillon, S. C., Zhang, X., Trievel, R. C. & Cheng, X. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome biology* **6**, 227, doi:10.1186/gb-2005-6-8-227 (2005).

155 van Leeuwen, F., Gafken, P. R. & Gottschling, D. E. Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* **109**, 745-756 (2002).

156 Feng, Q. *et al.* Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Current biology : CB* **12**, 1052-1058 (2002).

157 Jones, B. *et al.* The Histone H3K79 Methyltransferase Dot1L Is Essential for Mammalian Development and Heterochromatin Structure. *PLoS genetics* **4**, e1000190, doi:10.1371/journal.pgen.1000190 (2008).

158 Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941-953, doi:10.1016/j.cell.2004.12.012 (2004).

159 Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811-816, doi:10.1038/nature04433 (2006).

160 Whetstine, J. R. *et al.* Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**, 467-481, doi:10.1016/j.cell.2006.03.028 (2006).

161 Shen, H. *et al.* Suppression of Enhancer Overactivation by a RACK7-Histone Demethylase Complex. *Cell* **165**, 331-342, doi:10.1016/j.cell.2016.02.064 (2016).

162 Metzger, E. *et al.* LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**, 436-439, doi:10.1038/nature04020 (2005).

163 Shi, Y. J. *et al.* Regulation of LSD1 histone demethylase activity by its associated factors. *Mol Cell* **19**, 857-864, doi:10.1016/j.molcel.2005.08.027 (2005).

164 Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes & development* **16**, 2893-2905, doi:10.1101/gad.1035902 (2002).

165 Rao, R. C. & Dou, Y. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat Rev Cancer* **15**, 334-346, doi:10.1038/nrc3929 (2015).

166 Trojer, P. & Reinberg, D. Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell* **28**, 1-13, doi:10.1016/j.molcel.2007.09.011 (2007).

167 Ezhkova, E. *et al.* Ezh2 Orchestrates Gene Expression for the Stepwise Differentiation of Tissue-Specific Stem Cells. *Cell* **136**, 1122-1135, doi:10.1016/j.cell.2008.12.043.

168 Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110-113, doi:10.1126/science.1060118 (2001).

169 Martens, J. H. A. *et al.* The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *The EMBO journal* **24**, 800-812, doi:10.1038/sj.emboj.7600545 (2005).

170 Ng, H. H., Robert, F., Young, R. A. & Struhl, K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* **11**, 709-719 (2003).

171 Bernstein, B. E. *et al.* Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc Natl Acad Sci U S A* **99**, 8695-8700, doi:10.1073/pnas.082249499 (2002).

172 Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407-411, doi:10.1038/nature01080 (2002).

173 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* **107**, doi:10.1073/pnas.1016071107 (2010).

174 Wagner, E. J. & Carpenter, P. B. Understanding the language of Lys36 methylation at histone H3. *Nature reviews. Molecular cell biology* **13**, 115-126, doi:10.1038/nrm3274 (2012).

175 Musselman, C. A., Khorasanizadeh, S. & Kutateladze, T. G. Towards understanding methyllysine readout. *Biochimica et biophysica acta* **1839**, 686-693, doi:10.1016/j.bbagrm.2014.04.001 (2014).

176 Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491-496, doi:10.1038/20974 (1999).

177 Jacobson, R. H., Ladurner, A. G., King, D. S. & Tjian, R. Structure and function of a human TAFII250 double bromodomain module. *Science* **288**, 1422-1425 (2000).

178 Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120-124, doi:10.1038/35065138 (2001).

179 Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* **17**, 1870-1881, doi:10.1101/gad.1110503 (2003).

180 Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116-120, doi:10.1038/35065132 (2001).

181 Min, J., Zhang, Y. & Xu, R. M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* **17**, 1823-1828, doi:10.1101/gad.269603 (2003).

182 Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R., 3rd & Grant, P. A. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* **433**, 434-438, doi:10.1038/nature03242 (2005).

183 Huyen, Y. *et al.* Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* **432**, 406-411, doi:10.1038/nature03114 (2004).

184 Sanders, S. L. *et al.* Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. *Cell* **119**, 603-614, doi:10.1016/j.cell.2004.11.009 (2004).

185 Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859-872, doi:10.1016/j.cell.2005.03.036 (2005).

186 Yun, M., Wu, J., Workman, J. L. & Li, B. Readers of histone modifications. *Cell Res* **21**, 564-578, doi:10.1038/cr.2011.42 (2011).

187 Vermeulen, M. *et al.* Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58-69, doi:10.1016/j.cell.2007.08.016 (2007).

188 Lauberth, S. M. *et al.* H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* **152**, 1021-1036, doi:10.1016/j.cell.2013.01.052 (2013).

189 Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326, doi:10.1016/j.cell.2006.02.041 (2006).

190 Roh, T. Y., Cuddapah, S., Cui, K. & Zhao, K. The genomic landscape of histone modifications in human T cells. *Proc Natl Acad Sci U S A* **103**, 15782-15787, doi:10.1073/pnas.0607617103 (2006).

191 Pan, G. *et al.* Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* **1**, 299-312, doi:10.1016/j.stem.2007.08.003 (2007).

192 Alder, O. *et al.* Ring1B and Suv39h1 delineate distinct chromatin states at bivalent genes during early mouse lineage commitment. *Development (Cambridge, England)* **137**, 2483-2492, doi:10.1242/dev.048363 (2010).

193 Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* **8**, 532-538, doi:10.1038/ncb1403 (2006).

194 Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-353, doi:10.1038/nature04733 (2006).

195 Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M. & Helin, K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Molecular and cellular biology* **27**, 3769-3779, doi:10.1128/mcb.01432-06 (2007).

196 Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev* **24**, 265-276, doi:10.1101/gad.544410 (2010).

197 Chamberlain, S. J., Yee, D. & Magnuson, T. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem cells (Dayton, Ohio)* **26**, 1496-1505, doi:10.1634/stemcells.2008-0102 (2008).

198 Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell* **32**, 491-502, doi:10.1016/j.molcel.2008.10.016 (2008).

199 McGuire, E. A., Rintoul, C. E., Sclar, G. M. & Korsmeyer, S. J. Thymic overexpression of Ttg-1 in transgenic mice results in T-cell acute lymphoblastic leukemia/lymphoma. *Molecular and Cellular Biology* **12**, 4186-4196 (1992).

200 Oram, S. H. *et al.* Bivalent promoter marks and a latent enhancer may prime the leukaemia oncogene LMO1 for ectopic expression in T-cell leukaemia. *Leukemia* **27**, 1348-1357, doi:10.1038/leu.2013.2 (2013).

201 Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299-308 (1981).

202 Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human molecular genetics* **12**, 1725-1735, doi:10.1093/hmg/ddg180 (2003).

203 Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. & Shiroishi, T. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development (Cambridge, England)* **132**, 797-803, doi:10.1242/dev.01613 (2005).

204 Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283, doi:10.1038/nature09692 (2011).

205 Zentner, G. E., Tesar, P. J. & Scacheri, P. C. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res* **21**, 1273-1283, doi:10.1101/gr.122382.111 (2011).

206 Ghisletti, S. *et al.* Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity* **32**, 317-328, doi:10.1016/j.immuni.2010.02.008 (2010).

207 Heintzman, N. D. *et al.* Histone Modifications at Human Enhancers Reflect Global Cell Type-Specific Gene Expression. *Nature* **459**, 108-112, doi:10.1038/nature07829 (2009).

208 Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854-858, doi:10.1038/nature07730 (2009).

209 Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**, 825-837, doi:10.1016/j.molcel.2013.01.038 (2013).

210 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).

211 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).

212 Pott, S. & Lieb, J. D. What are super-enhancers? *Nat Genet* **47**, 8-12, doi:10.1038/ng.3167 (2015).

213 De Santa, F. *et al.* A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers. *PLoS biology* **8**, e1000384, doi:10.1371/journal.pbio.1000384 (2010).

214 Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, doi:10.1038/nature10006 (2011).

215 Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-187, doi:10.1038/nature09033 (2010).

216 Li, W., Notani, D. & Rosenfeld, M. G. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nature reviews. Genetics*, doi:10.1038/nrg.2016.4 (2016).

217 Arner, E. *et al.* Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* **347**, 1010-1014, doi:10.1126/science.1259418 (2015).

218 Alvarez-Dominguez, J. R. *et al.* Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. *Blood* **123**, 570-581, doi:10.1182/blood-2013-10-530683 (2014).

219 Hsieh, C. L. *et al.* Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc Natl Acad Sci U S A* **111**, 7319-7324, doi:10.1073/pnas.1324151111 (2014).

220 Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497-501, doi:10.1038/nature11884 (2013).

221 Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516-520, doi:10.1038/nature12210 (2013).

222 Melo, C. A. *et al.* eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol Cell* **49**, 524-535, doi:10.1016/j.molcel.2012.11.021 (2013).

223 Hatzis, P. & Talianidis, I. Dynamics of enhancer-promoter communication during differentiation-induced gene activation. *Mol Cell* **10**, 1467-1477 (2002).

224 Wang, Q., Carroll, J. S. & Brown, M. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol Cell* **19**, 631-642, doi:10.1016/j.molcel.2005.07.018 (2005).

225 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

226 Cirillo, L. A. *et al.* Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* **9**, 279-289 (2002).

227 Mullen, A. C. *et al.* Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell* **147**, 565-576, doi:10.1016/j.cell.2011.08.050 (2011).

228 Trompouki, E. *et al.* Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* **147**, 577-589, doi:10.1016/j.cell.2011.09.044 (2011).

229 Hengartner, C. J. *et al.* Association of an activator with an RNA polymerase II holoenzyme. *Genes Dev* **9**, 897-910 (1995).

230 Ito, M. *et al.* Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol Cell* **3**, 361-370 (1999).

231 Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435, doi:10.1038/nature09380 (2010).

232 Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**, 1573-1588.e1528, doi:10.1016/j.cell.2017.11.008.

233 Begley, C. G. *et al.* Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript. *Proc Natl Acad Sci U S A* **86**, 2031-2035 (1989).

234 Wang, J. *et al.* The t(14;21)(q11.2;q22) chromosomal translocation associated with T-cell acute lymphoblastic leukemia activates the BHLHB1 gene. *Proceedings of the National Academy of Sciences* **97**, 3497-3502, doi:10.1073/pnas.97.7.3497 (2000).

235 Groschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* **157**, 369-381, doi:10.1016/j.cell.2014.02.019 (2014).

236 Yamazaki, H. *et al.* A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. *Cancer Cell* **25**, 415-427, doi:10.1016/j.ccr.2014.02.008 (2014).

237 Herranz, D. *et al.* A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat Med* **20**, 1130-1137, doi:10.1038/nm.3665 (2014).

238 Mansour, M. R. *et al.* Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373-1377, doi:10.1126/science.1259037 (2014).

239 He, Y. F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307, doi:10.1126/science.1210944 (2011).

240 Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-1303, doi:10.1126/science.1210597 (2011).

241 Neri, F. *et al.* Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell reports*, doi:10.1016/j.celrep.2015.01.008 (2015).

242 Zhu, C. *et al.* Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution. *Cell Stem Cell* **20**, 720-731.e725, doi:https://doi.org/10.1016/j.stem.2017.02.013 (2017).

243 Wyatt, G. R. Occurrence of 5-methylcytosine in nucleic acids. *Nature* **166**, 237-238 (1950).

244 Doerks, T., Copley, R. R., Schultz, J., Ponting, C. P. & Bork, P. Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res* **12**, 47-56, doi:10.1101/ (2002).

245 Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature reviews. Genetics* **13**, 484-492, doi:10.1038/nrg3230 (2012).

246 Gruenbaum, Y., Stein, R., Cedar, H. & Razin, A. Methylation of CpG sequences in eukaryotic DNA. *FEBS letters* **124**, 67-71 (1981).

247 Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322, doi:10.1038/nature08514 (2009).

248 Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477-481, doi:10.1038/nature12433 (2013).

249 Tykocinski, M. L. & Max, E. E. CG dinucleotide clusters in MHC genes and in 5' demethylated genes. *Nucleic Acids Res* **12**, 4385-4396 (1984).

250 Bird, A., Taggart, M., Frommer, M., Miller, O. J. & Macleod, D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40**, 91-99 (1985).

251 Larsen, F., Gundersen, G., Lopez, R. & Prydz, H. CpG islands as gene markers in the human genome. *Genomics* **13**, 1095-1107 (1992).

252 Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* **41**, 178-186, doi:10.1038/ng.298 (2009).

253 Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490-495, doi:10.1038/nature10716 (2011).

254 Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134-1148, doi:10.1016/j.cell.2013.04.022 (2013).

255 Jeong, M. *et al.* Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat Genet* **46**, 17-23, doi:10.1038/ng.2836 (2014).

256 Hermann, A., Goyal, R. & Jeltsch, A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem* **279**, 48350-48359, doi:10.1074/jbc.M403427200 (2004).

257 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).

258 Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* **19**, 219-220, doi:10.1038/890 (1998).

259 Okano, M., Xie, S. & Li, E. Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic acids research* **26**, 2536-2540, doi:10.1093/nar/26.11.2536 (1998).

260 Yoder, J. A. & Bestor, T. H. A Candidate Mammalian DNA Methyltransferase Related to pmt1p of Fission Yeast. *Human molecular genetics* **7**, 279-284, doi:10.1093/hmg/7.2.279 (1998).

261 Goll, M. G. *et al.* Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science* **311**, 395-398, doi:10.1126/science.1120976 (2006).

262 Chédin, F., Lieber, M. R. & Hsieh, C.-L. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proceedings of the National Academy of Sciences* **99**, 16916-16921, doi:10.1073/pnas.262443999 (2002).

263 Hata, K., Okano, M., Lei, H. & Li, E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development (Cambridge, England)* **129**, 1983-1993 (2002).

264 Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472-479, doi:10.1038/nature12750 (2013).

265 Cortazar, D. *et al.* Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419-423, doi:10.1038/nature09672 (2011).

266 Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67-79, doi:10.1016/j.cell.2011.06.020 (2011).

267 Maiti, A. & Drohat, A. C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* **286**, 35334-35338, doi:10.1074/jbc.C111.284620 (2011).

268 Tate, P. H. & Bird, A. P. Effects of DNA methylation on DNA-binding proteins and gene expression. *Current Opinion in Genetics & Development* **3**, 226-231, doi:https://doi.org/10.1016/0959-437X(93)90027-M (1993).

269 Jones, P. A. & Laird, P. W. Cancer epigenetics comes of age. *Nat Genet* **21**, 163-167, doi:10.1038/5947 (1999).

270 Spencer, D. H. *et al.* CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression. *Cell* **168**, 801-816 e813, doi:10.1016/j.cell.2017.01.021 (2017).

271 Farlik, M. *et al.* DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell Stem Cell* **19**, 808-822, doi:10.1016/j.stem.2016.10.019 (2016).

272 Bock, C. *et al.* DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Mol Cell* **47**, 633-647, doi:10.1016/j.molcel.2012.06.019 (2012).

273 Baylin, S. B., Herman, J. G., Graff, J. R., Vertino, P. M. & Issa, J. P. Alterations in DNA methylation: a fundamental aspect of neoplasia. *Advances in cancer research* **72**, 141-196 (1998).

274 Herman, J. G. *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A* **91**, 9700-9704 (1994).

275 Wolf, S. F., Jolly, D. J., Lunnen, K. D., Friedmann, T. & Migeon, B. R. Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. *Proc Natl Acad Sci U S A* **81**, 2806-2810 (1984).

276 Hellman, A. & Chess, A. Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141-1143, doi:10.1126/science.1136352 (2007).

277 Dhayalan, A. *et al.* The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem* **285**, 26114-26120, doi:10.1074/jbc.M109.089433 (2010).

278 Baubec, T. *et al.* Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243-247, doi:10.1038/nature14176 (2015).

279 Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72-77, doi:10.1038/nature21373 (2017).

280 Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562, doi:10.1038/nature01262 (2002).

281 Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends in genetics : TIG* **13**, 335-340 (1997).

282 Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**, 187-191, doi:10.1038/561 (1998).

283 Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386-389, doi:10.1038/30764 (1998).

284 Wyatt, G. R. & Cohen, S. S. The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. *The Biochemical journal* **55**, 774-782 (1953).

285 Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935, doi:10.1126/science.1170116 (2009).

286 Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8**, 1698-1710, doi:10.4161/cc.8.11.8580 (2009).

287 Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129-1133, doi:10.1038/nature09303 (2010).

288 Bachman, M. *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nature chemistry* **6**, 1049-1055, doi:10.1038/nchem.2064 (2014).

289 Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929-930, doi:10.1126/science.1169786 (2009).

290 Madzo, J. *et al.* Hydroxymethylation at gene regulatory regions directs stem/early progenitor cell commitment during erythropoiesis. *Cell reports* **6**, 231-244, doi:10.1016/j.celrep.2013.11.044 (2014).

291 Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S. & Jacobsen, S. E. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome biology* **12**, R54, doi:10.1186/gb-2011-12-6-r54 (2011).

292 Hon, G. C. *et al.* 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol Cell* **56**, 286-297, doi:10.1016/j.molcel.2014.08.026 (2014).

293 Akalin, A. *et al.* Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet* **8**, e1002781, doi:10.1371/journal.pgen.1002781 (2012).

294 Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553-567, doi:10.1016/j.ccr.2010.11.015 (2010).

295 Figueroa, M. E. *et al.* DNA Methylation Signatures Identify Biologically Distinct Subtypes in Acute Myeloid Leukemia. *Cancer cell* **17**, 13-27, doi:10.1016/j.ccr.2009.11.020 (2010).

296 Figueroa, M. E. *et al.* MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation. *Blood* **114**, 3448-3458, doi:10.1182/blood-2009-01-200519 (2009).

297 Glass, J. L. *et al.* Epigenetic Identity in AML Depends on Disruption of Nonpromoter Regulatory Elements and Is Affected by Antagonistic Effects of

Mutations in Epigenetic Modifiers. *Cancer discovery* **7**, 868-883, doi:10.1158/2159-8290.cd-16-1032 (2017).

298 Walter, M. J. *et al.* Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia* **25**, 1153-1158, doi:10.1038/leu.2011.44 (2011).

299 Patel, J. P. *et al.* Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *The New England journal of medicine* **366**, 1079-1089, doi:10.1056/NEJMoa1112304 (2012).

300 Ley, T. J. *et al.* DNMT3A mutations in acute myeloid leukemia. *The New England journal of medicine* **363**, 2424-2433, doi:10.1056/NEJMoa1005143 (2010).

301 Ribeiro, A. F. T. *et al.* Mutant DNMT3A: a marker of poor prognosis in acute myeloid leukemia. *Blood* **119**, 5824-5831, doi:10.1182/blood-2011-07-367961 (2012).

302 Kim, S. J. *et al.* A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. *Blood* **122**, 4086-4089, doi:10.1182/blood-2013-02-483487 (2013).

303 Russler-Germain, D. A. *et al.* The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer Cell* **25**, 442-454, doi:10.1016/j.ccr.2014.02.010 (2014).

304 Celik, H. *et al.* Enforced differentiation of Dnmt3a-null bone marrow leads to failure with c-Kit mutations driving leukemic transformation. *Blood* **125**, 619-628, doi:10.1182/blood-2014-08-594564 (2015).

305 Mayle, A. *et al.* Dnmt3a loss predisposes murine hematopoietic stem cells to malignant transformation. *Blood* **125**, 629-638, doi:10.1182/blood-2014-08-594648 (2015).

306 Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506-510, doi:10.1038/nature10738 (2012).

307 Krönke, J. *et al.* Clonal evolution in relapsed NPM1-mutated acute myeloid leukemia. *Blood* **122**, 100-108, doi:10.1182/blood-2013-01-479188 (2013).

308 Liu, W.-J. *et al.* Prognostic significance of Tet methylcytosine dioxygenase 2 (TET2) gene mutations in adult patients with acute myeloid leukemia: a meta-analysis. *Leukemia & lymphoma* **55**, 2691-2698 (2014).

309 Weissmann, S. *et al.* Landscape of TET2 mutations in acute myeloid leukemia. *Leukemia* **26**, 934-942, doi:10.1038/leu.2011.326 (2012).

310 Delhommeau, F. *et al.* Mutation in TET2 in myeloid cancers. *The New England journal of medicine* **360**, 2289-2301, doi:10.1056/NEJMoa0810069 (2009).

311 Langemeijer, S. M. *et al.* Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet* **41**, 838-842, doi:10.1038/ng.391 (2009).

312 Tefferi, A. *et al.* Detection of mutant TET2 in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN and AML. *Leukemia* **23**, 1343-1345, doi:10.1038/leu.2009.59 (2009).

313 Grossmann, V. *et al.* Molecular profiling of chronic myelomonocytic leukemia reveals diverse mutations in >80% of patients with TET2 and EZH2 being of high prognostic relevance. *Leukemia* **25**, 877-879, doi:10.1038/leu.2011.10 (2011).

314 Smith, A. E. *et al.* Next-generation sequencing of the TET2 gene in 355 MDS and CMML patients reveals low-abundance mutant clones with early origins, but indicates no definite prognostic value. *Blood* **116**, 3923-3932, doi:10.1182/blood-2010-03-274704 (2010).

315 Konstandin, N. *et al.* Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia* **25**, 1649, doi:10.1038/leu.2011.134 https://www.nature.com/articles/leu2011134 - supplementary-information (2011).

316 Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839-843, doi:10.1038/nature09586 (2010).

317 Rampal, R. *et al.* DNA Hydroxymethylation Profiling Reveals that WT1 Mutations Result in Loss of TET2 Function in Acute Myeloid Leukemia. *Cell reports* **9**, 1841-1855, doi:10.1016/j.celrep.2014.11.004 (2014).

318 Kroeze, L. I. *et al.* Characterization of acute myeloid leukemia based on levels of global hydroxymethylation. *Blood* **124**, 1110-1118, doi:10.1182/blood-2013-08-518514 (2014).

319 Marcucci, G. *et al.* IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**, 2348-2355, doi:10.1200/jco.2009.27.3730 (2010).

320 Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *The New England journal of medicine* **361**, 1058-1066, doi:10.1056/NEJMoa0903840 (2009).

321 Paschka, P. *et al.* IDH1 and IDH2 mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with NPM1 mutation without FLT3 internal tandem duplication. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**, 3636-3643, doi:10.1200/jco.2010.28.3762 (2010).

322 Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739-744, doi:10.1038/nature08617 (2009).

323 Ward, P. S. *et al.* The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**, 225-234, doi:10.1016/j.ccr.2010.01.020 (2010).

324 Kantarjian, H. *et al.* Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study. *Cancer* **106**, 1794-1803, doi:10.1002/cncr.21792 (2006).

325 Silverman, L. R. *et al.* Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **20**, 2429-2440, doi:10.1200/jco.2002.04.117 (2002).

326 Brocks, D. *et al.* DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat Genet* **49**, 1052-1060, doi:10.1038/ng.3889 (2017).

327 Chiappinelli, K. B. *et al.* Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* **162**, 974-986, doi:10.1016/j.cell.2015.07.011 (2015).

328 Roulois, D. *et al.* DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell* **162**, 961-973, doi:10.1016/j.cell.2015.07.056 (2015).

329 Popovici-Muller, J. *et al.* Discovery of the First Potent Inhibitors of Mutant IDH1 That Lower Tumor 2-HG in Vivo. *ACS medicinal chemistry letters* **3**, 850-855, doi:10.1021/ml300225h (2012).

330 Stein, E. M. *et al.* Enasidenib in mutant IDH2 relapsed or refractory acute myeloid leukemia. *Blood* **130**, 722-731, doi:10.1182/blood-2017-04-779405 (2017).

331 Feser, J. *et al.* Elevated histone expression promotes life span extension. *Mol Cell* **39**, 724-735, doi:10.1016/j.molcel.2010.08.015 (2010).

332 Hu, Z. *et al.* Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging. *Genes Dev* **28**, 396-408, doi:10.1101/gad.233221.113 (2014).

333 O'Sullivan, R. J., Kubicek, S., Schreiber, S. L. & Karlseder, J. Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres. *Nature structural & molecular biology* **17**, 1218-1225, doi:10.1038/nsmb.1897 (2010).

334 Polo, S. E., Roche, D. & Almouzni, G. New histone incorporation marks sites of UV repair in human cells. *Cell* **127**, 481-493, doi:10.1016/j.cell.2006.08.049 (2006).

335 Lyons, S. M. *et al.* A subset of replication-dependent histone mRNAs are expressed as polyadenylated RNAs in terminally differentiated tissues. *Nucleic Acids Res* **44**, 9190-9205, doi:10.1093/nar/gkw620 (2016).

336 Wen, H. *et al.* ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. *Nature* **508**, 263-268, doi:10.1038/nature13045 (2014).

337 Szenker, E., Ray-Gallet, D. & Almouzni, G. The double face of the histone variant H3.3. *Cell Res* **21**, 421-434, doi:10.1038/cr.2011.14 (2011).

338 Maze, I. *et al.* Critical Role of Histone Turnover in Neuronal Transcription and Plasticity. *Neuron* **87**, 77-94, doi:10.1016/j.neuron.2015.06.014 (2015).

339 Tvardovskiy, A., Schwammle, V., Kempf, S. J., Rogowska-Wrzesinska, A. & Jensen, O. N. Accumulation of histone variant H3.3 with age is associated with profound changes in the histone methylation landscape. *Nucleic Acids Res* **45**, 9272-9289, doi:10.1093/nar/gkx696 (2017).

340 Maures, T. J., Greer, E. L., Hauswirth, A. G. & Brunet, A. The H3K27 demethylase UTX-1 regulates C. elegans lifespan in a germline-independent, insulin-dependent manner. *Aging cell* **10**, 980-990, doi:10.1111/j.1474-9726.2011.00738.x (2011).

341 Greer, E. L. *et al.* Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in C. elegans. *Nature* **466**, 383-387, doi:10.1038/nature09195 (2010).

342 Alvares, S. M., Mayberry, G. A., Joyner, E. Y., Lakowski, B. & Ahmed, S. H3K4 demethylase activities repress proliferative and postmitotic aging. *Aging cell* **13**, 245-253, doi:10.1111/acel.12166 (2014).

343 Siebold, A. P. *et al.* Polycomb Repressive Complex 2 and Trithorax modulate Drosophila longevity and stress resistance. *Proc Natl Acad Sci U S A* **107**, 169-174, doi:10.1073/pnas.0907739107 (2010).

344 Li, L., Greer, C., Eisenman, R. N. & Secombe, J. Essential Functions of the Histone Demethylase Lid. *PLoS genetics* **6**, e1001221, doi:10.1371/journal.pgen.1001221 (2010).

345 Kanfi, Y. *et al.* The sirtuin SIRT6 regulates lifespan in male mice. *Nature* **483**, 218-221, doi:10.1038/nature10815 (2012).

346 Herranz, D. *et al.* Sirt1 improves healthy ageing and protects from metabolic syndrome-associated cancer. *Nature communications* **1**, 3, doi:10.1038/ncomms1001 (2010).

347 Wood, J. G. *et al.* Chromatin remodeling in the aging genome of Drosophila. *Aging cell* **9**, 971-978, doi:10.1111/j.1474-9726.2010.00624.x (2010).

348 Baumgart, M. *et al.* RNA-seq of the aging brain in the short-lived fish N. furzeri - conserved pathways and novel genes associated with neurogenesis. *Aging cell* **13**, 965-974, doi:10.1111/acel.12257 (2014).

349 Berdyshev, G. D., Korotaev, G. K., Boiarskikh, G. V. & Vaniushin, B. F. [Nucleotide composition of DNA and RNA from somatic tissues of humpback and its changes during spawning]. *Biokhimiia (Moscow, Russia)* **32**, 988-993 (1967).

350 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115, doi:10.1186/gb-2013-14-10-r115 (2013).

351 Vanyushin, B. F., Nemirovsky, L. E., Klimenko, V. V., Vasiliev, V. K. & Belozersky, A. N. The 5-methylcytosine in DNA of rats. Tissue and age specificity and the changes induced by hydrocortisone and other agents. *Gerontologia* **19**, 138-152 (1973).

352 Wilson, V. L., Smith, R. A., Ma, S. & Cutler, R. G. Genomic 5-methyldeoxycytidine decreases with age. *J Biol Chem* **262**, 9948-9951 (1987).

353 Chen, H., Dzitoyeva, S. & Manev, H. Effect of aging on 5-hydroxymethylcytosine in the mouse hippocampus. *Restorative neurology and neuroscience* **30**, 237-245, doi:10.3233/rnn-2012-110223 (2012).

354 Xiong, J. *et al.* DNA hydroxymethylation age of human blood determined by capillary hydrophilic-interaction liquid chromatography/mass spectrometry. *Clinical epigenetics* **7**, 72, doi:10.1186/s13148-015-0109-x (2015).

355 Bjornsson, H. T. *et al.* Intra-individual change over time in DNA methylation with familial clustering. *Jama* **299**, 2877-2883, doi:10.1001/jama.299.24.2877 (2008).

356 Rakyan, V. K. *et al.* Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* **20**, 434-439, doi:10.1101/gr.103101.109 (2010).

357 Teschendorff, A. E. *et al.* Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* **20**, 440-446, doi:10.1101/gr.103606.109 (2010).

358 Maegawa, S. *et al.* Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res* **20**, 332-340, doi:10.1101/gr.096826.109 (2010).

359 Armstrong, V. L., Rakoczy, S., Rojanathammanee, L. & Brown-Borg, H. M. Expression of DNA methyltransferases is influenced by growth hormone in the long-living Ames dwarf mouse in vivo and in vitro. *The journals of gerontology. Series A, Biological sciences and medical sciences* **69**, 923-933, doi:10.1093/gerona/glt133 (2014).

360 Day, K. *et al.* Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. *Genome biology* **14**, R102, doi:10.1186/gb-2013-14-9-r102 (2013).

361 Bollati, V. *et al.* Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mechanisms of ageing and development* **130**, 234-239, doi:10.1016/j.mad.2008.12.003 (2009).

362 McClay, J. L. *et al.* A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects. *Hum Mol Genet* **23**, 1175-1185, doi:10.1093/hmg/ddt511 (2014).

363 Truong, T. P. *et al.* Age-Dependent Decrease of DNA Hydroxymethylation in Human T Cells. *Journal of clinical and experimental hematopathology : JCEH* **55**, 1-6, doi:10.3960/jslrt.55.1 (2015).

364 Torano, E. G. *et al.* Age-associated hydroxymethylation in human bone-marrow mesenchymal stem cells. *Journal of translational medicine* **14**, 207, doi:10.1186/s12967-016-0966-x (2016).

365 Sun, D. *et al.* Epigenomic Profiling of Young and Aged HSCs Reveals Concerted Changes during Aging that Reinforce Self-Renewal. *Cell Stem Cell* **14**, 673-688, doi:10.1016/j.stem.2014.03.002 (2014).

366 Schworer, S. *et al.* Epigenetic stress responses induce muscle stem-cell ageing by Hoxa9 developmental signals. *Nature* **540**, 428-432, doi:10.1038/nature20603 (2016).

367 Liu, L. *et al.* Chromatin modifications as determinants of muscle stem cell quiescence and chronological aging. *Cell reports* **4**, 189-204, doi:10.1016/j.celrep.2013.05.043 (2013).

368 Zhang, W. *et al.* Aging stem cells. A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. *Science* **348**, 1160-1163, doi:10.1126/science.aaa1356 (2015).

369 Chen, Y. H. *et al.* Enhancer of Zeste Homolog 2 and Histone Deacetylase 9c Regulate Age-Dependent Mesenchymal Stem Cell Differentiation into Osteoblasts and Adipocytes. *Stem cells (Dayton, Ohio)* **34**, 2183-2193, doi:10.1002/stem.2400 (2016).

370 Yang, J. *et al.* Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Scientific reports* **5**, 15145, doi:10.1038/srep15145 (2015).

371 He, W., Goodkind, D. & Kowal, P. An Aging World: 2015. (2016).

372 Guralnik, J. M., Eisenstaedt, R. S., Ferrucci, L., Klein, H. G. & Woodman, R. C. Prevalence of anemia in persons 65 years and older in the United States: evidence for a high rate of unexplained anemia. *Blood* **104**, 2263-2268, doi:10.1182/blood-2004-05-1812 (2004).

373 Aul, C., Gattermann, N. & Schneider, W. Age-related incidence and other epidemiological aspects of myelodysplastic syndromes. *British journal of haematology* **82**, 358-367 (1992).

374 Montecino-Rodriguez, E., Berent-Maoz, B. & Dorshkind, K. Causes, consequences, and reversal of immune system aging. *The Journal of clinical investigation* **123**, 958-965, doi:10.1172/JCI64096 (2013).

375 Kwok, B. *et al.* MDS-associated somatic mutations and clonal hematopoiesis are common in idiopathic cytopenias of undetermined significance. *Blood* **126**, 2355-2361, doi:10.1182/blood-2015-08-667063 (2015).

376 Andrews, S. *FastQC: a quality control tool for high throughput sequence data*, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (2010).

377 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, doi:10.14806/ej.17.1.200
pp. 10-12 (2011).

378 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).

379 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

380 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, doi:10.1093/bioinformatics/btp352 (2009).

381 Warnes, G. R. *et al.* gplots: Various R programming tools for plotting data. *R package version* **2**, 1 (2009).

382 Welch, R. P. *et al.* ChIP-Enrich: gene set enrichment testing for ChIP-seq data. *Nucleic Acids Res* **42**, e105, doi:10.1093/nar/gku463 (2014).

383 Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**, 5116-5121, doi:10.1073/pnas.091062498 (2001).

384 Akalin, A. *et al.* Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigeneticlandscapes in acute myeloid leukemia. *PLoS Genet* **8**, doi:10.1371/journal.pgen.1002781 (2012).

385 Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572, doi:10.1093/bioinformatics/btr167 (2011).

386 Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome biology* **13**, doi:10.1186/gb-2012-13-10-r87 (2012).

387 R: A language and environment for statistical computing (Vienna, Austria, 2012).

388 Li, S. *et al.* An optimized algorithm for detecting and annotating regional differential methylation. *BMC bioinformatics* **14 Suppl 5**, S10, doi:10.1186/1471-2105-14-s5-s10 (2013).

389 Meldi, K. *et al.* Specific molecular signatures predict decitabine response in chronic myelomonocytic leukemia. *The Journal of clinical investigation* **125**, 1857-1872, doi:10.1172/jci78752 (2015).

390 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

391 Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13, doi:10.1093/nar/gkn923 (2009).

392 Culhane, A. C., Thioulouse, J., Perriere, G. & Higgins, D. G. MADE4: an R package for multivariate analysis of gene expression data. *Bioinformatics* **21**, 2789-2790, doi:10.1093/bioinformatics/bti394 (2005).

393 Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer, 2009).

394 Glass, J. *et al.* Epigenetic Identity in AML Depends on Disruption of Non-promoter Regulatory Elements and is Affected by Antagonistic Effects of Mutations in Epigenetic Modifiers. *Cancer discovery*, doi:10.1158/2159-8290.cd-16-1032 (2017).

395 factoextra: Extract and Visualize the Results of Multivariate Data Analyses v. 1.0.5.999 (2017).

396 Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849, doi:10.1093/bioinformatics/btw313 (2016).

397 Zhang, Y., Lin, Y. H., Johnson, T. D., Rozek, L. S. & Sartor, M. A. PePr: a peak-calling prioritization pipeline to identify consistent or differential peaks from replicated ChIP-Seq data. *Bioinformatics* **30**, 2568-2575, doi:10.1093/bioinformatics/btu372 (2014).

398 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

399 Hartley, S. W. & Mullikin, J. C. QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. *BMC bioinformatics* **16**, 224, doi:10.1186/s12859-015-0670-5 (2015).

400 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome biology* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).

401 Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

402 Crews, L. A. *et al.* RNA Splicing Modulation Selectively Impairs Leukemia Stem Cell Maintenance in Secondary Human AML. *Cell Stem Cell* **19**, 599-612, doi:10.1016/j.stem.2016.08.003 (2016).

403 Shen, S. *et al.* rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A* **111**, E5593-5601, doi:10.1073/pnas.1419161111 (2014).

404 Shirai, C. L. *et al.* Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome. *Nature communications* **8**, 14060, doi:10.1038/ncomms14060 (2017).

405 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).

406 Emig, D. *et al.* AltAnalyze and DomainGraph: analyzing and visualizing exon expression data. *Nucleic Acids Res* **38**, W755-762, doi:10.1093/nar/gkq405 (2010).

407 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

408 Akalin, A., Franke, V., Vlahovicek, K., Mason, C. E. & Schubeler, D. Genomation: a toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics* **31**, 1127-1129, doi:10.1093/bioinformatics/btu775 (2015).

409 Li, Q. *et al.* Oncogenic Nras has bimodal effects on stem cells that sustainably increase competitiveness. *Nature* **504**, 143-147, doi:10.1038/nature12830 (2013).

410 pwr: Basic Functions for Power Analysis v. R package version 1.2-2 (2018).

411 Beck, D. *et al.* Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. *Blood* **122**, e12-e22, doi:10.1182/blood-2013-03-490425 (2013).

412 Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* **44**, W160-W165, doi:10.1093/nar/gkw257 (2016).

413 Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284, doi:10.1186/1471-2164-15-284 (2014).

414 Lindsley, R. C. *et al.* Prognostic Mutations in Myelodysplastic Syndrome after Stem-Cell Transplantation. *The New England journal of medicine* **376**, 536-547, doi:10.1056/NEJMoa1611604 (2017).

415 Lippert, R. A. Space-efficient whole genome comparisons with Burrows-Wheeler transforms. *J Comput Biol* **12**, doi:10.1089/cmb.2005.12.407 (2005).

416 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

417 Wilm, A. *et al.* LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res* **40**, 11189-11201, doi:10.1093/nar/gks918 (2012).

418 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164, doi:10.1093/nar/gkq603 (2010).

419 Jian, X., Boerwinkle, E. & Liu, X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res* **42**, 13534-13544, doi:10.1093/nar/gku1206 (2014).

420 Abecasis, G. R. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073, doi:10.1038/nature09534 (2010).

421 Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* **39**, D945-950, doi:10.1093/nar/gkq929 (2011).

422 Liau, W. S. *et al.* Aberrant activation of the GIMAP enhancer by oncogenic transcription factors in T-cell acute lymphoblastic leukemia. *Leukemia* **31**, 1798-1807, doi:10.1038/leu.2016.392 (2017).

423 Loven, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-334, doi:10.1016/j.cell.2013.03.036 (2013).

424 Georgantas, R. W., 3rd *et al.* CD34+ hematopoietic stem-progenitor cell microRNA expression and function: a circuit diagram of differentiation control. *Proc Natl Acad Sci U S A* **104**, 2750-2755, doi:10.1073/pnas.0610983104 (2007).

425 Celik, H. *et al.* Jarid2 Restricts Long-Term Repopulating Stem Cell Capacity in Multipotent Progenitors and Acts As Tumor Suppressor in Chronic Myeloid Neoplasms. *Blood* **130**, 488-488 (2017).

426 Ma, X.-Y. *et al.* Malat1 as an evolutionarily conserved lncRNA, plays a positive role in regulating proliferation and maintaining undifferentiated status of early-stage hematopoietic cells. *BMC Genomics* **16**, 676, doi:10.1186/s12864-015-1881-x (2015).

427 Hock, H. *et al.* Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. *Genes & development* **18**, 2336-2341, doi:10.1101/gad.1239604 (2004).

428 Hahn, M. A. *et al.* Loss of the Polycomb mark from bivalent promoters leads to activation of cancer-promoting genes in colorectal tumors. *Cancer research* **74**, 3617-3629, doi:10.1158/0008-5472.CAN-13-3147 (2014).

429 Bernhart, S. H. *et al.* Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer. *Scientific reports* **6**, 37393, doi:10.1038/srep37393 (2016).

430 Voigt, P., Tee, W. W. & Reinberg, D. A double take on bivalent promoters. *Genes Dev* **27**, 1318-1338, doi:10.1101/gad.219626.113 (2013).

431 Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560, doi:10.1038/nature06008 (2007).

432 Pang, W. W., Schrier, S. L. & Weissman, I. L. Age-associated changes in human hematopoietic stem cells. *Semin Hematol* **54**, 39-42, doi:10.1053/j.seminhematol.2016.10.004 (2017).

433 Naeim, F., Nagesh Rao, P., Song, S. X. & Grody, W. W. in *Atlas of Hematopathology* 201-212 (Academic Press, 2013).

434 Annangi, S. & Kota, V. Incidence and Survival Outcomes Of Myelodysplastic Syndrome In The United States: A SEER Analysis 2001 - 2010. *Blood* **122**, 5205-5205 (2013).

435 Shin, J. W. *et al.* Lamins regulate cell trafficking and lineage maturation of adult human hematopoietic cells. *Proc Natl Acad Sci U S A* **110**, 18892-18897, doi:10.1073/pnas.1304996110 (2013).

436 De Sandre-Giovannoli, A. *et al.* Lamin A Truncation in Hutchinson-Gilford Progeria. *Science* **300**, 2055-2055, doi:10.1126/science.1084125 (2003).

437 Eriksson, M. *et al.* Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. *Nature* **423**, 293-298, doi:10.1038/nature01629 (2003).

438 Beerman, I. *et al.* Proliferation-Dependent Alterations of the DNA Methylation Landscape Underlie Hematopoietic Stem Cell Aging. *Cell Stem Cell* **12**, 413-425, doi:http://dx.doi.org/10.1016/j.stem.2013.01.017 (2013).

439 Chouliaras, L. *et al.* Age-related increase in levels of 5-hydroxymethylcytosine in mouse hippocampus is prevented by caloric restriction. *Current Alzheimer research* **9**, 536-544 (2012).

440 Tammen, S. A. *et al.* Aging alters hepatic DNA hydroxymethylation, as measured by liquid chromatography/mass spectrometry. *Journal of cancer prevention* **19**, 301-308, doi:10.15430/jcp.2014.19.2.301 (2014).

441 Wen, L. & Tang, F. Genomic distribution and possible functions of DNA hydroxymethylation in the brain. *Genomics* **104**, 341-346, doi:10.1016/j.ygeno.2014.08.020 (2014).

442 Steiner, L. A., Schulz, V., Makismova, Y., Lezon-Geyda, K. & Gallagher, P. G. CTCF and CohesinSA-1 Mark Active Promoters and Boundaries of Repressive Chromatin Domains in Primary Human Erythroid Cells. *PLOS ONE* **11**, e0155378, doi:10.1371/journal.pone.0155378 (2016).

443 Juliusson, G. *et al.* Age and acute myeloid leukemia: real world data on decision to treat and outcomes from the Swedish Acute Leukemia Registry. *Blood* **113**, 4179-4187, doi:10.1182/blood-2008-07-172007 (2009).

444 Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function,and impair hematopoietic differentiation. *Cancer Cell* **18**, doi:10.1016/j.ccr.2010.11.015 (2010).

445 Figueroa, M. E. *et al.* DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17**, doi:10.1016/j.ccr.2009.11.020 (2010).

446 Gazit, R., Weissman, I. L. & Rossi, D. J. Hematopoietic stem cells and the aging hematopoietic system. *Semin Hematol* **45**, 218-224, doi:10.1053/j.seminhematol.2008.07.010 (2008).

447 Farlik, M. *et al.* DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell Stem Cell* **19**, 808-822, doi:10.1016/j.stem.2016.10.019.

448 Moran-Crusio, K. *et al.* Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20**, 11-24, doi:10.1016/j.ccr.2011.06.001 (2011).

449 Kim, G. D. *et al.* Kruppel-like Factor 6 Promotes Macrophage-mediated Inflammation by Suppressing B Cell Leukemia/Lymphoma 6 Expression. *J Biol Chem* **291**, 21271-21282, doi:10.1074/jbc.M116.738617 (2016).

450 Merideth, M. A. *et al.* Phenotype and Course of Hutchinson–Gilford Progeria Syndrome. *The New England journal of medicine* **358**, 592-604, doi:10.1056/NEJMoa0706898 (2008).

451 Dechat, T. *et al.* Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin. *Genes Dev* **22**, 832-853, doi:10.1101/gad.1652708 (2008).

452 van Steensel, B. & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780-791, doi:10.1016/j.cell.2017.04.022 (2017).

453 Kim, E. *et al.* SRSF2 Mutations Contribute to Myelodysplasia Through Mutant-Specific Effects on Exon Recognition. *Cancer cell* **27**, 617-630, doi:10.1016/j.ccell.2015.04.006 (2015).

454 Shirai, C. L. *et al.* Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer cell* **27**, 631-643, doi:10.1016/j.ccell.2015.04.008 (2015).

455 Hu, D. *et al.* The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Molecular and cellular biology* **33**, 4745-4754, doi:10.1128/mcb.01181-13 (2013).

456 Hsieh, J. J. D., Cheng, E. H. Y. & Korsmeyer, S. J. Taspase1: A Threonine Aspartase Required for Cleavage of MLL and Proper HOX Gene Expression. *Cell* **115**, 293-303, doi:https://doi.org/10.1016/S0092-8674(03)00816-X (2003).

457 Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nature reviews. Genetics* **7**, 715-727, doi:10.1038/nrg1945 (2006).

458 Villar, D. *et al.* Enhancer evolution across 20 mammalian species. *Cell* **160**, 554-566, doi:10.1016/j.cell.2015.01.006 (2015).

459 de Haan, G., Nijhof, W. & Van Zant, G. Mouse Strain-Dependent Changes in Frequency and Proliferation of Hematopoietic Stem Cells During Aging: Correlation Between Lifespan and Cycling Activity. *Blood* **89**, 1543-1550 (1997).

460 Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2017. (Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services, Atlanta, GA, 2017).

461 Geerlings, S. E. & Hoepelman, A. I. Immune dysfunction in patients with diabetes mellitus (DM). *FEMS immunology and medical microbiology* **26**, 259-265 (1999).

462 Ling, C. & Groop, L. Epigenetics: A Molecular Link Between Environmental Factors and Type 2 Diabetes. *Diabetes* **58**, 2718 (2009).

463 Davis, J. S. *et al.* Use of non-steroidal anti-inflammatory drugs in US adults: changes over time and by demographic. *Open heart* **4** (2017).

464 Hoggatt, J. *et al.* Differential stem- and progenitor-cell trafficking by prostaglandin E2. *Nature* **495**, 365-369, doi:10.1038/nature11929 (2013).

465 Bancos, S., Bernard, M. P., Topham, D. J. & Phipps, R. P. Ibuprofen and other widely used non-steroidal anti-inflammatory drugs inhibit antibody production in human cells. *Cellular immunology* **258**, 18-28, doi:10.1016/j.cellimm.2009.03.007 (2009).

466 Austad, S. N. & Fischer, K. E. Sex Differences in Lifespan. *Cell metabolism* **23**, 1022-1033, doi:10.1016/j.cmet.2016.05.019 (2016).

467 Nakada, D. *et al.* Estrogen increases haematopoietic stem cell self-renewal in females and during pregnancy. *Nature* **505**, 555-558, doi:10.1038/nature12932 (2014).

468 Heo, H.-R. *et al.* Hormonal Regulation of Hematopoietic Stem Cells and Their Niche: A Focus on Estrogen. *International journal of stem cells* **8**, 18-23, doi:10.15283/ijsc.2015.8.1.18 (2015).

469 Ray, R. *et al.* Sex Steroids and Stem Cell Function. *Molecular Medicine* **14**, 493-501, doi:10.2119/2008-00004.Ray (2008).

470 Mendez-Ferrer, S., Chow, A., Merad, M. & Frenette, P. S. Circadian rhythms influence hematopoietic stem cells. *Current opinion in hematology* **16**, 235-242, doi:10.1097/MOH.0b013e32832bd0f5 (2009).

471 Puram, R. V. *et al.* Core circadian clock genes regulate leukemia stem cells in AML. *Cell* **165**, 303-316, doi:10.1016/j.cell.2016.03.015 (2016).

472 Oh, G. *et al.* Cytosine modifications exhibit circadian oscillations that are involved in epigenetic diversity and aging. *Nature communications* **9**, 644, doi:10.1038/s41467-018-03073-7 (2018).

473 Feil, R. & Fraga, M. F. Epigenetics and the environment: emerging patterns and implications. *Nature reviews. Genetics* **13**, 97-109, doi:10.1038/nrg3142 (2012).

474 Will, B. *et al.* Minimal PU.1 reduction induces a preleukemic state and promotes development of acute myeloid leukemia. *Nat Med* **21**, 1172-1181, doi:10.1038/nm.3936 (2015).

475 Chapuy, B. *et al.* Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777-790, doi:10.1016/j.ccr.2013.11.003 (2013).

476 Heyn, H. *et al.* Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A.* **109**, doi:10.1073/pnas.1120658109 (2012).

477 Henry, C. J. *et al.* Aging-associated inflammation promotes selection for adaptive oncogenic events in B cell progenitors. *The Journal of clinical investigation* **125**, 4666-4680, doi:10.1172/jci83024 (2015).

478 Rozhok, A. I., Salstrom, J. L. & DeGregori, J. Stochastic modeling reveals an evolutionary mechanism underlying elevated rates of childhood leukemia. *Proc Natl Acad Sci U S A* **113**, 1050-1055, doi:10.1073/pnas.1509333113 (2016).

479 Ergen, A. V., Boles, N. C. & Goodell, M. A. Rantes/Ccl5 influences hematopoietic stem cell subtypes and causes myeloid skewing. *Blood* **119**, 2500-2509, doi:10.1182/blood-2011-11-391730 (2012).

480 Guidi, N. *et al.* Osteopontin attenuates aging-associated phenotypes of hematopoietic stem cells. *Embo j* **36**, 840-853, doi:10.15252/embj.201694969 (2017).

481 Poulos, M. G. *et al.* Endothelial transplantation rejuvenates aged hematopoietic stem cell function. *The Journal of clinical investigation*, doi:10.1172/JCI93940 (2017).

482 Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**, 1348-1354, doi:10.1038/ng1896 (2006).

483 Notta, F. *et al.* DISTINCT ROUTES OF LINEAGE DEVELOPMENT RESHAPE THE HUMAN BLOOD HIERARCHY ACROSS ONTOGENY. *Science (New York, N.Y.)* **351**, aab2116-aab2116, doi:10.1126/science.aab2116 (2016).

484 Poleshko, A. *et al.* Genome-Nuclear Lamina Interactions Regulate Cardiac Stem Cell Lineage Restriction. *Cell* **171**, 573-587 e514, doi:10.1016/j.cell.2017.09.018 (2017).

485 Isoda, T. *et al.* Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate. *Cell* **171**, 103-119 e118, doi:10.1016/j.cell.2017.09.001 (2017).

486 Mateos, J. *et al.* Lamin A deregulation in human mesenchymal stem cells promotes an impairment in their chondrogenic potential and imbalance in their

response to oxidative stress. *Stem cell research* **11**, 1137-1148, doi:10.1016/j.scr.2013.07.004 (2013).

487 Wang, C. Q. *et al.* Runx3 deficiency results in myeloproliferative disorder in aged mice. *Blood* **122**, 562-566, doi:10.1182/blood-2012-10-460618 (2013).