HHS Public Access

Author manuscript

Cancer Discov. Author manuscript; available in PMC 2020 August 01.

Published in final edited form as:

Cancer Discov. 2019 August ; 9(8): 1080–1101. doi:10.1158/2159-8290.CD-18-1474.

Aging Human Hematopoietic Stem Cells Manifest Profound Epigenetic Reprogramming of Enhancers That May Predispose to Leukemia

Emmalee R. Adelman[1,2,3], **Hsuan-Ting Huang**[1,2,#], **Alejandro Roisman**[1,2,#], **André Olsson**[4], **Antonio Colaprico**[1,2], **Tingting Qin**[5], **R. Coleman Lindsley**[6], **Rafael Bejar**[7], **Nathan Salomonis**[8], **H. Leighton Grimes**[4], **Maria E. Figueroa**[1,2,*]

[1]Dept. of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, 33136.

[2]Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL, 33136.

[3]Dept. of Pathology, University of Michigan Medical School, Ann Arbor, MI, 48109.

[4]Div. of Immunobiology and Center for Systems Immunology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229.

[5]Dept. of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, 48109.

[6]Dept. of Medical Oncology, Div. of Hematological Malignancies, Dana-Farber Cancer Institute, Boston, MA, 02215.

[7]Division of Hematology and Oncology, Moores Cancer Center, University of California San Diego, La Jolla, CA, 92023.

[8]Div. of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229.

Abstract

Aging is associated with functional decline of hematopoietic stem cells (HSC) as well as an increased risk of myeloid malignancies. We performed an integrative characterization of epigenomic and transcriptomic changes, including single-cell RNA-seq, during normal human aging. Lineage^{-}CD34^{+}CD38^{-} cells (HSC-enriched, HSCe) undergo age-associated epigenetic reprogramming consisting of redistribution of DNA methylation and reductions in H3K27ac,

[*]Correspondence to: María E. Figueroa, M.D. Associate Professor, Department of Human Genetics, University of Miami Miller School of Medicine, 1501 NW 10th Ave, BRB 723, Miami, FL, 33136, mef162@miami.edu, Phone: 305-243-7333.

[#]These authors contributed equally

AUTHOR CONTRIBUTIONS:

ERA conceived and designed the study, performed the experiments, performed data analysis and wrote the manuscript. HTH and AR performed experiments and analyzed data. AO performed experiments. AC and TQ contributed to data analysis. RB performed experiments and analyzed data. RCL contributed key reagents. NS and HLG performed data analysis and contributed to study design and data interpretation. MEF conceived and designed the study, oversaw data analysis and interpretation, and wrote the manuscript. The authors declare no competing financial interests.

CONTACT FOR REAGENT AND RESOURCE SHARING:

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Maria Figueroa (mef162@miami.edu).

H3K4me1 and H3K4me3. This reprogramming of aged HSCe globally targets developmental and cancer pathways which are comparably altered in AML of all ages; encompassing loss of 4,646 active enhancers, 3,091 bivalent promoters, and deregulation of several epigenetic modifiers and key hematopoietic transcription factors, such as KLF6, BCL6 and RUNX3. Notably, in vitro downregulation of KLF6 results in impaired differentiation, increased colony forming potential and changes in expression that recapitulate aging and leukemia signatures. Thus, age-associated epigenetic reprogramming may form a predisposing condition for the development of age-related AML.

Statement of significance:

Acute Myeloid Leukemia, which is more frequent in the elderly, is characterized by epigenetic deregulation. We demonstrate that epigenetic reprogramming of human hematopoietic stem cells occurs with age, affecting cancer and developmental pathways. Downregulation of genes epigenetically altered with age leads to impairment in differentiation and partially recapitulates aging phenotypes.

Keywords

Aging; Epigenetics; Enhancers; HSC; Leukemia

Introduction:

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 (1). This change in demographics will require a better understanding of aging-associated features and medical challenges, including the higher incidence of cancer. At the hematopoietic level, aging is associated with an increased rate of isolated cytopenias of unknown significance, loss of adaptive immunity, and an increased predisposition to develop myeloid malignancies such as myelodysplastic syndromes (MDS) and acute myeloid leukemia (AML) (2-5). Additionally, aging correlates with an increased incidence of clonal hematopoiesis of indeterminate potential (CHIP) with frequent mutations in the epigenetic modifiers *DNMT3A*, *TET2*, and *ASXL1,* as well as in proteins involved in alternative splicing such as *SF3B1* and *SRSF2*. However, while patients carrying these mutations experience an overall increased likelihood of developing hematological malignancies, this occurs at a rate of only 0.5–1% per year and does not affect every carrier (6-10). Moreover, not all age-related myeloid malignancies occur in the context of CHIP mutations. Thus, additional events must be necessary to initiate malignant transformation in healthy individuals.

Aged HSC display decreased self-renewal, reduced homing ability, and an ultimate myeloid differentiation bias due to reduced lymphoid potential (11-15). This loss of function is due in part to altered DNA damage response and accumulation of DNA damage, increased reactive oxygen species (ROS), and modifications in WNT signaling (16-21). Notably, reprogramming of HSC through a pluripotent intermediate partially rejuvenates aged murine HSC, suggesting that epigenetic changes with age contribute to impaired HSC function (22).

In line with this, changes in the epigenome have been reported with normal aging. Aged murine HSC display regional gains in DNA methylation as well as an increase in the number and breadth of H3K4me3 peaks, suggesting that epigenetic shifts with age may contribute to HSC loss of function (23,24). In human hematopoietic progenitors, global changes in histone modification levels have been reported (25), but detailed studies into chromatin profiles of aged human HSC are still lacking. Likewise, while DNA methylation has been profiled in human specimens, these studies utilized either peripheral blood, bulk mononuclear cells (MNC), or unfractionated CD34+ bone marrow cells (26,27). Thus, it is as yet unknown if the age-related epigenetic differences observed in aged murine HSC are representative of those which occur with human HSC aging, or whether these epigenetic changes contribute to the increased risk of malignant transformation.

In recent years, there has been increasing evidence for the role of deregulated enhancers and other regulatory elements in hematological malignancies (28-32). Thus, given the epigenetic changes observed with mouse HSC aging, we hypothesized that aging of human HSCs is associated with changes in the epigenetic landscape that likely impact enhancers and other regulatory elements, and contribute to aged HSC loss of function and increased risk of malignant transformation.

Results:

Aged HSCs display systematic changes in histone and cytosine modifications

We hypothesized that epigenetic changes in human HSCs with aging may underlie the age-related functional decline observed in these cells. Therefore, in order to perform a comprehensive characterization of the epigenetic and transcriptional landscape of these cells, bone marrow MNC were collected from a total of 41 young (18–30 yo) and 55 aged (65–75 yo) healthy donors. We first evaluated the frequencies of the HSC-enriched lineage-negative (Lin^-) $CD34^+$ $CD38^-$ (herein HSCe) fraction, 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 our cohort. As previously shown, we observed an increase in HSCe frequency with age ($p=1.59e-5$) as well as a decrease in the frequency of GMP with an increase in MEP amongst the $Lin^-CD34^+CD38^+$ fraction ($p<0.001$) (33,34) (Supplementary Fig. S1A-C).

In order to understand the incidence of clonal hematopoiesis in our aging cohort and how this may impact our epigenomic studies, we performed targeted sequencing of a panel of 128 genes. From a total of 59 donors (n=27 young and 32 aged donors) used to complete all the genome-wide epigenetic and transcriptome profiling, 20 had sufficient material available to also perform mutational profiling. Only one of these donors presented a mutation, 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 (Supplementary Table S1).

In order to capture dynamic age-associated epigenetic changes in human HSCe, we characterized histone and cytosine modifications in young and aged HSCe isolated from healthy donors. We first sought to determine how histone modifications change with normal HSCe aging. Using a low-input protocol for chromatin immunoprecipitation followed by massively parallel sequencing (micro ChIP-seq), we determined the genome-wide distribution of H3K4me1, H3K27ac, H3K4me3 and H3K27me3, on 4–7 independent donors per age, per mark. Differential peak calling analysis for each of these four histone marks across the two age groups revealed that aging is associated with significant reductions in H3K4me1, H3K27ac and H3K4me3, affecting 14,077 (11% of all young peaks), 10,512 (24%) peaks and 18,195 (31%) peaks, respectively ($\log_{10}$ likelihood ratio >3, absolute fold-change ≥1.5). Notably, relatively few regions had an increase in signal intensity of H3K4me1 (n=229 peaks), H3K4me3 (n=17 peaks), or H3K27ac (n=35 peaks) with age (Figure 1A **and** Supplementary Fig. S2A-B). Importantly, while age-related changes in H3K27ac, H3K4me3 and H3K4me1 were distributed across the genome, these changes were not random, but were instead focalized at specific areas of the genome that were reproducibly affected across the different donors. By contrast, H3K27me3 displayed only 1,564 H3K27me3 peaks changed with aging, consisting of both gains (25% of differential peaks) and losses of H3K27me3 (75% of differential peaks). However, despite fewer H3K27me3 peaks being affected with aging, these peaks displayed the greatest magnitude of change in signal intensity (Supplementary Fig. S2A-B **and** Supplementary Table S2).

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 <0.05 for all marks except for H3K4me3 for which it was 0.14. Moreover, both Western blot and ChIP-seq analyses of histone 3 (H3) confirm age-related differences in H3 modifications are not due to a reduction in total H3 levels with age, which is in line with the findings of Cheung et al by mass-cytometry (25) (Supplementary Fig. S3A-D). Finally, to control for any potential gender effects, differential peak calling was also performed in a sex-dependent manner and showed that even when restricted to only male or female donors, the above observations were still present with significant reductions in H3K4me1, H3K4me3, and H3K27ac, and only minimal changes in H3K27me3, suggesting that age-related reductions in these marks are not sex-dependent (Supplementary Fig. S3E-F).

Functional annotation of age-related differences in histone profiles revealed that sites with decreased H3K4me1 were linked to genes involved in myeloid and erythroid differentiation and functions, while loss of H3K27ac was linked to genes associated with leukocyte activation, apoptotic signaling and histone modifications. Changes in H3K4me3 and H3K27me3 were associated with genes involved in developmental processes, including cell fate commitment and the WNT receptor signaling pathway (Figure 1B **and** Supplementary Table S2). These findings further indicate that age-related changes in histone modifications are not governed by random changes, but instead represent a systematic change in the epigenetic program of HSCe.

Next, to determine whether cytosine modifications are also reprogrammed during aging, we analyzed DNA methylation landscapes in young and aged HSCe by ERRBS, a reduced

representation bisulfite sequencing approach which captures ~3 million CpGs across the genome (35) (n=5–7 per age group). We identified 529 differentially methylated regions (DMR) (beta binomial test, Benjamini-Höchberg corrected p-value < 0.05 and absolute difference ≥20%), encompassing 2,249 differentially methylated cytosines (DMCs) annotated to 748 unique genes (Figure 1C **and** Supplementary Table S3). Notably, despite the reduced representation aspect of the ERRBS assay as well as the focal nature of the changes observed, we were still able to capture biologically relevant DMRs associated with aging, which were significantly enriched for pathways relevant to HSC biology such as WNT, Cadherin and cell adhesion pathways (Figure 1D).

Enhancer deregulation is a feature of aged HSCe

In order to obtain a more comprehensive picture of the epigenetic changes observed and identify groups of genes with similar age-related changes in human HSCe, we performed an integrative analysis across all regions that displayed changes in the histone marks studied. Using k-means clustering we were able to identify 12 different genomic clusters based on their age-associated epigenetic changes (Figure 2A **and** Supplementary Table S4). Genomic annotation of these clusters revealed that clusters A through F were associated with epigenetic changes at promoter regions, while clusters G through L were associated with changes within gene bodies and distal intergenic regions.

To further characterize these age-related changes in the context of specific regulatory regions, we used the histone profiles from young HSCe to establish the normal landscape of poised enhancers, active enhancers, and bivalent promoters in HSCe. 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. 2,066 poised enhancers and 11,569 active enhancers were identified in young HSCe. Bivalent promoters were defined as promoter regions with overlapping enrichment for H3K4me3 and H3K27me3, and 3,604 genes with such promoters were identified in young HSCe (Supplementary Table S5). Annotation of the 12 clusters of differentially regulated regions in aged HSCe to these genomic regulatory regions revealed 5 categories of genomic regions that are altered with age: Active TSS I, Active TSS II, Bivalent, Non-TSS, and Enhancer. Clusters J, K and L, which were defined by non-promoter regions with loss of H3K27ac in the absence of any H3K4me3 changes, were enriched for regions overlapping with active enhancers (n=4,124 enhancer peaks) (Figure 2B **and** Supplementary Table S4). Notably, fewer poised enhancers (n=225 peaks) were annotated to any of the clusters that showed differential histone enrichment with aging, indicating that this type of regulatory element is less impacted by age than active enhancers. Genes annotated to enhancers displaying age-related deregulation were involved in pathways associated with lymphoid and immune signaling as well as myeloid leukemias (Figure 2C **and** Supplementary Table S4). Amongst the genes associated with loss of active enhancers were several transcription factors such as *RUNX1/2/3, HIF1A, IKZF1, MEIS1* and *KLF6,* as well as *ETV6* and *GFI1,* which have known tumor suppressor functions. In addition to transcription factors, enhancers affected with aging were also annotated to several epigenetic modifiers such as *KAT6B, KDM8, CBX7,* and *DOT1L* (Figure 2B).

DNA motif analysis of the enhancer clusters identified strong enrichment for FLI1, ERG, ETV1, IRF2, and RUNX1 motifs, indicating that the establishment of new epigenetic programs in aged HSCe may be mediated through changes in the function of these transcription factors. We next took advantage of publicly available ChIP-seq datasets, to validate binding of these transcription factors to these regions in human CD34+ cells (36) and found that indeed, enhancers lost with age overlapped with FLI1, ERG and RUNX1 binding (Figure 2D **and** E). Together these results suggest that enhancer deregulation of hematopoietic transcription factors and immune pathways may be one mechanism contributing to HSC decline with age.

Bivalent promoters in HSCe display a switch to repression during aging

Amongst the three categories of age-associated epigenetic changes that were enriched for promoter regions, two (Active TSS I and Active TSS II) were defined by reduction of H3K4me3 at non-bivalently marked promoters while the third one was annotated to bivalent promoters (Figure 2A **and** 3A-B). Active TSS I (clusters A and B) included active promoters marked by the presence of both H3K4me3 and H3K27ac, and predominantly lost the acetylation mark with aging, with a milder loss of H3K4me3. These promoters are enriched in pathways associated with B and T cell receptor signaling, apoptosis, cell cycle, mTOR and MAPK signaling, as well as several cancer-related pathways. In contrast, peaks within the Active TSS II category (clusters C and D) were marked by the presence of H3K4me3, which was significantly lost with aging. These promoters were enriched almost exclusively in cancer related pathways (Figure 3B-D).

Like the active promoter clusters, clusters E and F showed reduction of H3K4me3 with age. However, these clusters were highly enriched for bivalent promoters, 79% (n=2,171) of genes within clusters E and F had bivalent promoters that were altered with age (Figure 3A-B **and** Supplementary Table S4). These clusters were annotated to developmental pathways such as WNT and Hedgehog signaling, as well as cancer-related pathways (Figure 3E-F), and were enriched for TCF4 and homeobox motifs, including HOXA9 (Figure 3G). Notably, aging was associated with a marked loss of H3K4me3 at bivalent promoters, with only minimal changes in H3K27me3, thus resulting in an effective switch from bivalency to repression, an epigenetic change comparable to that seen in bivalent genes during malignant transformation (37,38). Such a switch, while having only minimal impact on the steady state expression of those genes which are normally expressed at very low levels, changes instead their potential for activation in response to signaling cues.

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

In order to identify the genes that may contribute to the marked histone and methylation changes we observed in aged HSCe, we performed a supervised analysis of RNA-seq profiles of young and aged HSCe (n=10 per age group) and determined that 1,133 genes were differentially expressed with age (B-H adjusted p-value <0.05 and absolute fold change ≥ 1.5) with 517 genes upregulated and 616 genes downregulated with aging (Figure 4A **and** Supplementary Table S6). Age-related expression changes in HSCe were strongly associated with epigenetic reprogramming at these loci. 732 genes presented changes at regions within

the active promoter clusters A-C and 290 at regions within the enhancer-enriched clusters J-L (Figure 4B-C).

Amongst the genes which were differentially expressed were genes previously reported as deregulated in aged HSCe, such as *CDC42*, *JUN*, and *FOSL (24,39).* 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 *GATA2, 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 levels of several epigenetic modifiers and co-repressors (upregulation of *CITED2 and HDAC11,* with downregulation of *KAT7*, *KAT8*, *KDM3A, SETD6, SETD8 and SETD1A).* To verify that these results were not biased by donor sex, we restricted the differential gene expression analysis to only female donors (n=3 young and n=9 aged) and found that 88% of the age-associated differentially expressed genes were still detected, confirming that the aging gene signature reflects differences in donor age rather than sex (Supplementary Fig. S4A-B).

Gene set enrichment analysis revealed downregulation of genes involved in 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. (40) was also confirmed in our cohort (Figure 4D **and** Supplementary Table S7). 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 (41). 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 (Supplementary Table S8).

Age-related epigenetic changes are the result of HSC reprogramming

To determine whether 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 or whether instead, these changes stemmed from true reprogramming, we performed single-cell RNA-seq (scRNA-seq) of HSCe from young and aged donors (n=5 donors <40 yo and n=4 donors > 60 yo). As the HSCe compartment contains both LT-HSC and more mature progenitors, we analyzed the composition changes of this population with age. Classification of the HSCe compartment (n=338 young HSCe and n=310 aged HSCe) using cellHarmony (42) and gene expression data from 18 reference human bone marrow progenitor cell types from the Human Cell Atlas (HCA) (43), revealed that the majority of young and aged HSCe (78.4% and 87.4%, respectively) corresponded to non-cycling HSC (nc-HSC). Within both the young and aged HSCe populations we identified low frequencies of MEP (3.6% and 3.5%, respectively), as well as multi-linage progenitors (0.3% and 0.6%, respectively). While we observed a decrease in cycling-HSC (12% in young vs. 7% in aged) and lymphoid-primed multipotent progenitors (LMPP; 5.9% in young and 1.6% in aged) with age, these small shifts in percent composition are not sufficient to explain the

magnitude of the epigenetic changes detected in bulk HSCe with aging (Supplementary Fig. S5A **and** Supplementary Table S9). Since the majority of the cells profiled represented nc-HSC and there was an insufficient number of the other types, we focused the rest of our analysis on nc-HSC. Differential analysis of these cells identified 364 differentially expressed genes (absolute fold-change > 1.5, eBayes t-test $p<0.05$, FDR <0.05) with age (Supplementary Fig. S5B **and** Supplementary Table S9). These included genes that were also differentially expressed at the bulk RNA-seq level, such as *EGR1*, *KLF6*, *HIF1a*, and *JUN*. HOPACH clustering of all genes differentially expressed in aged nc-HSC compared to young nc-HSC identified 3 clusters, or gene expression programs, that largely separated young and aged nc-HSC (Figure 5A **and** Supplementary Fig. S5B). To determine the frequency of these different gene expression programs in young and aged nc-HSC and determine whether a detectable sub-population of young nc-HSC exists that presents an expression profile comparable to aged nc-HSC, we calculated expression centroids for the young (n=1) and aged (n=2) nc-HSC clusters and then applied cellHarmony to re-classify each nc-HSC individually, using these references. Notably, while nc-HSC isolated from aged donors were frequently classified as young based on their expression profile (26/271; 9.6%), nc-HSC isolated from young donors were only exceptionally classified as aged (3/265; 1.1%) with only a subset of young donors showing this phenotype (Figure 5B). Thus, a defined aged-like population could not be reproducibly identified within the sequenced young nc-HSC and, while we cannot definitively rule out that the rare aged-like HSC detected in some of the young donors may contribute to some degree of population shift over time, these findings are compatible with a scenario in which the observed epigenetic and transcriptional changes are largely the result of reprogramming during aging.

Age-associated epigenetic changes may predispose for AML

Given that alterations in DNA methylation and deregulation of enhancer elements are observed in age-associated myeloid malignancies such as AML (32,44-46), we next sought to identify changes in DNA methylation or histone modifications with age that may predispose for this malignancy. Using our published ERRBS methylation data from 119 AML patients (15–77 yo) (32) and our young and aged HSCe samples, we sought to identify aging DMRs that were also differentially methylated in AML, regardless of patient age. We stipulated that if an age-related DMR was only aberrantly methylated in aged AML patients, it could simply be reflective of a normal aging change, like that of our aged HSCe donors. However, if the DMR is likewise aberrantly methylated in young and middle-aged patients it would indicate that the DMR is a universal feature of AML and may therefore be part of the process of malignant transformation such that AML in the young also display this change. Thus, starting with our age-related DMRs, we first used k-means clustering to identify groups of DMRs that behave in a similar manner. This resulted in 11 different DMR clusters, 8 of which corresponded to DMRs with hypermethylation with aging and 3 with age-related hypomethylation. A total of 7 DMR clusters (5 with age-related hypermethylation and 2 with hypomethylation ($p< 0.05$, Mann-Whitney rank sum test, corrected for multiple testing) were identified that consistently presented the same methylation changes across all AML age groups as detected in aged HSCe, while the remaining 5 clusters were either variable in their behavior or consistently displayed methylation changes in the opposite direction (Figure 6A-B **and** Supplementary Table 10). These AML-like DMRs were enriched for

genes associated with cell cycle regulation as well as transcription factors such as *KLF6*, *RUNX1*, and *HOXC4/6* in the hypermethylated clusters, while hypomethylated DMRs included *RXRA*, *SKI*, and *SOCS1* (Supplementary Table 10).

Similarly, we took advantage of two publicly available cohorts of AMLs that had been profiled for H3K27ac alone (n=52) (46), or H3K27ac in conjunction with H3K4me1, H3K4me3, and H3K27me3 (n=19 for H3K27ac, n=16 for H3K4me1, n=19 for H3K4me3, and n=17 for H3K27me3) (47). Using the categories of regulatory elements that we had identified as being altered with HSCe aging, we examined the enrichment or depletion of the defining histone marks for these categories in AML. Similar to our analysis of the age-related DNA methylation changes, we identified 6 clusters of H3K27ac peaks (n=4,582 peaks) and 4 clusters of H3K4me1 peaks (n=2,559 peaks) that had decreased enrichment of these marks at enhancers (p-adjusted < 0.005, Mann-Whitney rank sum test) in normal aged HSCe and AML blasts of all ages compared to young HSCe. In addition, 1 cluster of H3K27ac and 1 of H3K4me1 with enrichment of these marks at enhancers showed a similar pattern of mimicking the age-related increases seen in HSCe in AMLs of all ages (Figure 6C-D **and** Supplementary Fig. S6A, **and** Supplementary Table 10). These enhancers were associated with RAS signaling, cell adhesion, immune cell activation, and myeloid differentiation, and included genes such as *BCL6B*, *NFATC2*, *PML*, *RXRA*, *GADD45A*, *HES1*, *MEIS1* and *DNMT3B* (Figure 6E-F **and** Supplementary Fig. S6B). In addition, we also observed chromatin alterations at Active TSS sites in aged HSCe that were also present in AML, with 7 clusters (n=6,028 peaks) having reduced H3K27ac enrichment at active TSS I, and 4 clusters (n=2,661 peaks) showing loss of H3K4me3 at active TSS I and II in aged HSCe and across AMLs of all ages. These Active TSS were associated with VEGF, RAS, and p53 signaling, and developmental processes, respectively. Finally, we identified 9 clusters that showed loss of H3K4me3 and 3 clusters with loss of H3K27me3 (n=4,780 and n=596 peaks, respectively) at bivalent promoters in aged HSCe and AMLs, which were associated with cell fate and specification (Supplementary Fig. S6A-B).

Downregulation of KLF6 impairs differentiation and recapitulates aging and malignant expression profiles

Since *KLF6* is amongst the genes most strongly downregulated with HSCe aging (Figure 4A) and the DNA methylation changes seen at the locus are also detected in AMLs across all age groups, we next evaluated the impact of KLF6 downregulation on normal CD34+ cells. *KLF6* encodes for the Kruppel-like factor 6, a transcription factor essential for yolk sac hematopoiesis and which is involved in inflammation and myeloid differentiation (48,49). We first tested the effect of CRISPR-Cas9 mediated knockout of *KLF6* on the phenotype of CD34+ cells (Supplementary Table S11). Knockout of KLF6 resulted in an increase in both total colony numbers and granulocyte-monocyte colonies when plated on methylcellulose, demonstrating that downregulation of this gene in CD34+ induces an increase in myeloid colony-forming potential ($p < 0.05$) (Figure 7A **and** Supplementary Fig. S7A-B). We further tested the impact of KLF6 downregulation on myeloid and erythroid differentiation potential in liquid culture. After 7 days in myeloid or erythroid promoting conditions, CD34+ cells transfected with sgKLF6 had decreased expression of CD11b and CD71, respectively, with increased expression of CD34 compared to control sgRNA (n=5,

p<0.05) (Figure 7B). Persistent robust knockdown of KLF6 at this time point was confirmed using intracellular flow cytometry staining (Supplementary Fig. S7C).

RNA-seq profiling of CD34+ cells after downregulation of KLF6 confirmed a partial recapitulation of the aging HSCe signature with enrichment of both up and downregulated age-related expression changes (Figure 7C). Furthermore, gene set enrichment analysis of RNA-seq of CD34+ with *KLF6* knockout showed strong enrichment of several AML- and cancer-associated gene sets (FDR < 0.05) (Figure 7D **and** Supplementary Table S6). Together, these findings indicate that age-related loss of KLF6 results in an aberrant differentiation phenotype that may predispose to the development of myeloid neoplasms.

Of note, while many other genes change in a manner similar to *KLF6*, not all age-related changes may predispose to malignancy in a manner similar to that induced by *KLF6* knockout. *LMNA*, which encodes for the nuclear Lamin A/C protein, and is one of the genes mutated in Hutchinson-Gilford progeria syndrome, was also one of the most downregulated genes in aged HSCe, and age-related H3K27ac changes at this locus parallel those seen in AMLs of all ages. Downregulation of LMNA by shRNA in CD34+ cells recapitulated features of aging, which consisted of partial recapitulation of the aging signature and increased colony-forming potential, and variable impact on the differentiation potential, with CD34+ cells from different donors experiencing different degrees of myeloid or erythroid differentiation impairment (Supplementary Fig. S7D-G). Thus, we show that both *LMNA* and *KLF6* are epigenetically and transcriptionally deregulated with age, and as expected, only age-associated alteration of select genes, such as *KLF6*, may predispose to myeloid malignancy.

Discussion:

In the hematopoietic system, aging is associated with an increase in the prevalence of CHIP and the risk of developing myeloid leukemias (7,9,50). 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 (51,52). Here we present, to our knowledge, the first map of regulatory elements, including active and repressed promoters, enhancers, and bivalent promoters in normal human HSCe. In addition, we show that aging results in reproducible genome-wide changes at the epigenomic level, affecting both cytosine and histone modifications and targeting key developmental and cancer pathways, including several known tumor suppressor genes in AML and MDS. This shift in epigenetic program of HSCe with aging was particularly evident at active enhancers, indicating that, similar to what has been observed for cancer (28,31,53,54), age-related dysfunction of human HSCe may also be mediated through the dysfunction of these regulatory elements. As such, many of the putative active enhancers that lose H3K27ac with age were associated with genes involved in lymphoid and immune signaling, opening the possibility that through a change in their activation potential, HSCe enhancer dysfunction may be one of the mechanisms through which immune function is compromised with aging. Importantly, despite certain degree of interpersonal variability, age-related epigenetic

changes targeted a core set of regions across all donors, which was reproducibly detected despite the relatively small number of donors analyzed, indicating that these changes are unlikely to be stochastic. Moreover, scRNA-seq analysis further confirmed that the observed changes in the epigenetic program are unlikely due to the expansion of a pre-existing subpopulation within the bone marrow, but rather the consequence of true epigenetic reprogramming. Thus, the epigenetic characterization within this study will not only allow for the identification of those epigenetic regulatory elements that are vital for maintaining normal human HSC function, but also those with the potential to be targeted to rejuvenate aged HSC.

While additional work will be required to determine the mechanism driving age-associated epigenetic reprogramming, 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 3 lysine 4 methyltransferase *SETD1A*, which may explain the reduced number of H3K4me1 and H3K4me3 peaks observed in aged HSCe. Since we also observed gene expression alterations in *KAT5/7/8*, *KDM2A/3A*, *SETD6/8*, and *HDAC11*, it is possible that HSCe aging may also present focal alterations in other histone modifications, as has been observed at the global level by others (25). Additionally, as alterations in RNA splicing have been observed with HSC aging (40), and the *U2AF1* and *SREK1* splicing factors become downregulated with age, it is possible that aberrant alternative splicing of specific epigenetic modifiers with aging may be contributing to age-associated epigenetic reprogramming. However, whether differential expression or usage of epigenetic modifiers leads to alterations in the epigenome or the reverse instead is true, is yet to be determined.

In addition, multiple transcription factors become epigenetically deregulated with aging, including *FLI1*, *RUNX3*, *BCL6*, *MEF2C*, and *KLF6*. Furthermore, regions with age-related epigenetic changes are enriched for motifs of many of these same transcription factors. Therefore, it is possible that changes in the binding of these transcription factors drives epigenetic reprogramming with age by altering the recruitment of epigenetic modifiers to these target sites.

Notably, a subset of the regions that show chromatin changes with normal aging are also altered in both elderly and young AML. These regions include enhancers that putatively regulate immune and cell-adhesion associated genes, as well as active promoters annotated to genes involved in RAS and VEGF signaling, and bivalent promoters associated with developmental genes. Likewise, a subset of age-associated DMRs involved in cell cycle regulation, apoptosis and WNT signaling also appear to be aberrantly methylated in all AMLs. These findings would suggest that alterations in DNA methylation and enhancer regulation with age may also contribute to leukemogenesis. Among these age-associated DMRs, we identified KLF6, which not only is amongst the most strongly downregulated transcription factors with aging, but it is also aberrantly methylated in AMLs. Notably, downregulation of KLF6 in normal CD34+ cells was sufficient to change the differentiation phenotype of the cells and recapitulate expression profiles seen in AMLs, indicating that deregulation of these transcription factors with aging is likely to contribute to both aging HSCe dysfunction and the increased risk of malignant transformation. However, given that

not all people develop myeloid malignancies with age, it is likely that age-associated epigenetic deregulation alone it not sufficient for transformation, but rather other cooperating events are also required. Future functional screens in the context of known clonal hematopoiesis or leukemic drivers will also be helpful for identifying which age-associated epigenetic changes are most vital for increasing risk of myeloid malignancy with age.

While our study focused on intrinsic characteristics of HSCe aging, extrinsic signals from the bone marrow niche should not be ignored. It is possible that the changes in the epigenetic program of aged HSCe are the result of microenvironment cues from an aged bone marrow niche, which becomes more adipogenic, with less bone cell formation and several studies have demonstrated the influence of the aged niche on HSC function (55-57). However, the precise mechanism by which the bone marrow niche may contribute to HSC epigenetic reprogramming with aging remains to be elucidated.

In summary, human HSCe experience dynamic age-related changes in multiple layers of the epigenome which result in a shift in the epigenetic program of aged HSCe likely to impact their functional capabilities. Most notably, scRNA-seq of HSCe points to a true reprogramming mechanism underlying these age-related changes, as opposed to an expansion of pre-existing subpopulations already present in the young bone marrow compartment. However, why cells would undergo age-related reprogramming and whether this reprogramming inevitably results in detrimental consequences for the cells, or instead, forms part of a protective mechanism to help compensate for other deleterious changes in the cells and the microenvironment is yet to be answered. Moreover, whether remodeling of the epigenome with age favors the development of ARCH/CHIP is yet to be determined. Given the low incidence of ARCH in our cohort and that reproducible age-associated epigenetic changes were observed regardless of the presence of ARCH accompanying mutations, it is possible that the remodeling of the epigenome with age may precede the development of clonal hematopoiesis. Further functional studies utilizing models of clonal hematopoiesis and single-cell experiments will be required in the future to help establish this as well as to discern how any deleterious changes contribute to driving the aged HSCe phenotype and predisposition to myeloid malignancies.

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 Medical School or the University of Miami Miller School of Medicine. The study was approved by the Institutional Review Boards of the University of Michigan Medical School and University of Miami Miller School of Medicine. All material included in this study was de-identified, leftover material to be discarded after pathological review of the hip specimens, and was therefore exempt from human subjects research regulations. 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) donors. Supplemental Table 1 describes detailed donor information.

FACS isolation of HSCe

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-) 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 (primer sequences in Supplementary Table S11). 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 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Illumina adapters were trimmed using Cutadapt (version 1.12) (58). Reads were aligned to hg19 using Bowtie (version 2.0.5) (59). All samples had at least 28,000,000 aligned reads. Only reads mapping to unique locations were retained for downstream analysis.

ChIP-seq analysis and differential 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 (60). For visual representation and differential peak calling, aligned reads for each mark for each age group were pooled using samtools v1.3.1 (61). 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 (62) 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 (63). Significant peaks for each age group and differential peaks (FDR<0.05, Log_{10} likelihood ratio>3 and absolute fold change >1.5, respectively) were annotated to the nearest TSS of Refseq hg19 using the R packages ChIP-enrich, and Broad-enrich for H3K27me3 (v1.8.0) (64). 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 (65). 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.

Clustering of aging histone changes

k-means clustering was performed using all peaks that had significant (LLR > 3, absolute fold-change >1.5) H3K4me1, H3K4me3, H3K27ac, or H3K27me3 changes with age (n=37,058 peaks) and R version 3.5.1. For the input matrix, scores for each histone modification, for each age group, were calculated using deeptools2 multiBigwigSummary (62) with bigwig files that were read normalized and contained the fold-enrichment of the Pooled IP/Pooled Input. The fold-change of aged/young was calculated for each modification for each peak using gtools (63). Peaks with low-counts for a given histone

modification (score aged <3 and score young <3) were filtered by converting the fold-change to zero for that respective mark. K-means clustering of the fold-change values was then performed using base R, with set.seed(123), centers=12, and nstart=50. Genomic annotation of each peak was performed using the R-package Genomation (v1.2.2) (66), with promoter regions being defined as +/− 3000 bp of the TSS. Genomation was also used to annotate peaks to bivalent promoters and active enhancers identified in young HSCe. The heatmap was plotted using ComplexHeatmap (67), splitting between each cluster.

Comparison to AML ChIP-seq

ChIP-seq data from AML blasts was downloaded from SRA (SRP103200) (46) and EGA (EGAD00001002340 and EGAD00001002418) (47) and processed as described above. In total, there were 71 patients for H3K27ac, 16 for H3K4me1, 19 for H3K4me3, and 17 for H3K27me3, available for download and used in the analysis, Only patients for which donor age was annotated were kept for downstream analysis. Deeptools2 bamCompare (62) was then used to create $\log_2$ normalized bigwig files for each sample and for young and aged HSCe samples with the following parameters: bamCompare --bamfile1 ip --bamfile2 input --outFileName name_log2.bw --ratio log2 -p 30 --verbose --outFileFormat bigwig --scaleFactorsMethod readCount. For each category of interest (enhancer, active TSS I and II, and bivalent promoter), a matrix containing the counts for each region within the category was generated using the bigwig files for the relevant histone modification and deeptools2 multiBigwigSummary. H3K27ac samples from both Yi et al. and McKeown et al were used together. The optimal number of k-means clusters was calculated using the gap statistic method in factoextra (68). K-means clustering was performed using R, and the heatmap was plotted using ComplexHeatmap (67), splitting between each cluster, and using light grey for NA values. Boxplots were plotted using ggplot2 (69). Clusters that had significantly different enrichment for the relative histone mark with age and AML were identified using the one-sided Mann-Whitney rank-sum test (implemented in python 3). Only those clusters where young HSCe was significantly ($p<0.05$ and Bonferroni p-adjusted < 0.005) different from all other conditions were considering to be predisposing to AML.

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 (35). 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 0.12.8 and 0.4.1, respectively) (70). 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) (71) and R statistical software (version 3.2.1, R-Core Team,

http://www.R-project.org/). Differentially methylated cytosines (DMC) were then inputted into the R package eDMR (v.0.6.3.1) (72), 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 (73). Functional annotation was performed using DAVID version 6.7 (74,75). 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 light grey for NA values.

Comparison to AML DNA methylation

The percent methylation for each age-associated hyper or hypo DMR was calculated for each young and aged HSCe donor and AML patient (32). NA values were removed prior to clustering. The optimal number of k-means clusters was calculated using the gap statistic method in factoextra (68). K-means clustering was performed using R. Boxplots were plotted using ggplot2 (69). Clusters that had significantly different methylation with age and AML were identified using the one-sided Mann-Whitney rank-sum test (implemented in python 3). Only those clusters where young HSCe methylation was significantly (Bonferroni p-adjusted < 0.00625 for hyper and p-adjusted < 0.017 for hypo) different from all other conditions were considering to be predisposing to AML.

RNA-seq

FACS isolated HSCe were sorted into Qiagen RLT+ (n=10 biological replicates per age group). 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 (58). 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 (76).

Differential gene expression analysis

Gene counts were calculated using QoRTs (v1.0.7) (77). 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 (78). 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.

GSEA

Gene set enrichment analysis (GSEA) was performed using a list of genes pre-ranked by the Wald statistic (stat column from DESeq2 output) (79). 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 (40).

Cell isolation for scRNA-seq

Mononuclear cells were isolated from 5 young (24–37 years old) donors and 4 aged (64–71 years old) donors 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.

scRNA-seq

Single-cell Lin-CD34+CD38-isolated cells were captured using the C_1™ Single-Cell Auto Prep System (Fluidigm, San Fransisco, CA), according to the manufacturer's instructions. In short, flow-sorted cells were counted and resuspended at a concentration of 35,000 cells per 100 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 50–84 single cells were normally captured. Cells were lysed on chip and reverse transcription was performed using Clontech SMARTer® Kit using the mRNA-Seq: RT + Amp (1771×) according to the manual. After the reverse transcriptase step, cDNAs were transferred to a 96 well plate and diluted with 5 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 using 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 75 bp RNA-Sequencing on a HiSeq 2500 (Illumina Inc., San Diego, CA). 96 scRNA-Seq libraries were sequenced per HiSeq 2500 cell (~300 million bp/cell).

scRNA-seq Analysis

The Fluidigm C1 RNA-seq data was aligned to the human transcriptome (hg19) and quantified using the software RSEM (80). Cells with a minimum of 100,000 aligned reads were retained for further analysis. To assign cell-types to these profiles, the software cellHarmony (42) was used to classify individual cells according to 18 cell populations from the bone marrow single-cell human cell atlas (HCA) (43). Differential expression analyses were performed in the software AltAnalyze (81), using an empirical Bayes moderated t-test, following FDR correction where indicated. Clustering for the heatmap of non-cyling young vs. aged DEG was performed using HOPACH in AltAnalyze. To determine if any young HSC had an aged gene-signature profile, expression centroids from the 3 clusters identified by HOPACH (n=1 young cluster and n=2 aged clusters) were used to re-classify young and aged HSC.

Enhancer analysis

To identify young poised enhancers ((H3K4me1>H3K4me3), H3K27me3+, H3K27ac-, and not within the promoter region), peak calling of young pooled H3K4me1 was performed as described above. The enrichment of H3K4me1 and H3K4me3 at these regions was calculated using deeptools2 multiBigwigSummary (62) with bigwig files that were read normalized and contained the fold-enrichment of the Pooled IP/Pooled Input. Peaks where H3K4me1 enrichment was not greater than twice that of H3K4me3 were discarded. Genomic annotation of the remaining peaks was performed using the R-package Genomation (v1.2.2) (66), and peaks within 3 kb of a TSS were removed from further analysis. Peaks overlapping (by at least 1 basepair) H3K27me3, but not H3K27ac peaks, were found with bedtools intersect. Active enhancers were defined as H3K27ac+, (H3K4me1 > H3K4me3), H3K27me3-, and not within the promoter region. To identify young active enhancers, 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, but did not overlap with H3K27me3 peaks. Active and poised enhancers were annotated to the nearest TSS using the R package ChIP-enrich with hg19 RefSeq (64).

Bivalent promoter analysis

Peak calling of pooled young HSCe H3K4me3 and H3K27me3 was performed as described above. Promoter peaks (+/− 3000 bp of TSS) were identified using the R-package Genomation (v1.2.2) (66). The bedtools intersect function was used to identify regions that were bound by both H3K4me3 and H3K27me3 in young HSCe (overlap >25%). Peaks

overlapping bivalent promoters (n=4,662) were annotated to the nearest TSS using hg19 RefSeq and using ChIP-enrich to identify the total number of genes with bivalent promoters (n=3,604) (66).

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 (82). 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. GAPDH normalized values were then normalized to the average young intensity. 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).

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 clusters altered with age (83). 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 the peaks within the enhancer clusters J-L (36).

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, age-associated clusters, and regions predisposing to AML was performed using the R package ChIP-enrich (64). Regions were annotated to RefSeq hg19 Gene Ontology Biological Processes data sets of less than 500 genes, using the method=chipenrich function for H3K4me1, H3K4me3, H3K27ac, and clusters. H3K27me3 was annotated using method=broadenrich. All ChIP-seq peaks and age-associated clusters were annotated to the nearest TSS, while DMRs predisposing to AML were annotated to the nearest gene. Results with FDR<0.05 were considered highly significant.

Heatmaps and density plots of ChIP-seq

Heatmaps of differential ChIP-seq peaks, enhancers, and bivalent promoters were created using deepTools2 (v2.5.0.1) (62). 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 (62). 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.

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.

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 (84). 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 (85). 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 (86). Somatic variants were called using LoFreq v2.1.1 for all variants at >=1% variant frequency (87). 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 (88). Variants predicted to alter splicing were assessed as described in Jian et al (89). 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 (90) or the Exome Aggregation Consortium (ExAC v.3.1) were similarly excluded unless they were also listed as confirmed somatic mutations in COSMIC (91). 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 shRNA targeting LMNA (*5'-GATGATCCCTTGCTGACTTAC-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 or purchased from Fred Hutchinson Hematopoietic Cell Procurement Services. 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 unit assay

Sorted CD34+ were seeded in methylcellulose, MethoCult H4435 (StemCell Technologies, #04435), in duplicate onto a 6-well SmartDish (StemCell Technologies, #27302) at varying

densities of 500 to 1,500 cells. Colonies were scored on a STEMvision after 14 days of incubation (StemCell Technologies) (n=7 and n=5 for shLMNA and sgKLF6, respectively).

sgRNA Synthesis

Protospacer sequences were identified using the CRISPRdesign algorithm (http://crispr.mit.edu) or CRISPRScan (www.crisprscan.org) (92). DNA templates for sgRNAs contain a T7 promoter, the protospacer sequence, and the sgRNA scaffold sequence (93). They were produced by PCR using custom forward primers and a reverse primer that amplifies the sgRNA scaffold of the plasmid pKLV-U6gRNA-PGKpuro2ABFP (Addgene #62348). The PCR products were purified and in vitro transcribed with the HiScribe T7 High Yield RNA Synthesis Kit (NEB #E2040S) following manufacturer instructions. In vitro transcribed sgRNA products were purified using RNA Clean & Concentrator kit (Zymo Research #R1015). See also Supplementary Table S11.

CRISPR/Cas9 targeting of $CD34^+$ cells

PB CD34 cells were thawed and cultured in stem-cell-promoting media (IMDM + 20% Human Serum AB + hTPO (100 ng/ml), hSCF (100 ng/ml), hIL-6 (20 ng/ml), hFLT-3 (10 ng/ml) and non-essential amino acids) under low oxygen conditions to maximize maintenance and expansion of LT-HSCs (94-96). After 72 hours in culture, cells were then transfected with equimolar amounts of Cas9 protein (PNA Bio, 1 μg/μL in PBS) and sgRNAs using the Neon Transfection System (Thermo Fisher). Electroporation conditions used were 1600 volts, 10 milliseconds, and 3 pulses. A non-targeting sgRNA was used as control. Following transfection, cells were cultured in stem-cell-promoting media for 72 hrs. under low oxygen conditions (94-96) before sorting for CD34+ cells. After an additional 72 hrs. in culture, cells were used for RNA extraction, CFU assay, or plated in myeloid or erythroid differentiation media (see conditions below). Viability was assessed by trypan blue exclusion and was generally above 55% 48–72 hours post-transfection. Cutting was validated using the T7E1 assay as follows: PCR amplicon spanning the cleavage site was gradually hybridized in a thermal cycler. Hybridized fragments were then digested with 1.25 U of T7 endonuclease I (NEB) for up to 30 minutes at 37°C. Cleavage at heteroduplex mismatch sites were detected by agarose gel electrophoresis. Knock-out of KLF6 was also verified using flow-cytometry with an anti-KLF6 antibody (Millipore, #MABN119) and goat-anti-mouse IgG-AF594 (Thermo-Fisher, #A-11032) or goat-anti-mouse IgG-AF488 (Thermo-Fisher, #A-11059).

RNA-seq of LMNA knockdown

RNA from FACS isolated transduced CD34+ cells (n=8 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. For GSEA, the Wald statistic ranked list was used with the top (ranked by $\log_2$FoldChange) 500 genes up- and down-regulated with HSCe aging as gene sets.

RNA-seq of KLF6 knockdown

RNA from transfected CD34+ cells (n=3 biological replicates and 1–3 technical replicates for a total of n=4 sgCTRL and n=4 sgKLF6) was extracted using the Qiagen Allprep Micro kit according to manufacturer's instructions (Qiagen, #80204). 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 HiSeq-3000 with 75 bp paired-end sequencing. Data was aligned and processed as described above. For GSEA, the Wald statistic ranked list was used with the top (ranked by $\log_2$FoldChange) 500 genes up- and down-regulated with HSCe aging as gene sets as well as the c2.all.v6.2.symbols gene set using the weighted enrichment score.

Myeloid and erythroid liquid culture

Cells were plated and cultured for 7 days under myeloid-promoting conditions: SCF (100 ng/ml), FLT-3 ligands (10 ng/ml), IL-3 (20 ng/ml), IL-6 (20 ng/ml), GM-CSF (20 ng/ml), and G-CSF (20 ng/ml) and erythroid-promoting conditions: Epo (6 IU/ml) and SCF (100 ng/ml). At day 7 of myeloid and erythroid expansion, cells were stained for CD34 (BD Pharmigen, #304441), myeloid CD11b (BD Pharmigen, #301324), and erythroid CD71 (Pharmigen, #563769) and CD235a (BD Pharmigen, #559943) markers respectively, and also anti-KLF6 for cells transfected with sgKLF6 (Millipore, #MABN119)

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 the Western Blot analysis, a two-sided parametric unpaired t-test was used to calculate p-values with Prism software (v7.0).

DATA AND SOFTWARE AVAILIBILITY:

The ChIP-seq, ERRBS, and bulk RNA-seq data have been deposited in the NCBI Gene Expression Omnibus under ID code GSE104408. The scRNA-seq data has been deposited in the NCBI Gene Expression Omnibus under ID code GSE104379. Code used for analyses is available on Github: https://github.com/exa564/HSCe_Aging.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS:

E.R. Adelman was supported by the T32-AG11431 Biology of Aging grant to the University of Michigan and an award from the University of Michigan Department of Pathology Pilot Training Program in Translational Research. This work was supported by a Quest for Cures grant (0858–15 and R0858–18) as well as a Scholar Award (1357–19) from The Leukemia and Lymphoma Society to M.E. Figueroa. M.E. Figueroa, N. Salomonis, and H.L. Grimes were also supported by R01 CA196658. The authors would like to acknowledge the support of the tissue procurement services of the University of Michigan Comprehensive Cancer Center Tissue Core, the University of

Michigan Medical School (UMMS) DNA Sequencing Core, the UMMS Flow Cytometry Core, the University of Miami Sylvester Comprehensive Cancer Center (UM SCCC) Onco-Genomics Shared Resource and the UM SCCC Flow Cytometry Shared Resource as well as the Cincinnati Children's Hospital (CCH) Cell Processing Core which is supported by the CCH Research Foundation.

ERA was supported by the T32-AG11431 Biology of Aging grant to the University of Michigan and an award from the University of Michigan Department of Pathology Pilot Training Program in Translational Research. This work was supported by a Quest for Cures grant (0858–15 and R0858–18) as well as a Scholar Award (1357–19) from The Leukemia and Lymphoma Society to MEF. MEF, NS, and HLG were also supported by R01 CA196658. The authors declare no competing financial interests.

REFERENCES:

1. He W, Goodkind D, Kowal P. An Aging World: 2015. 2016.
2. Guralnik JM, Eisenstaedt RS, Ferrucci L, Klein HG, Woodman RC. Prevalence of anemia in persons 65 years and older in the United States: evidence for a high rate of unexplained anemia. Blood 2004;104(8):2263–8 doi 10.1182/blood-2004-05-1812. [PubMed: 15238427]
3. Aul C, Gattermann N, Schneider W. Age-related incidence and other epidemiological aspects of myelodysplastic syndromes. British journal of haematology 1992;82(2):358–67. [PubMed: 1419819]
4. Montecino-Rodriguez E, Berent-Maoz B, Dorshkind K. Causes, consequences, and reversal of immune system aging. The Journal of clinical investigation 2013;123(3):958–65 doi 10.1172/JCI64096. [PubMed: 23454758]
5. Kwok B, Hall JM, Witte JS, Xu Y, Reddy P, Lin K, et al. MDS-associated somatic mutations and clonal hematopoiesis are common in idiopathic cytopenias of undetermined significance. Blood 2015;126(21):2355–61 doi 10.1182/blood-2015-08-667063. [PubMed: 26429975]
6. Busque L, Patel JP, Figueroa ME, Vasanthakumar A, Provost S, Hamilou Z, et al. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. Nat Genet 2012;44(11):1179–81 doi 10.1038/ng.2413. [PubMed: 23001125]
7. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. New England Journal of Medicine 2014;371(26):2488–98 doi doi:10.1056/NEJMoa1408617. [PubMed: 25426837]
8. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. New England Journal of Medicine 2014;371(26):2477–87 doi doi:10.1056/NEJMoa1409405. [PubMed: 25426838]
9. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med 2014;20(12):1472–8 doi 10.1038/nm.3733. [PubMed: 25326804]
10. Jaiswal S, Natarajan P, Silver AJ, Gibson CJ, Bick AG, Shvartz E, et al. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. The New England journal of medicine 2017;377(2):111–21 doi 10.1056/NEJMoa1701719. [PubMed: 28636844]
11. Morrison SJ, Wandycz AM, Akashi K, Globerson A, Weissman IL. The aging of hematopoietic stem cells. Nat Med 1996;2(9):1011–6. [PubMed: 8782459]
12. Rossi DJ, Bryder D, Zahn JM, Ahlenius H, Sonu R, Wagers AJ, et al. Cell intrinsic alterations underlie hematopoietic stem cell aging. Proc Natl Acad Sci U S A 2005;102(26):9194–9 doi 10.1073/pnas.0503280102. [PubMed: 15967997]
13. Chambers SM, Shaw CA, Gatza C, Fisk CJ, Donehower LA, Goodell MA. Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation. PLoS biology 2007;5(8):e201 doi 10.1371/journal.pbio.0050201. [PubMed: 17676974]
14. Dykstra B, Olthof S, Schreuder J, Ritsema M, de Haan G. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. The Journal of experimental medicine 2011;208(13):2691–703 doi 10.1084/jem.20111490. [PubMed: 22110168]
15. Pang WW, Price EA, Sahoo D, Beerman I, Maloney WJ, Rossi DJ, et al. Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. Proc Natl Acad Sci U S A 2011;108(50):20012–7 doi 10.1073/pnas.1116110108. [PubMed: 22123971]

16. Mohrin M, Bourke E, Alexander D, Warr MR, Barry-Holson K, Le Beau MM, et al. Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis. Cell Stem Cell 2010;7(2):174–85 doi 10.1016/j.stem.2010.06.014. [PubMed: 20619762]
17. Rossi DJ, Bryder D, Seita J, Nussenzweig A, Hoeijmakers J, Weissman IL. Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. Nature 2007;447(7145):725–9 doi 10.1038/nature05862. [PubMed: 17554309]
18. Moehrle BM, Nattamai K, Brown A, Florian MC, Ryan M, Vogel M, et al. Stem Cell-Specific Mechanisms Ensure Genomic Fidelity within HSCs and upon Aging of HSCs. Cell reports 2015;13(11):2412–24 doi 10.1016/j.celrep.2015.11.030. [PubMed: 26686632]
19. Jang YY, Sharkis SJ. A low level of reactive oxygen species selects for primitive hematopoietic stem cells that may reside in the low-oxygenic niche. Blood 2007;110(8):3056–63 doi 10.1182/blood-2007-05-087759. [PubMed: 17595331]
20. Florian MC, Nattamai KJ, Dorr K, Marka G, Uberle B, Vas V, et al. A canonical to non-canonical Wnt signalling switch in haematopoietic stem-cell ageing. Nature 2013;503(7476):392–6 doi 10.1038/nature12631. [PubMed: 24141946]
21. Flach J, Bakker ST, Mohrin M, Conroy PC, Pietras EM, Reynaud D, et al. Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. Nature 2014;512(7513):198–202 doi 10.1038/nature13619. [PubMed: 25079315]
22. Wahlestedt M, Erlandsson E, Kristiansen T, Lu R, Brakebusch C, Weissman IL, et al. Clonal reversal of ageing-associated stem cell lineage bias via a pluripotent intermediate. Nature communications 2017;8:14533 doi 10.1038/ncomms14533.
23. Beerman I, Bock C, Garrison Brian S, Smith Zachary D, Gu H, Meissner A, et al. Proliferation-Dependent Alterations of the DNA Methylation Landscape Underlie Hematopoietic Stem Cell Aging. Cell Stem Cell 2013;12(4):413–25 doi 10.1016/j.stem.2013.01.017. [PubMed: 23415915]
24. Sun D, Luo M, Jeong M, Rodriguez B, Xia Z, Hannah R, et al. Epigenomic Profiling of Young and Aged HSCs Reveals Concerted Changes during Aging that Reinforce Self-Renewal. Cell Stem Cell 2014;14(5):673–88 doi 10.1016/j.stem.2014.03.002. [PubMed: 24792119]
25. Cheung P, Vallania F, Warsinske HC, Donato M, Schaffert S, Chang SE, et al. Single-Cell Chromatin Modification Profiling Reveals Increased Epigenetic Variations with Aging. Cell 2018;173(6):1385–97.e14 doi 10.1016/j.cell.2018.03.079. [PubMed: 29706550]
26. Bocker MT, Hellwig I, Breiling A, Eckstein V, Ho AD, Lyko F. Genome-wide promoter DNA methylation dynamics of human hematopoietic progenitor cells during differentiation and aging. Blood 2011;117(19):e182–9 doi 10.1182/blood-2011-01-331926. [PubMed: 21427290]
27. Bacalini MG, Boattini A, Gentilini D, Giampieri E, Pirazzini C, Giuliani C, et al. A meta-analysis on age-associated changes in blood DNA methylation: results from an original analysis pipeline for Infinium 450k data. Aging 2015;7(2):97–109. [PubMed: 25701668]
28. Yamazaki H, Suzuki M, Otsuki A, Shimizu R, Bresnick EH, Engel JD, et al. A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. Cancer Cell 2014;25(4):415–27 doi 10.1016/j.ccr.2014.02.008. [PubMed: 24703906]
29. Mansour MR, Abraham BJ, Anders L, Berezovskaya A, Gutierrez A, Durbin AD, et al. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science 2014;346(6215):1373–7 doi 10.1126/science.1259037. [PubMed: 25394790]
30. Herranz D, Ambesi-Impiombato A, Palomero T, Schnell SA, Belver L, Wendorff AA, et al. A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. Nat Med 2014;20(10):1130–7 doi 10.1038/nm.3665. [PubMed: 25194570]
31. Groschel S, Sanders MA, Hoogenboezem R, de Wit E, Bouwman BA, Erpelinck C, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell 2014;157(2):369–81 doi 10.1016/j.cell.2014.02.019. [PubMed: 24703711]
32. Glass JL, Hassane D, Wouters BJ, Kunimoto H, Avellino R, Garrett-Bakelman FE, et al. Epigenetic Identity in AML Depends on Disruption of Nonpromoter Regulatory Elements and Is Affected by Antagonistic Effects of Mutations in Epigenetic Modifiers. Cancer discovery 2017;7(8):868–83 doi 10.1158/2159-8290.Cd-16-1032. [PubMed: 28408400]

33. Rundberg Nilsson A, Soneji S, Adolfsson S, Bryder D, Pronk CJ. Human and Murine Hematopoietic Stem Cell Aging Is Associated with Functional Impairments and Intrinsic Megakaryocytic/Erythroid Bias. PLoS One 2016;11(7):e0158369 doi 10.1371/journal.pone.0158369. [PubMed: 27368054]
34. Pang WW, Schrier SL, Weissman IL. Age-associated changes in human hematopoietic stem cells. Semin Hematol 2017;54(1):39–42 doi 10.1053/j.seminhematol.2016.10.004. [PubMed: 28088986]
35. Akalin A, Garrett-Bakelman FE, Kormaksson M, Busuttil J, Zhang L, Khrebtukova I, et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigeneticlandscapes in acute myeloid leukemia. PLoS Genet 2012;8 doi 10.1371/journal.pgen.1002781.
36. Beck D, Thoms JAI, Perera D, Schütte J, Unnikrishnan A, Knezevic K, et al. Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. Blood 2013;122(14):e12–e22 doi 10.1182/blood-2013-03-490425. [PubMed: 23974199]
37. Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, et al. Epigenetic stem cell signature in cancer. Nat Genet 2007;39(2):157–8 doi 10.1038/ng1941. [PubMed: 17200673]
38. Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet 2007;39(2):232–6 doi 10.1038/ng1950. [PubMed: 17200670]
39. Florian MC, Dorr K, Niebel A, Daria D, Schrezenmeier H, Rojewski M, et al. Cdc42 activity regulates hematopoietic stem cell aging and rejuvenation. Cell Stem Cell 2012;10(5):520–30 doi 10.1016/j.stem.2012.04.007. [PubMed: 22560076]
40. Crews LA, Balaian L, Delos Santos NP, Leu HS, Court AC, Lazzari E, et al. RNA Splicing Modulation Selectively Impairs Leukemia Stem Cell Maintenance in Secondary Human AML. Cell Stem Cell 2016;19(5):599–612 doi 10.1016/j.stem.2016.08.003. [PubMed: 27570067]
41. Yang J, Huang T, Petralia F, Long Q, Zhang B, Argmann C, et al. Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. Scientific reports 2015;5:15145 doi 10.1038/srep15145. [PubMed: 26477495]
42. DePasquale EAK, Ferchen K, Hay S, Muench DE, Grimes HL, Salomonis N. CellHarmony: Cell-level matching and comparison of single-cell transcriptomes. bioRxiv 2018.
43. Hay SB, Ferchen K, Chetal K, Grimes HL, Salomonis N. The Human Cell Atlas bone marrow single-cell interactive web portal. Experimental hematology 2018;68:51–61 doi 10.1016/j.exphem.2018.09.004. [PubMed: 30243574]
44. Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function,and impair hematopoietic differentiation. Cancer Cell 2010;18 doi 10.1016/j.ccr.2010.11.015.
45. Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer Cell 2010;17 doi 10.1016/j.ccr.2009.11.020.
46. McKeown MR, Corces MR, Eaton ML, Fiore C, Lee E, Lopez JT, et al. Superenhancer Analysis Defines Novel Epigenomic Subtypes of Non-APL AML, Including an RARalpha Dependency Targetable by SY-1425, a Potent and Selective RARalpha Agonist. Cancer discovery 2017;7(10):1136–53 doi 10.1158/2159-8290.cd-17-0399. [PubMed: 28729405]
47. Yi G, Wierenga ATJ, Petraglia F, Narang P, Janssen-Megens EM, Mandoli A, et al. Chromatin-Based Classification of Genetically Heterogeneous AMLs into Two Distinct Subtypes with Diverse Stemness Phenotypes. Cell reports 2019;26(4):1059–69 e6 doi 10.1016/j.celrep.2018.12.098. [PubMed: 30673601]
48. Kim GD, Das R, Goduni L, McClellan S, Hazlett LD, Mahabeleshwar GH. Kruppel-like Factor 6 Promotes Macrophage-mediated Inflammation by Suppressing B Cell Leukemia/Lymphoma 6 Expression. J Biol Chem 2016;291(40):21271–82 doi 10.1074/jbc.M116.738617. [PubMed: 27539853]
49. Matsumoto N, Kubo A, Liu H, Akita K, Laub F, Ramirez F, et al. Developmental regulation of yolk sac hematopoiesis by Krüppel-like factor 6. Blood 2006;107(4):1357 doi 10.1182/blood-2005-05-1916. [PubMed: 16234353]

50. Desai P, Mencia-Trinchant N, Savenkov O, Simon MS, Cheang G, Lee S, et al. Somatic mutations precede acute myeloid leukemia years before diagnosis. Nature medicine 2018;24(7):1015–23 doi 10.1038/s41591-018-0081-z.
51. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, et al. Enhancer evolution across 20 mammalian species. Cell 2015;160(3):554–66 doi 10.1016/j.cell.2015.01.006. [PubMed: 25635462]
52. 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 1997;89(5):1543–50. [PubMed: 9057635]
53. Will B, Vogler TO, Narayanagari S, Bartholdy B, Todorova TI, da Silva Ferreira M, et al. Minimal PU.1 reduction induces a preleukemic state and promotes development of acute myeloid leukemia. Nat Med 2015;21(10):1172–81 doi 10.1038/nm.3936. [PubMed: 26343801]
54. Chapuy B, McKeown MR, Lin CY, Monti S, Roemer MG, Qi J, et al. Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. Cancer Cell 2013;24(6):777–90 doi 10.1016/j.ccr.2013.11.003. [PubMed: 24332044]
55. Ergen AV, Boles NC, Goodell MA. Rantes/Ccl5 influences hematopoietic stem cell subtypes and causes myeloid skewing. Blood 2012;119(11):2500–9 doi 10.1182/blood-2011-11-391730. [PubMed: 22289892]
56. Guidi N, Sacma M, Standker L, Soller K, Marka G, Eiwen K, et al. Osteopontin attenuates aging-associated phenotypes of hematopoietic stem cells. Embo j 2017;36(7):840–53 doi 10.15252/embj.201694969. [PubMed: 28254837]
57. Poulos MG, Ramalingam P, Gutkin MC, Llanos P, Gilleran K, Rabbany SY, et al. Endothelial transplantation rejuvenates aged hematopoietic stem cell function. The Journal of clinical investigation 2017;127(11):4163–78 doi 10.1172/jci93940. [PubMed: 29035282]
58. Martin M Cutadapt removes adapter sequences from high-throughput sequencing reads EMBnetjournal; Vol 17, No 1: Next Generation Sequencing Data Analysis 2011 doi 10.14806/ej.17.1.200.
59. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 2009;10(3):R25 doi 10.1186/gb-2009-10-3-r25. [PubMed: 19261174]
60. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome biology 2008;9(9):R137 doi 10.1186/gb-2008-9-9-r137. [PubMed: 18798982]
61. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25(16):2078–9 doi 10.1093/bioinformatics/btp352. [PubMed: 19505943]
62. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic acids research 2016;44(W1):W160–W5 doi 10.1093/nar/gkw257. [PubMed: 27079975]
63. Warnes GR, Bolker B, Lumley T. gtools: Various R Programming Tools. R package version 3.8.1.
64. Welch RP, Lee C, Imbriano PM, Patil S, Weymouth TE, Smith RA, et al. ChIP-Enrich: gene set enrichment testing for ChIP-seq data. Nucleic Acids Res 2014;42(13):e105 doi 10.1093/nar/gku463. [PubMed: 24878920]
65. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98(9):5116–21 doi 10.1073/pnas.091062498. [PubMed: 11309499]
66. Akalin A, Franke V, Vlahovicek K, Mason CE, Schubeler D. Genomation: a toolkit to summarize, annotate and visualize genomic intervals. Bioinformatics 2015;31(7):1127–9 doi 10.1093/bioinformatics/btu775. [PubMed: 25417204]
67. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32(18):2847–9 doi 10.1093/bioinformatics/btw313. [PubMed: 27207943]
68. Kassambara A, Mundt F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. 1.0.5.9992017.

69. Wickham H ggplot2: elegant graphics for data analysis. New York: Springer; 2009.
70. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 2011;27(11):1571–2 doi 10.1093/bioinformatics/btr167. [PubMed: 21493656]
71. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome biology 2012;13 doi 10.1186/gb-2012-13-10-r87.
72. Li S, Garrett-Bakelman FE, Akalin A, Zumbo P, Levine R, To BL, et al. An optimized algorithm for detecting and annotating regional differential methylation. BMC bioinformatics 2013;14 Suppl 5:S10 doi 10.1186/1471-2105-14-s5-s10.
73. Meldi K, Qin T, Buchi F, Droin N, Sotzen J, Micol JB, et al. Specific molecular signatures predict decitabine response in chronic myelomonocytic leukemia. The Journal of clinical investigation 2015;125(5):1857–72 doi 10.1172/jci78752. [PubMed: 25822018]
74. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols 2009;4(1):44–57 doi 10.1038/nprot.2008.211. [PubMed: 19131956]
75. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009;37(1):1–13 doi 10.1093/nar/gkn923. [PubMed: 19033363]
76. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29(1):15–21 doi 10.1093/bioinformatics/bts635. [PubMed: 23104886]
77. Hartley SW, Mullikin JC. QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. BMC bioinformatics 2015;16(1):224 doi 10.1186/s12859-015-0670-5. [PubMed: 26187896]
78. Anders S, Huber W. Differential expression analysis for sequence count data. Genome biology 2010;11(10):R106 doi 10.1186/gb-2010-11-10-r106. [PubMed: 20979621]
79. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences 2005;102(43):15545–50 doi 10.1073/pnas.0506580102.
80. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics 2011;12:323 doi 10.1186/1471-2105-12-323. [PubMed: 21816040]
81. Emig D, Salomonis N, Baumbach J, Lengauer T, Conklin BR, Albrecht M. AltAnalyze and DomainGraph: analyzing and visualizing exon expression data. Nucleic Acids Res 2010;38(Web Server issue):W755–62 doi 10.1093/nar/gkq405. [PubMed: 20513647]
82. Li Q, Bohin N, Wen T, Ng V, Magee J, Chen SC, et al. Oncogenic Nras has bimodal effects on stem cells that sustainably increase competitiveness. Nature 2013;504(7478):143–7 doi 10.1038/nature12830. [PubMed: 24284627]
83. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 2010;38(4):576–89 doi 10.1016/j.molcel.2010.05.004. [PubMed: 20513432]
84. Lindsley RC, Saber W, Mar BG, Redd R, Wang T, Haagenson MD, et al. Prognostic Mutations in Myelodysplastic Syndrome after Stem-Cell Transplantation. The New England journal of medicine 2017;376(6):536–47 doi 10.1056/NEJMoa1611604. [PubMed: 28177873]
85. Lippert RA. Space-efficient whole genome comparisons with Burrows-Wheeler transforms. J Comput Biol 2005;12 doi 10.1089/cmb.2005.12.407.
86. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20(9):1297–303 doi 10.1101/gr.107524.110. [PubMed: 20644199]
87. Wilm A, Aw PP, Bertrand D, Yeo GH, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput

sequencing datasets. Nucleic Acids Res 2012;40(22):11189–201 doi 10.1093/nar/gks918. [PubMed: 23066108]

88. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38(16):e164 doi 10.1093/nar/gkq603. [PubMed: 20601685]
89. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res 2014;42(22):13534–44 doi 10.1093/nar/gku1206. [PubMed: 25416802]
90. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature 2010;467(7319):1061–73 doi 10.1038/nature09534. [PubMed: 20981092]
91. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 2011;39(Database issue):D945–50 doi 10.1093/nar/gkq929. [PubMed: 20952405]
92. Moreno-Mateos MA, Vejnar CE, Beaudoin JD, Fernandez JP, Mis EK, Khokha MK, et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat Methods 2015;12(10):982–8 doi 10.1038/nmeth.3543. [PubMed: 26322839]
93. Gundry MC, Brunetti L, Lin A, Mayle AE, Kitano A, Wagner D, et al. Highly Efficient Genome Editing of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9. Cell reports 2016;17(5):1453–61 doi 10.1016/j.celrep.2016.09.092. [PubMed: 27783956]
94. Danet GH, Pan Y, Luongo JL, Bonnet DA, Simon MC. Expansion of human SCID-repopulating cells under hypoxic conditions. The Journal of clinical investigation 2003;112(1):126–35 doi 10.1172/jci17669. [PubMed: 12840067]
95. Koller MR, Bender JG, Miller WM, Papoutsakis ET. Expansion of primitive human hematopoietic progenitors in a perfusion bioreactor system with IL-3, IL-6, and stem cell factor Bio/technology (Nature Publishing Company) 1993;11(3):358–63. [PubMed: 7680209]
96. Bak RO, Porteus MH. CRISPR-mediated Integration of Large Gene Cassettes using AAV Donor Vectors. Cell reports 2017;20(3):750–6 doi 10.1016/j.celrep.2017.06.064. [PubMed: 28723575]

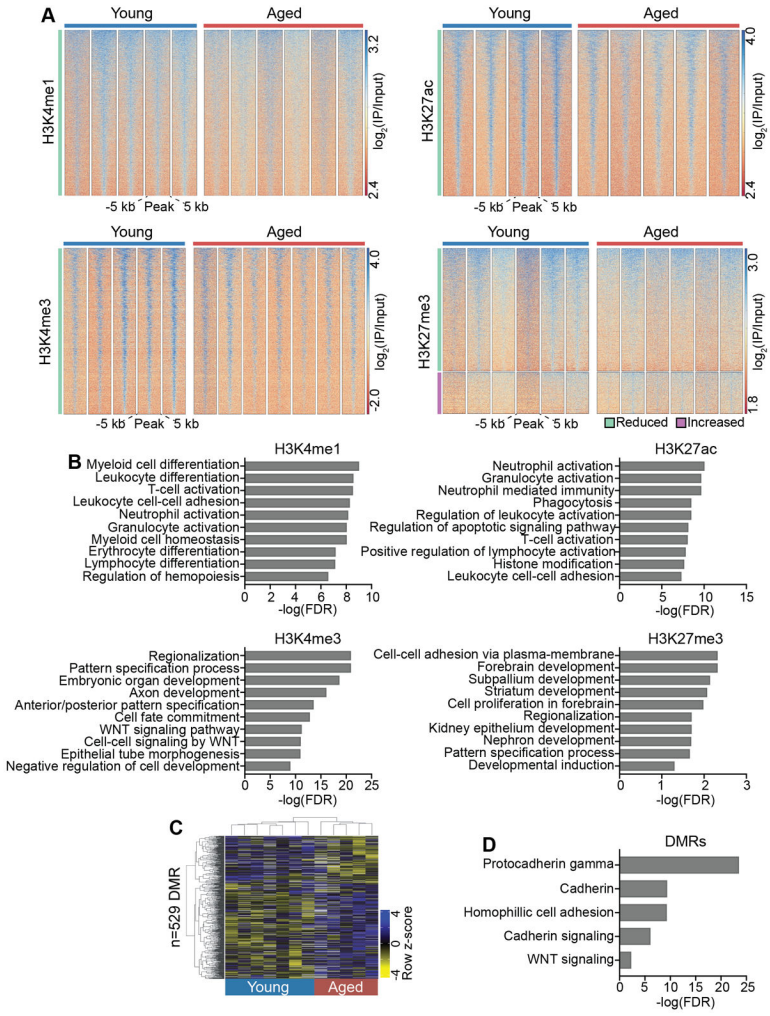

Fig. 1: Focal histone and DNA methylation alterations with HSCe aging.

(A) Heatmap representation of regions with either loss or gain (log_{10} likelihood ratio > 3, absolute fold-change >1.5) 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 donor. **(B)** Functional annotation using ChIP-enrich Gene Ontology Biological Processes for 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. **(C)** Row-scaled 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 corresponds to a unique region with differential methylation and each column corresponds to one donor. **(D)** Bar plot representation of significant (FDR <0.05) pathways associated with genes that are differentially methylated in aged HSCe.

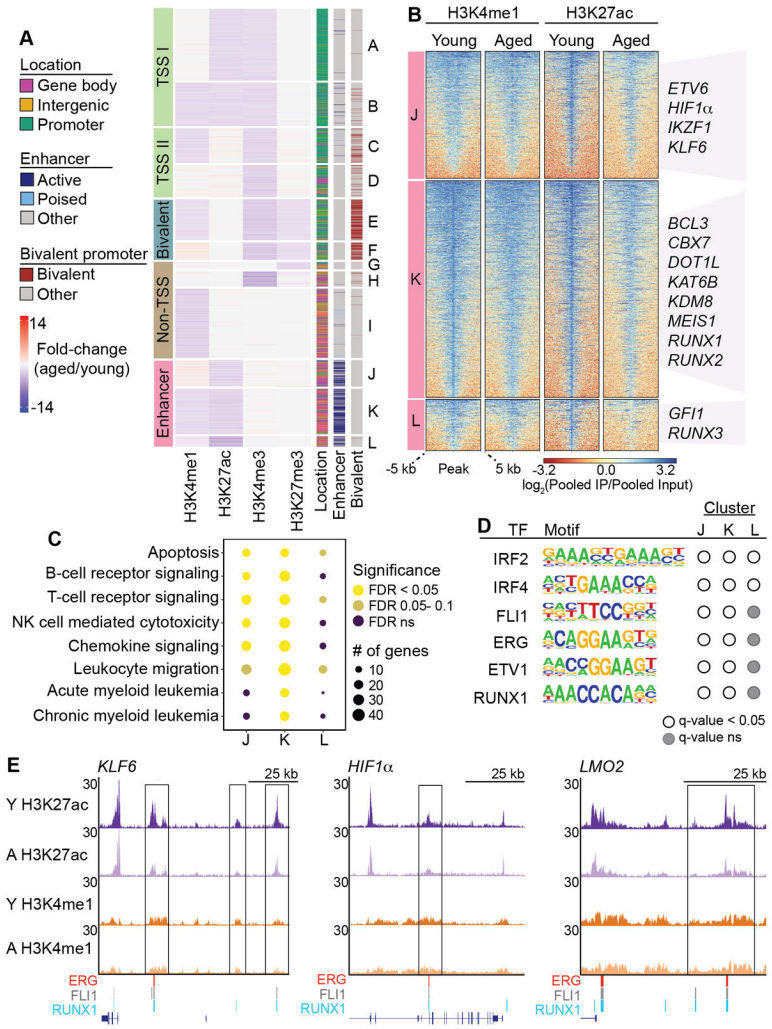

Fig. 2: Decreased H3K27ac at immune and cancer associated enhancers with age.

(A) Heatmap depicting the 12 clusters identified using k-means clustering of regions with significant (LLR > 3 and absolute fold-change > 1.5) changes of H3K4me1, H3K4me3, H3K27me3, or H3K27ac with age (n=37,058 peaks). The fold-change(aged/young) signal is plotted for each histone modification for each peak. Annotation to active and poised enhancers as well as bivalent promoters identified in young HSCe is also shown. **(B)** Heatmap of H3K4me1 and H3K27ac signal at the enhancer enriched clusters J-L. The $\log_2$(Pooled IP/Pooled Input) signal is plotted for each age group, centered on the differential peak +/− 5kb. Select genes within each cluster are denoted on the right of the heatmap. **(C)** Bubble plot representation of select KEGG pathways that are enriched in clusters J-L. The size of each bubble corresponds to the number of genes within the cluster that are within the given gene set. Highly significant (FDR <0.05) categories are colored in yellow. **(D)** Example of transcription factor binding DNA motifs enriched in clusters J-L. Motifs with significant enrichment (q-value < 0.05) are denoted with a white circle. **(E)** UCSC genome browser track examples of active enhancers within clusters J-L that overlap transcription factor ChIP-Seq peaks. Tracks are of pooled replicates for each age group,

normalized to reads per million and to the corresponding Input. Lines below tracks represent transcription factor peaks annotated in publicly available datasets of CD34+ cells.

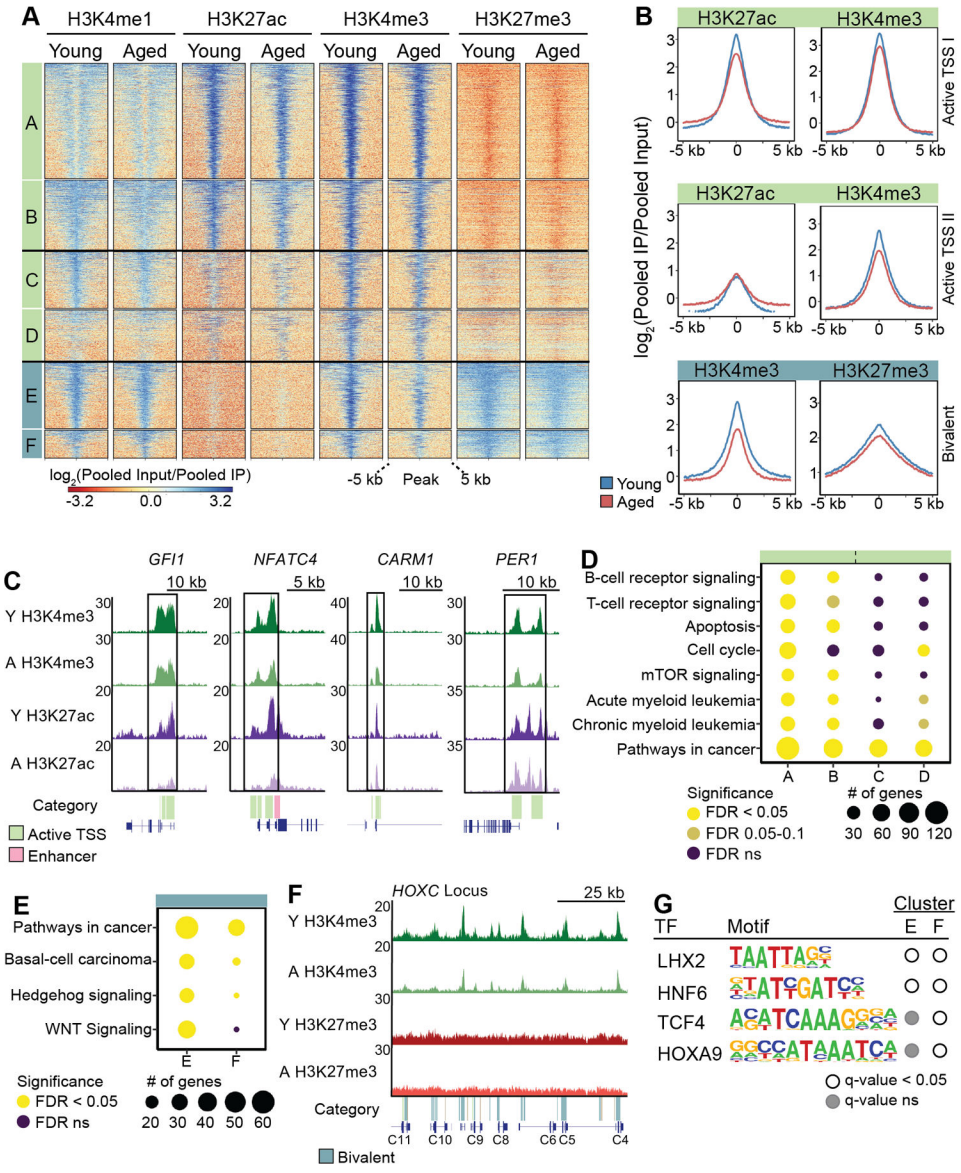

Fig. 3: Loss of activating histone modifications at promoter regions with age.

(A) Heatmap of H3K4me1, H3K27ac, H3K4me3 and H3K27me3 signals at age-associated clusters that are enriched for active promoters *(top and middle)* and bivalent promoters *(bottom)*. The $\log_2$(Pooled IP/Pooled Input) signal is plotted for each age group, centered on the differential peak +/− 5kb. **(B)** Density plots of the $\log_2$(Pooled IP/Pooled Input) signal for the characteristic histone marks for the peaks within the Active TSS I (*top*), Active TSS II (*middle*) and Bivalent promoter (*bottom*) categories. **(C)** UCSC genome browser tracks of genes with altered promoters from the Active TSS I cluster (*GFI1* and *NFATC4*) and Active TSS II cluster (*CARM1* and *PER1*). Tracks are of pooled replicates for each age group, normalized to reads per million and to the corresponding Input for ChIP-seq. Light green bars below tracks represent the differential promoter regions identified from the cluster analysis. **(D)** Bubble plot representation of select KEGG pathways that are enriched in the active promoter clusters (clusters A-D). The size of each bubble corresponds to the number

of genes within the cluster that are within the given gene set. Highly significant (FDR <0.05) categories are colored in yellow. **(E)** Bubble plot representation of select KEGG pathways that are enriched in clusters E and F. The size of each bubble corresponds to the number of genes within the cluster that are within the given gene set. Highly significant (FDR <0.05) categories are colored in yellow. **(F)** UCSC genome browser tracks of the HOXC cluster, which contains bivalent promoters within clusters E-F that are altered with age. Tracks are of pooled replicates for each age group, normalized to reads per million and to the corresponding Input for ChIP-seq. Teal bars below tracks represent the differential bivalent promoter regions identified from the cluster analysis. **(G)** Example of transcription factor binding DNA motifs enriched in clusters E and F. Motifs with significant enrichment (q-value < 0.05) are denoted with a white circle.

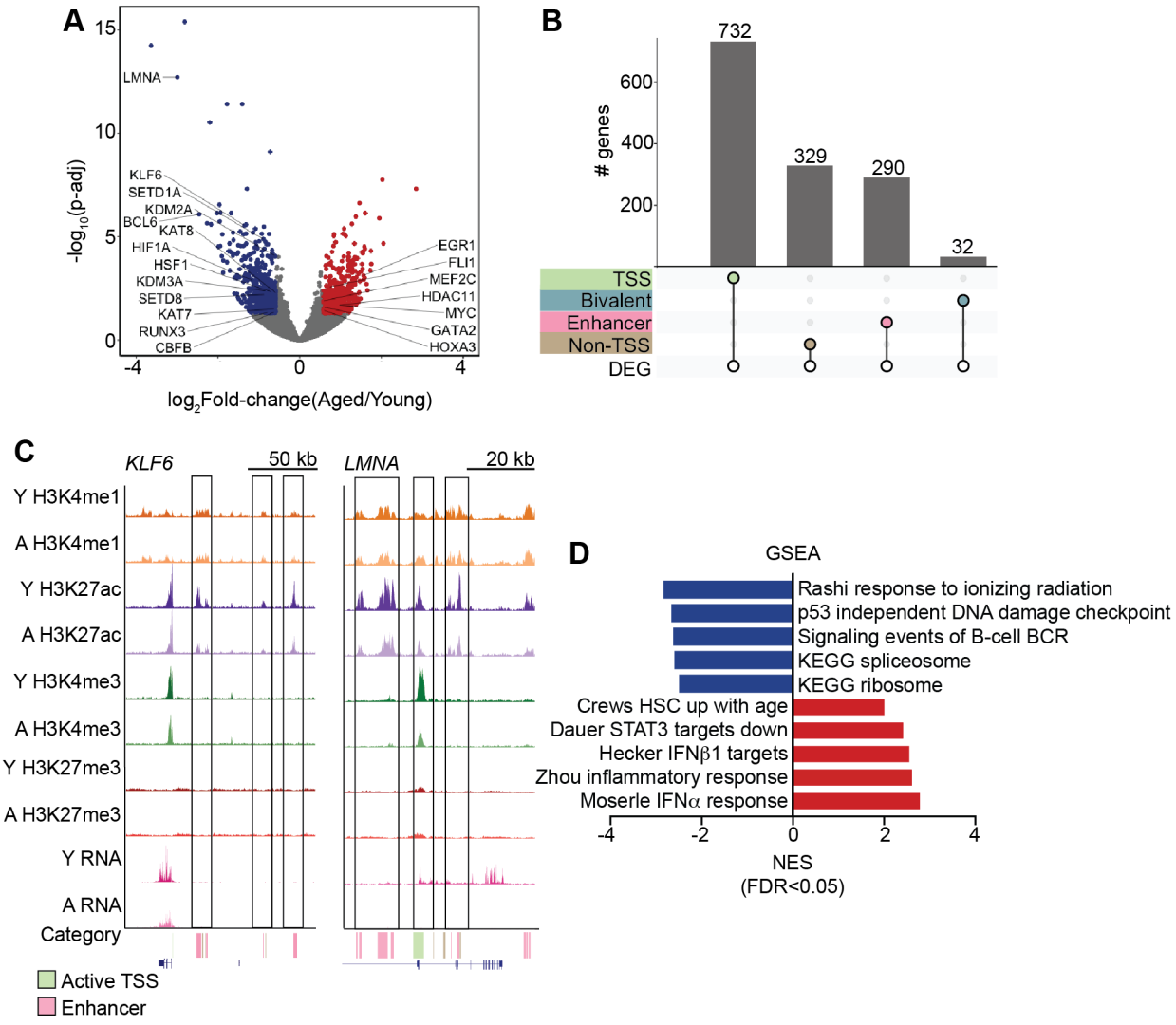

Fig. 4: Differential gene expression of transcription factors and epigenetic modifiers in aged HSCe.

(A) Volcano plot of the $\log_2$ fold-change (aged/young) gene expression versus the $-\log_{10}$(p-adj). Significant downregulated and upregulated genes in aged HSCe compared to young are colored in blue and red, respectively (FDR < 0.05 and fold-change< −1.5 or fold-change > 1.5, respectively). Select differentially expressed genes are labeled. **(B)** Bar plot depicting the number of age-associated differential genes that also epigenetically deregulated with age at the different genomic regions identified in Figure 2. The numbers above the bars denote the number of genes that are both differentially expressed and within the epigenomic category. **(C)** UCSC tracks of pooled replicates for young and aged HSCe at the *KLF6* and *LMNA* loci. Colored bars below the tracks denote differential epigenetic regions as identified in Figure 2A. **(D)** Bar plot representation of the normalized enrichment score (NES) for select gene set enrichment pathways that are up or downregulated with HSCe aging (n=10 donors per age group; FDR< 0.05).

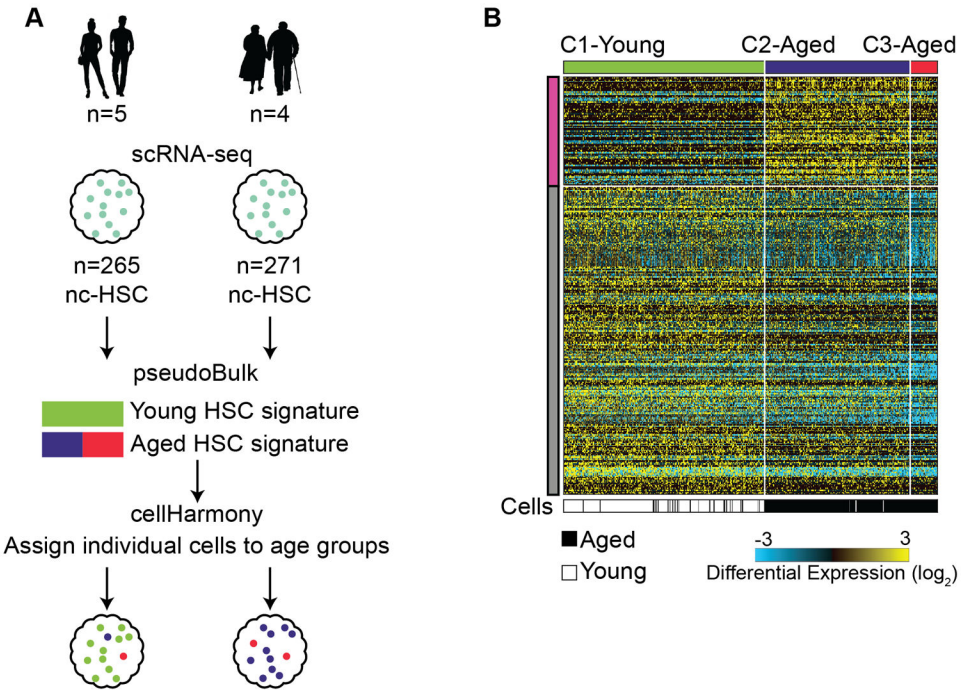

Fig. 5: Age-associated changes in epigenetic landscapes are driven by reprogramming.

(A) Schematic of the analysis used to identify and classify aged HSC using scRNAseq data from young (24-37 yo) and aged (64-71 yo) HSCe. **(B)** Heatmap of young and aged nc-HSC. Expression centroids were calculated from the 3 clusters (C1-Young in green, C2-Aged in purple, and C3-Aged in red) identified using the genes differentially expressed at the single-cell level (shown in Figure S5B) and used to cluster nc-HSC. Each column corresponds to an individual cell, and age group of the donor is depicted in the bottom bar.

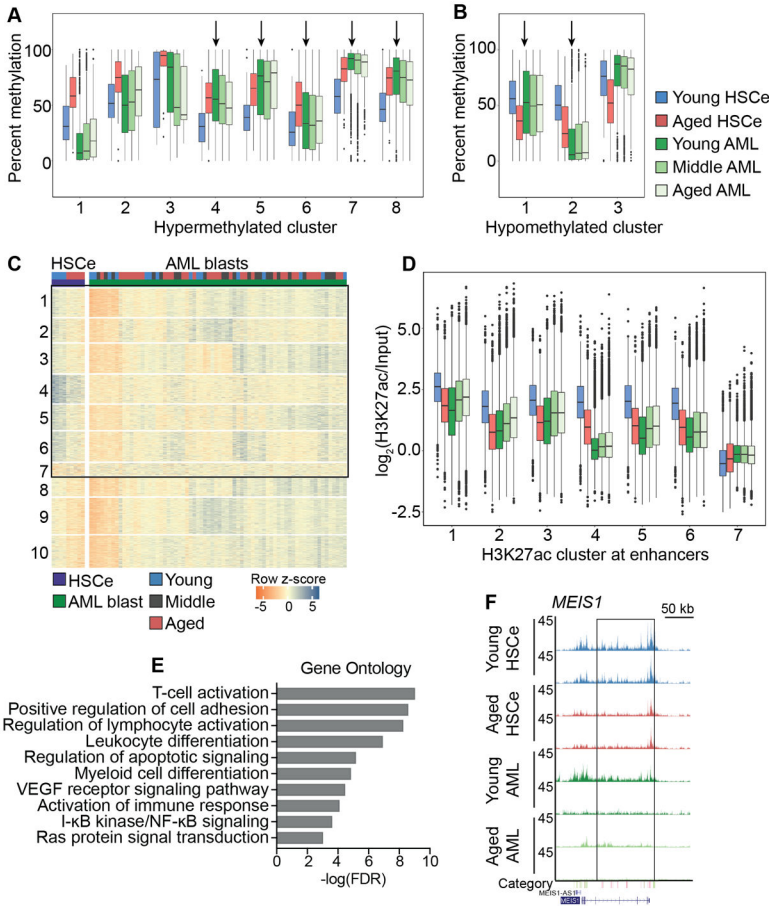

Fig. 6: Age-associated epigenetic changes may predispose for AML.

(A & B) Boxplots of the percent methylation for k-means clusters of regions that are **(A)** hypermethylated (k=8) or **(B)** hypomethylated (k=3) 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 ($p< 0.05$, Mann-Whitney rank sum test, and corrected for multiple testing). **(C)** Heatmap of H3K27ac signal at peaks within the enhancer category that is altered with HSCe aging. Rows are ordered using k-means clustering (k=10) performed using young and aged HSCe and AML blasts (n=71 patients). Bars above heatmap denote donor age and cell type. **(D)** Boxplots of the $\log_2$(H3K27ac/Input) enrichment for peaks (n=4,931) within clusters 1-7 (C) which showed consistent changes in H3K27ac in aged HSCe and AML compared to young HSCe ($p < 0.05$, Mann-Whitney rank-sum test, and corrected for multiple testing). **(E)** Select significant (FDR <0.05) gene ontology biological processes associated with peaks in clusters 1–7 (C). **(F)** UCSC genome browser tracks of the read-normalized fold-enrichment (H3K27ac/Input) signal at *MEIS1,* a gene predicted by k-means clustering to have reduced H3K27ac at its intergenic enhancer. Each track is derived from 1 donor and 2 biological replicates are shown for each condition.

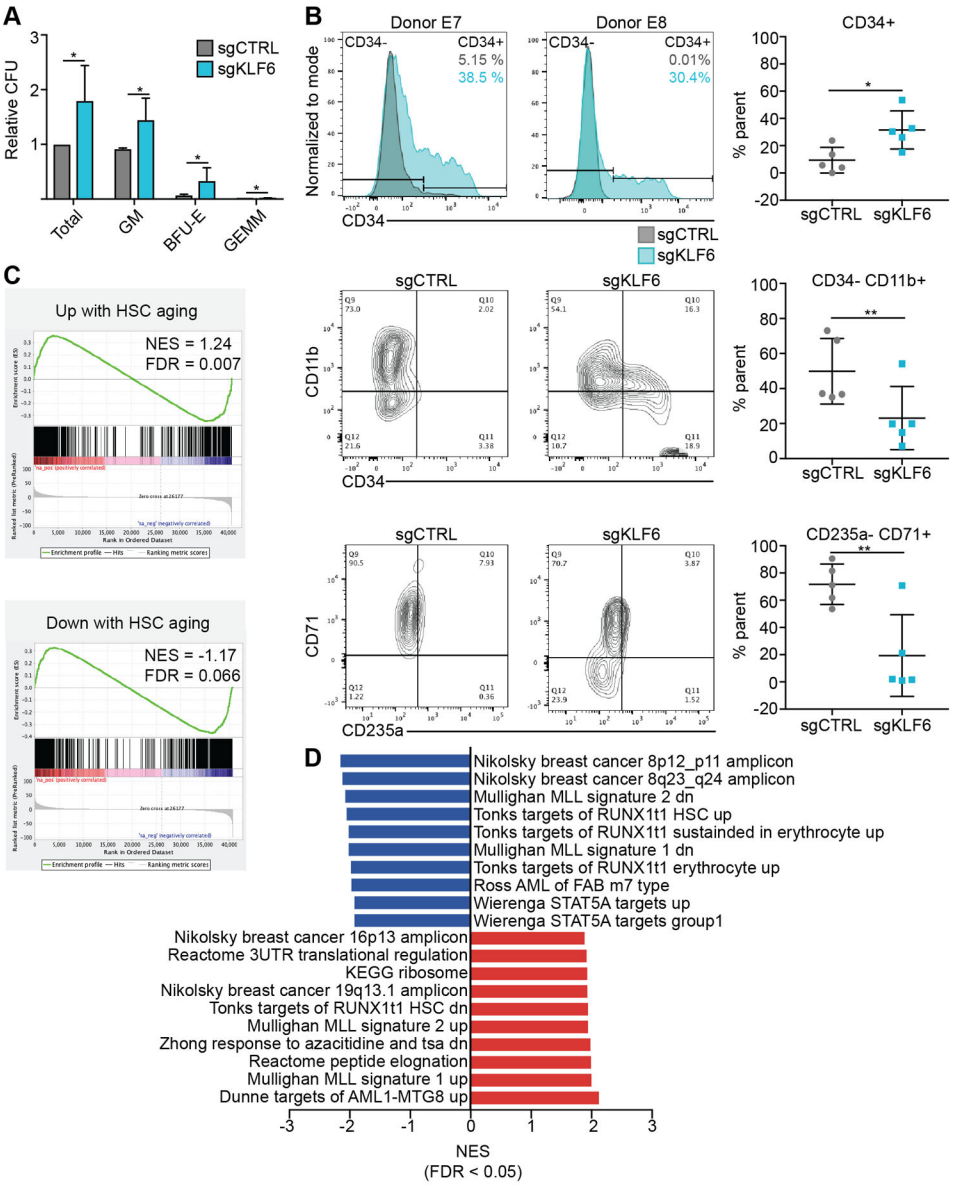

Fig. 7: Loss of KLF6 impairs differentiation and leads to expression changes reminiscent of leukemia.

(A) Colony-forming unit assays of CD34+ cells with CRISPR-Cas9 knockout of KLF6. Normalized colony numbers per 500 CD34+ cells plated are plotted for total colony number, granulocyte-macrophage (GM), granulocyte-erythrocyte-macrophage-megakaryocyte (GEMM) and burst-forming unit erythroid (BFU-E). Colony numbers for each biological replicate (n=5) are normalized to the total colony number for that replicate. **(B)** Representative flow cytometry histograms for 2 donors and contour plots from 1 donor for CD34+ cells transfected with sgCTRL or sgKLF6 and cultured in myeloid (*top and middle*) or erythroid (*bottom*) promoting conditions for 7 days. Dot plot representation of the percentage of CD34+ cells (top) CD34− CD11b+ (*middle*) and CD235a− CD71+ (*bottom*) cells is also shown (n=5). Cells transfected with sgCTRL were gated on KLF6+, while cells transfected with sgKLF6 were gated on KLF6−. **(C)** GSEA leading edge plots showing the

enrichment of the gene sets for genes up- or down-regulated with HSCe aging in CD34+ cells with sgKLF6 knockout (n=4 replicates). GSEA was ran using a list pre-ranked by the Wald-statistic (DESeq2), with the weighted enrichment score. **(D)** Bar plot representation of the normalized enrichment score (NES) for the topmost gene set enrichment pathways that are up- or down-regulated with sgKLF6 knockout (FDR< 0.05). p-values from paired one-tailed t-test, calculated with Prism, p>0.05=ns, p ≤ 0.05=*, p≤0.01=**, p≤0.001=***, p≤0.0001=****.