

DNMT3A^{R882H} Is Not Required for Disease Maintenance in Primary Human AML, but Is Associated With Increased Leukemia Stem Cell Frequency

5 Thomas Köhnke^{1,†}, Daiki Karigane^{1,†}, Eleanor Hilgart², Amy C. Fan¹, Kensuke Kayamori¹, Masashi Miyauchi¹, Cailin T. Collins¹, Fabian P. Suchy¹, Athreya Rangavajhula¹, Yang Feng¹, Yusuke Nakauchi¹, Eduardo Martinez-Montes², Jonas L. Fowler³, Kyle M. Loh³, Hiromitsu Nakauchi¹, Michael A. Koldobskiy², Andrew P. Feinberg², and Ravindra Majeti^{1,*}

10 ¹Department of Medicine, Division of Hematology, Cancer Institute, and Institute for Stem Cell Biology and Regenerative Medicine, Stanford University; Stanford, CA, 94305, USA.

²Center for Epigenetics and Department of Medicine, Johns Hopkins University School of Medicine; Baltimore, MD, 21205, USA.

15 ³Department of Developmental Biology and Institute for Stem Cell Biology and Regenerative Medicine, Stanford University; Stanford, CA, 94305, USA

[†]These authors contributed equally to this work

*Corresponding author. Email: rmajeti@stanford.edu

DNMT3A^{R882H} Is Not Required for Disease Maintenance in Primary Human AML, but Is Associated With Increased Leukemia Stem Cell Frequency

ABSTRACT

5 Genetic mutations are being thoroughly mapped in human cancers, yet a fundamental question in cancer biology is whether such mutations are functionally required for cancer initiation, maintenance of established cancer, or both. Here, we study this question in the context of human acute myeloid leukemia (AML), where *DNMT3A*^{R882} missense mutations often arise early, in pre-leukemic clonal hematopoiesis, and corrupt the DNA methylation landscape to initiate leukemia. We developed CRISPR-based methods to directly correct *DNMT3A*^{R882} mutations in leukemic cells obtained from patients. Surprisingly, *DNMT3A*^{R882} mutations were largely dispensable for disease maintenance. Replacing *DNMT3A*^{R882} mutants with wild-type *DNMT3A* did not impair the ability of AML cells to engraft *in vivo*, and minimally altered DNA methylation. Taken together, *DNMT3A*^{R882} mutations are initially necessary for AML initiation, but are largely dispensable for disease maintenance. The notion that initiating oncogenes differ from those that maintain cancer has important implications for cancer evolution and therapy. 15

STATEMENT OF SIGNIFICANCE

20 Understanding which driver mutations are required for cancer initiation, maintenance, or both phases remains poorly understood. Here, we uncover that highly prevalent pre-leukemic *DNMT3A* mutations are only required during disease initiation, but become dispensable after leukemic transformation, uncovering the context-specific role of this driver mutation with important therapeutic implications.

INTRODUCTION

An important and understudied question in cancer pathogenesis is whether a given genetic mutation is required for disease initiation, maintenance of established disease, or both. In myeloid malignancies, previous studies in mouse models interrogating the role of the driver mutations *IDH2*^{R140Q} and *JAK2*^{V617F} have suggested that their continued presence is required for both disease initiation and maintenance.^{1,2} Likewise, within *in vivo* mouse models of colorectal cancer, correction of the *Apc* driver mutation in established cancer cells leads to disease regression, again suggesting that the driver mutation is required for both cancer initiation and maintenance.³ However, whether this principle extends to all genetic mutations found in cancer, especially those acquired early during pre-malignant evolution, is unclear. Delineating the specific contribution of individual mutations across the disease course directly in patient specimens is hampered by substantial variability in co-mutations, germline background, inter-patient variability, and technical challenges in molecularly manipulating primary human cells. Thus, prospective, functional interrogation of how specific somatic mutations affect these cellular phenotypes, as well as the biology of cancer stem cells, have been performed in surrogate systems such as mouse disease models and cell lines. To overcome these limitations and explore this question in human acute myeloid leukemia (AML), we developed an approach to perform functional genetic experiments directly in AML patient specimens by leveraging CRISPR/Cas9 to selectively and specifically correct individual somatic mutations and assay the effects on disease initiation, disease maintenance and leukemia stem cell phenotypes.

We focused on *DNMT3A*^{R882} missense mutations, which are among the most frequently occurring founding mutations in human clonal hematopoiesis and acute myeloid leukemia (AML).⁴⁻⁶ Mutations in *DNMT3A* impair its ability to methylate DNA CpGs⁷⁻⁹, leading to increased self-renewal of hematopoietic stem cells (HSCs),¹⁰ and have been associated with poor treatment outcomes in AML.^{4,11} During disease evolution, *DNMT3A*^{R882} missense mutations are frequently acquired early, preceding leukemic transformation,¹² suggesting a critical role in disease initiation. However, it is possible that such mutations might be required to provide the epigenetic background necessary for disease initiation but may no longer be required for maintenance of established disease in the presence of other somatic driver mutations. Since current approaches are unable to discern whether *DNMT3A*^{R882} missense mutations contribute to pre-leukemic and leukemic phases of human AML equally, we aimed to functionally interrogate the disease-phase specific contribution directly in patient specimens.

RESULTS

Correction of *DNMT3A*^{R882H} in pre-leukemic HSCs results in normalization of aberrant self-renewal

Functional studies in murine models^{10,13,14} and descriptive studies in human clonal hematopoiesis¹⁵ suggest that the *DNMT3A*^{R882H} missense mutation induces increased self-renewal in HSCs. To prospectively interrogate whether *DNMT3A*^{R882H} is continuously required for increased self-renewal in patient-derived HSCs, we utilized CRISPR/Cas9 in combination with recombinant AAV6 to perform allele-specific correction of the *DNMT3A*^{R882H} missense mutation in pre-leukemic HSCs isolated from AML patient

DNMT3A^{R882H} Is Not Required for Disease Maintenance in Primary Human AML, but Is Associated With Increased Leukemia Stem Cell Frequency

specimens. We hypothesized that correction of the mutant allele would promptly normalize HSC self-renewal, thus demonstrating that *DNMT3A*^{R882H} directly drives and is continuously required for this phenotype. Building on our previous approaches to *introduce* mutations into primary human hematopoietic stem and progenitor cells (HSPCs),¹⁶⁻²¹ we developed an approach to *revert* existing mutations directly in mutant, patient-derived cells. First, we used our published sorting strategy to isolate rare pre-leukemic HSCs from AML patient samples that contain the *DNMT3A*^{R882H} mutation, but lack the leukemic *NPM1c* and/or *FLT3-ITD* mutations.^{12,22} After isolation by flow cytometry, cells were electroporated with an R882H-specific single guide RNA (sgRNA) pre-complexed with Cas9 protein and transduced with recombinant adeno-associated vector (rAAV6) encoding either the wildtype *DNMT3A*^{R882} codon (correction) or, as a control, the mutant *DNMT3A*^{R882H} codon (re-mutation, **Figure 1A, S1**). Each donor vector also encodes for a fluorescent marker driven by a separate promoter to allow for purification and tracking of successfully engineered cells (**Figure 1B**). Importantly, since we employ an R882H-specific sgRNA, only cells containing the *DNMT3A*^{R882H} mutation are subjected to Cas9-induced double stranded DNA break and repair, and only on the mutant allele. We achieved editing efficiencies of ~5-20% in cells from several primary samples (**Figure 1C**) and single-cell derived colonies from either re-mutated (*DNMT3A*^{WT/R882H}) or corrected (*DNMT3A*^{WT/WT}) HSPCs expressed the appropriate fluorescent marker (**Figure 1D**). We validated on-target integration (**Figure S2A**), in-frame expression of the corrected or re-mutated cDNA (**Figure S2B**), and that our edited pre-leukemic HSCs lacked mutations found in the leukemic fraction (*NPM1c*, *FLT3-ITD*; **Figure S2C**). Taken together, this approach demonstrates that *DNMT3A*^{R882H} can be corrected in pre-leukemic HSCs, and we sought to functionally characterize the corrected cells.

We first performed a serial colony replating assay to determine whether correction of *DNMT3A*^{R882H} alters pre-leukemic HSC self-renewal. During the first plating, gene-edited pre-leukemic cells showed no difference in colony number. However, after serial replating, cells with *DNMT3A*^{R882H} showed a clear and consistent replating advantage compared to gene-corrected *DNMT3A*^{WT/WT} HSPCs (**Figure 1E**). Next, to prospectively elucidate the effect of *DNMT3A*^{R882H} on mature hematopoietic cells, we performed single-cell RNA-sequencing on the progeny of corrected or re-mutated pre-leukemic HSCs from two patients differentiated *in vitro* for 14 days into the erythroid and myeloid lineages. After quality control, 58,036 cells were available for downstream analysis (**Figure 1F**). We assigned cell types based on known marker genes (**Figure 1G**, Methods) and quantified relative abundance of genotypes within each cell type. Here, we noted no major differences in cell abundances between corrected and re-mutated *DNMT3A*^{R882H} (**Figure 1H**), consistent with the lack of large effects on hematopoietic differentiation observed in mouse models^{10,11} and individuals with clonal hematopoiesis harboring this mutation. However, gene set enrichment analysis within individual cell types revealed several differences. First, we observed an enrichment across early myelo-erythroid progenitors for heme metabolism in re-mutated *DNMT3A*^{R882H} cells, in line with previous observations of erythroid lineage bias of *DNMT3A*-mutant HPSCs.^{15,23} Next, we observed a strong enrichment of gene sets linked to inflammation, particularly response to interferon both in myeloid-lineage committed progenitors (granulocyte/monocyte progenitors and promyelocytes) and in mature monocytes and neutrophils (**Figure 1I**), as has been observed in patients with clonal hematopoiesis and zebrafish models.²⁴⁻²⁷ Notably,

DNMT3A^{R882H} Is Not Required for Disease Maintenance in Primary Human AML, but Is Associated With Increased Leukemia Stem Cell Frequency

however, this was not the case in the earliest progenitors detected in this assay (common myeloid progenitors, CMP), which instead showed a marked de-enrichment of the G2M checkpoint signature in re-mutated (*DNMT3A*^{WT/R882H}) cells, in line with observations in murine models¹¹ and our functional observation that stem cell features, such as increased replating capacity, are lost upon correction of *DNMT3A*^{R882H}.

Taken together, these data support the idea that the *DNMT3A*^{R882H} mutation both drives and is continuously required for increased self-renewal in human pre-leukemic HSCs. Importantly, our approach demonstrates that correction of this mutation rapidly eliminates this phenotype, within weeks. Furthermore, this approach prospectively demonstrates the relationship between *DNMT3A*^{R882H} and aberrant HSC self-renewal, increased expression of erythroid programs in HSPCs, as well as a hyperinflammatory state in late progenitors and mature myeloid cells which previously had been noted in descriptive or murine studies.

Correction of *DNMT3A*^{R882H} in primary AML blasts does not eliminate leukemogenicity

Next, we aimed to determine whether *DNMT3A*^{R882H} is still required for the maintenance of established leukemia. We isolated leukemic blasts from six patient specimens with *DNMT3A*^{R882H}, as well as one specimen with *DNMT3A*^{R882C}, and performed either re-mutation or correction as described above (**Figure 2A**). Importantly, unlike pre-leukemic HSCs, these cells carry additional mutations recurrently found in AML, such as in *NPM1*, *FLT3*, *IDH1/2*, and *TET2* (**Figure 2B**). Notably, the correction efficiency varied between patient samples and ranged from 2-38% (**Figure 2C-D**). Next, we performed primary transplantation of FACS-purified, edited leukemic blasts into irradiated NSGS mice and assessed human leukemic engraftment 12-16 weeks later (**Figure 2E**). Strikingly, correction of *DNMT3A*^{R882H} did not eliminate leukemogenicity, with equal levels of human leukemic engraftment regardless of *DNMT3A* genotype (**Figure 2F**). We confirmed that co-mutation patterns were preserved in the engrafted, edited cells, determining that all mutations, except for *DNMT3A*^{R882H}, were present (**Figure 2G, S3A**). Finally, we confirmed that our editing approach correctly replaced the endogenous, mutant allele with either the inserted wildtype *DNMT3A*^{R882} codon (correction) or the inserted mutant *DNMT3A*^{R882H} codon (re-mutation) while leaving the endogenous wildtype-allele unaltered (**Figure S3B-D**).

Phenotypically, correction of *DNMT3A*^{R882H} in leukemic blasts did not alter the morphological appearance (**Figure 2H**) or immunophenotype (**Figure 2I**) assessed in engrafted cells 4 months after gene-correction. To further assess the ability of *DNMT3A*-corrected cells to initiate leukemia, we determined the leukemia-initiating cell (LIC) frequency from two patient samples by performing limiting dilution transplantation assays with re-mutated or corrected leukemic blasts (**Figure 2J**). Similar to engraftment levels, we found no difference in LIC frequency between the re-mutated (*DNMT3A*^{WT/R882H}) and corrected (*DNMT3A*^{WT/WT}) genotypes (**Figure 2K**). Taken together, our approach allows for the selective and specific correction of *DNMT3A*^{R882H} missense mutation in primary human AML blasts and demonstrates that correction of this mutation does not impair leukemic engraftment potential, human chimerism, immunophenotype, or leukemia-initiating cell frequency upon primary transplantation.

In vivo correction of DNMT3A^{R882H} shows no effect on leukemia disease maintenance

In our experiments above, gene-correction of the DNMT3A^{R882H} allele was performed *ex vivo*, and re-mutated or corrected leukemic blasts were transplanted into NSGS mice. Since this approach cannot delineate between the effect of DNMT3A^{R882H} on initial engraftment and subsequent maintenance/proliferation, we established an approach where gene correction of DNMT3A^{R882H} was inducibly performed *in vivo* after the engraftment of human leukemia. For this approach, we utilized an AML-derived induced pluripotent stem cell (iPSC) model, which allows for more extensive genome engineering. We and others have previously demonstrated that iPSCs derived from human AML can be re-differentiated into a transplantable human leukemia that re-acquires both phenotypic and epigenetic features of the original patient sample.^{28,29} To assess the role of DNMT3A^{R882H} in leukemic maintenance, we established an AML-derived iPSC line (iSU444) that carries the DNMT3A^{R882H} mutation along with *FLT3-ITD*, but otherwise has a normal karyotype (**Figure S4A**). These AML-derived iPSCs were then differentiated into HSC-like cells,³⁰ and then xenotransplanted into mice, whereupon they developed into leukemic cells *in vivo* (**Figure S4B**). Using this line, we engineered LoxP-sites flanking the mutant (R882H) exon 23 of DNMT3A and introduced a corrected (WT) exon 23 downstream. Further, we engineered these cells to express doxycycline-inducible Cre recombinase. These cells are mutant (DNMT3A^{WT/R882H}) at baseline, but upon treatment with doxycycline will recombine to correct the endogenous DNMT3A mutant allele to wildtype (DNMT3A^{WT/WT}, **Figure 2L**). We transplanted leukemic blasts derived from iSU444 engineered with this DNMT3A correction cassette into NSGS mice and waited for 3 weeks to allow for initial engraftment. Then, half of the mice were injected with two doses of doxycycline and recombination rates, human chimerism, immunophenotype, and survival were monitored (**Figure 2M**). 5 weeks after injection of doxycycline, we obtained bone marrow aspirates from these mice. Using a droplet digital PCR (ddPCR) strategy to quantify the recombination rate (**Figure S5A**), we found near complete recombination rates of the DNMT3A allele in the human leukemic cells (**Figure 2N**) while observing no difference in human leukemic engraftment (**Figure 2O**) or difference in the percentage of CD34+ human leukemic cells (**Figure S5B**). 14-18 weeks after *in vivo* correction, mice became moribund from fulminant leukemia, showing no difference in survival compared to mice carrying mutant control (DNMT3A^{WT/R882H}) cells (**Figure 2P**). Of note, we confirmed the presence of *FLT3-ITD* (**Figure S5C**) and lack of the mutant DNMT3A^{R882H} allele in most of these mice upon sacrifice (**Figure S5D**). Taken together, *in vivo* correction of DNMT3A^{R882H} showed no difference in human leukemic engraftment, phenotype, or overall survival, further demonstrating that DNMT3A^{R882H} is not required for maintenance of established AML.

Secondary transplantation reveals a reduction of leukemic stem cell frequency upon DNMT3A^{R882H} correction

Leukemic stem cells (LSCs) are a fraction of leukemia cells that exhibit stem cell features, including the ability to self-renew and to generate downstream progeny that are detected through secondary transplantation.³¹ In **Figure 2**, we assessed the ability of leukemia

cells to initiate disease in primary recipient mice and saw no difference between re-mutated (*DNMT3A*^{WT/R882H}) and corrected (*DNMT3A*^{WT/WT}) primary leukemic blasts, including in leukemia-initiating cell (LIC) frequency. To assess whether *DNMT3A*^{R882H} influences the frequency of LSCs, we performed secondary transplantation of leukemic cells from primary recipient mice 4 months after transplantation (**Figure 3A**). Specifically, we performed secondary limiting dilution transplantation assays of re-mutated (*DNMT3A*^{WT/R882H}) and corrected (*DNMT3A*^{WT/WT}) leukemic blasts for two cases. Mice were injected with decreasing doses of either re-mutated or corrected leukemic cells harvested from primary recipient mice and engraftment was assessed 4 months later. Strikingly, the LSC frequency was significantly reduced in cells that had been corrected (*DNMT3A*^{WT/WT}) compared to the re-mutated control (LSC frequency 1/216 in re-mutated versus 1/23,572 in corrected, **Figure 3B**, and 1/10,879 in re-mutated versus undetected in corrected, **Figure 3C**). We confirmed presence of the other clonal mutations in the human leukemia cells engrafted in these mice (**Figure 3D-E**). Taken together, we demonstrate that the frequency of leukemia stem cells is decreased upon correction of the *DNMT3A*^{R882H} mutation in established leukemia, while leukemia engraftment, leukemia-initiating cell frequency, and disease maintenance are not dependent on the persistence of the *DNMT3A*^{R882H} mutation.

Epigenetic and transcriptional identity is highly patient-specific with only minimal changes upon correction of *DNMT3A*^{R882H}

The R882H missense mutation in *DNMT3A* is thought to exert a dominant-negative effect on DNA methyltransferase activity, resulting in CpG hypomethylation.^{15,32} However, recent murine studies have suggested that the DNA methylation landscape can be restored upon overexpression of *DNMT3A* in *DNMT3A*-mutant cells.^{33,34} Inspired by this previous research, we performed whole-genome bisulfite sequencing (WGBS) and transcriptomic analysis (RNA-Seq) in leukemic blasts 4 months after re-mutation or correction of *DNMT3A*^{R882H} (**Figure 4A**). Global genome-wide CpG methylation levels were similar across three patient specimens between *DNMT3A* corrected or *DNMT3A* re-mutated cells (**Figure 4B**). Indeed, when assessing methylation patterns across the genome by principal component analysis (PCA), we found that samples from each patient grouped closely, with no detectable influence of *DNMT3A* genotype in the first two dimensions, although there was partial separation in PC3, which accounts for 4.7% of the variance (**Figure 4C**). Next, we identified differentially methylated regions (DMRs) between *DNMT3A* re-mutated and *DNMT3A* corrected samples within each patient specimen using dmrseq³⁵ (Methods). We detected DMRs in the three patients analyzed based on 2-3 biological replicates in each, with the number of significant DMRs between *DNMT3A* re-mutated and corrected samples in a given patient ranging from 5611 to 13,331 and encompassing 4.0 to 12.4 Mb of the genome. Correction of *DNMT3A*^{R882H} led to increased methylation at 94-99% of these loci (**Table S3-6, Data S1-3**), an effect that was consistent across patient samples and replicates (**Figure 4D**). DMRs were highly patient-specific, with only 89 genes containing DMRs in all three patients; however, this set included genes involved in leukemia, hematopoiesis, and multipotency, such as *FLI1*, *EGFL7*, and *TCF15* (**Table S7**).³⁶⁻⁴⁰ We then analyzed the genomic features associated

with DMRs across our cohort (**Figure 4E, S6A**). DMRs showed enrichment over CpG islands, shores, and shelves (**Figure S6B**).

Since *DNMT3A* mutations have previously been associated with erosion of the borders of large, normally hypomethylated genomic regions (“methylation canyons”),⁴¹ we next interrogated whether we could detect changes at these loci. We first identified human genomic regions analogous to these murine canyons (Methods). Next, we assessed the methylation levels over canyon regions between leukemic patient samples with re-mutated or corrected *DNMT3A*^{R882H}, as well as healthy donor-derived normal CD34+ HSPCs (**Figure 4F, S6C**). Over these regions, correction of *DNMT3A*^{R882H} did not consistently induce re-normalization of methylation levels towards the healthy-donor derived CD34+ HSPC reference. Only 6-21% of canyons displayed differential methylation between re-mutation and correction of *DNMT3A*^{R882H}, indicated by presence of a DMR in the annotated canyon, and only 1-4% of all DMRs lay within canyon regions (**Figure S6D, Table S8, Data S4-7**). These results are in stark contrast to the methylation changes seen in normal HSCs upon knockout of *DNMT3A*, where 75% of canyons display altered DNA methylation, and 30% of DMRs are located within canyons or canyon edges.⁴¹

We previously reported that hypomethylated blocks are a nearly universal feature across many cancer types, including colon, lung, breast, thyroid, and pancreatic cancer, as well as in premalignant lesions and EBV-transformed B-cells, and can range in size from 5 kb to over 2 Mb.⁴²⁻⁴⁴ CpG islands contained in blocks show flattening of methylation profiles, resulting in both the hypermethylation of islands and hypomethylation of shores and shelves characteristic of altered DNA methylation in cancer, suggesting that this disruption of methylation around CpG islands is associated with the loss of structural integrity of heterochromatin in blocks.⁴² We observed a striking relationship between these previously defined blocks and canyons, where canyon regions corresponded with the edges of blocks identified in EBV-transformed B-cells (**Figure S6E, Data S7**). Blocks of hypomethylation have not yet been reported in myeloid tissues, so we selected blocks identified in EBV-transformed B-cells, another hematopoietic cell type, for this comparison. Next, we searched for blocks between leukemic blasts with re-mutated or corrected *DNMT3A*^{R882H}. Putative blocks were identified using minimum sizes of 5 kb, as originally defined, and 3.5 kb, to encompass smaller canyon regions (Methods). Correction of *DNMT3A*^{R882H} did not induce large-scale restoration of methylation, with no putative block regions having significant differences in methylation (**Table S9-14**).

Finally, we determined whether correction of *DNMT3A*^{R882H} affects the transcriptome of AML by performing RNA-Seq on the same specimens utilized for methylome analysis. Similarly to DNA methylation, the global transcriptome primarily clustered by patient identity, rather than *DNMT3A* genotype (**Figure 4G**). To elucidate whether common transcriptional programs were modified upon *DNMT3A* correction, we identified differentially expressed genes and performed gene set enrichment analysis on a per-patient basis (**Figure 4H**). Here, we found that correction of *DNMT3A*^{R882H} consistently reduced the expression of gene sets associated with hematopoietic and leukemic stem cell function, in line with our functional observation of reduced LSC frequency with

DNMT3A^{R882H} Is Not Required for Disease Maintenance in Primary Human AML, but Is Associated With Increased Leukemia Stem Cell Frequency

correction of *DNMT3A*^{R882H}. We then aimed to correlate changes in DNA methylation and gene expression between *DNMT3A*^{R882H} re-mutation and correction. We plotted the fold change in gene expression against the differences in methylation over putative promoter DMRs between re-mutated and corrected *DNMT3A*^{R882H} (**Figure S6F**). While correlation between gene expression and methylation was modest (Pearson's R -0.02 to -0.04 per patient), correction of *DNMT3A*^{R882H} consistently led to increased methylation over genes associated with HSC and LSC function, linking DNA methylation to gene expression (**Figure 4I**).

Our analysis of the methylome and transcriptome in leukemic blasts revealed that correction of *DNMT3A*^{R882H} shows relatively modest influence on epigenetic maintenance in established AML. Both the methylome and transcriptome were highly patient-specific and showed retention of leukemia-specific patterns 4 months after correction of the *DNMT3A*^{R882H} mutation. However, in line with our functional experiments in secondary transplantation studies, we observed an enrichment of hematopoietic and leukemic stem cell programs in re-mutated versus corrected leukemic blasts.

DISCUSSION

During pre-malignant evolution, driver mutations confer a competitive advantage to individual clones within healthy tissue,⁴⁵ which then may acquire additional mutations to progress into frank malignancy. Several of these driver mutations found in pre-malignant stages of cancers have been identified,⁴⁶ such as biallelic inactivation of *APC* in colorectal cancer,⁴⁷ mutations in *TP53* and *CDKN2A* in Barrett's esophagus, melanoma, and pancreatic cancer,^{48,49} and *BRCA1* and *BRCA2* in breast and ovarian cancer.⁵⁰ In pre-leukemic clonal hematopoiesis, the expansion of clones carrying driver mutations has similarly been associated with an increased risk of developing hematologic cancers.⁵¹⁻⁵³ However, our understanding whether mutations acquired early during cancer evolution are required for disease initiation, maintenance of established disease, or both, is limited. While mouse models of driver mutations such as *JAK2*^{V617F}, *IDH2*^{R140Q} or *APC* loss have suggested that these drivers are continuously required for disease maintenance,¹⁻³ it is unclear whether this extends to other pre-malignant drivers, particularly those associated with clonal hematopoiesis.

We focus on the *DNMT3A*^{R882H} missense mutation, which is the most frequent *DNMT3A* missense mutation in human AML,^{4,54} and is also the most frequently affected gene in clonal hematopoiesis.^{6,55,56} The effects of mutant *DNMT3A* have been intensely studied in the context of hematopoiesis, where mutant *DNMT3A* confers increased self-renewal in HSCs^{13,23} and a hyperinflammatory phenotype in mature myeloid cells.²⁵ However, in leukemic cells, the role of *DNMT3A*^{R882H} is less clear. Overall, mutations in *DNMT3A*, including *DNMT3A*^{R882H}, are associated with worse prognosis in AML,^{4,57} however, both co-mutation patterns, as well as clinical attributes such as age, are also associated with *DNMT3A* mutations. Therefore, despite being one of the most frequently mutated genes in human AML, current clinical risk assessment does not take presence or absence of *DNMT3A* mutations into account for treatment decisions⁵⁸ and further investigation to delineate the specific contribution of *DNMT3A* mutations is warranted.

DNMT3A^{R882H} Is Not Required for Disease Maintenance in Primary Human AML, but Is Associated With Increased Leukemia Stem Cell Frequency

Cancer cell lines and model organisms have proven to be useful tools to investigate cancer biology and develop novel treatments. However, prospective, functional experiments with primary human cancers and cancer cells are critical to advance our understanding of human cancer biology. In cancer genetics, interrogation of individual somatic mutations in primary patient specimens is inherently challenging, as it is difficult to control for the combination of co-mutation patterns and germline background. In this study, we approach this challenge by employing tools derived from advances in gene therapy to cancer genetics. We aim to delineate the specific contribution of *DNMT3A*^{R882H}, clarifying whether the presence of this mutation is necessary once leukemia has been established through prospective, controlled genetic experiments directly in primary human leukemia specimens, avoiding surrogate systems for human cancer genetics. Using primary patient samples, we selectively and specifically revert *DNMT3A*^{R882H} back to wildtype, and demonstrate that correction has a pronounced effect in pre-leukemic HSPCs that rapidly lose features of aberrant self-renewal and mature myeloid cells that lose a hyperinflammatory transcriptional profile. However, in established disease, correction of *DNMT3A*^{R882H} does not affect leukemic engraftment potential, human chimerism, immunophenotype, or leukemia-initiating cell frequency. Nor does it affect leukemia disease maintenance when inducibly corrected in engrafted leukemia cells. Notably, correction of *DNMT3A*^{R882H} does result in a decrease in leukemia stem cell frequency, indicating more subtle effects in established disease. Finally, we demonstrate that reversion of the methylation landscape upon correction of *DNMT3A*^{R882H} is minimal, with global methylation levels unchanged 4 months after gene correction. Our results additionally clarify the relationship between DNA methylation canyons identified in normal hematopoietic tissues and blocks of hypomethylation identified in cancer, demonstrating that canyons correspond to the edges of blocks.

This data supports the concept that *DNMT3A*^{R882} mutations are only required during disease initiation, acting as landscaping mutations upon which secondary mutations can act. Importantly, however, once disease transformation has occurred, *DNMT3A*^{R882} mutations become non-essential, where correction does not immediately affect disease maintenance. This stands in stark contrast to other mutations such as *JAK2*^{V617F} or *IDH2*^{R140Q} in leukemia, or *APC* in colorectal cancer, where previous studies have suggested a continued requirement for disease maintenance.¹⁻³

Taken together, our data highlights that prospective, genetic experiments in primary patient samples are feasible, and reveal a surprising stability of leukemic phenotypes and methylation patterns months after reversion of a pre-leukemic driver mutation of human AML. By demonstrating the context-specific contribution of a driver mutation in human cancer, our experiments also have important therapeutic implications. Specifically, therapeutic targeting of *DNMT3A*^{R882} missense mutations might be most effective during pre-leukemic evolution and less effective in established disease. We envision that similar approaches to elucidate the interplay of driver mutations in cancer and improve the discovery of novel therapies.

REFERENCES

1. Kats, L.M., Reschke, M., Taulli, R., Pozdnyakova, O., Burgess, K., Bhargava, P., Straley, K., Karnik, R., Meissner, A., Small, D., et al. (2014). Proto-oncogenic role of mutant IDH2 in leukemia initiation and maintenance. *Cell Stem Cell* *14*, 329-341. 10.1016/j.stem.2013.12.016.
2. Dunbar, A.J., Bowman, R.L., Park, Y.C., O'Connor, K., Izzo, F., Myers, R.M., Karzai, A., Zaroogian, Z., Kim, W.J., Fernandez-Maestre, I., et al. (2024). Jak2V617F Reversible Activation Shows Its Essential Requirement in Myeloproliferative Neoplasms. *Cancer Discov* *14*, 737-751. 10.1158/2159-8290.CD-22-0952.
3. Dow, L.E., O'Rourke, K.P., Simon, J., Tschaharganeh, D.F., van Es, J.H., Clevers, H., and Lowe, S.W. (2015). Apc Restoration Promotes Cellular Differentiation and Reestablishes Crypt Homeostasis in Colorectal Cancer. *Cell* *161*, 1539-1552. 10.1016/j.cell.2015.05.033.
4. Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., Baty, J., Welch, J., et al. (2010). DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* *363*, 2424-2433. 10.1056/NEJMoa1005143.
5. Network, T.C.G.A.R. (2013). Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N. Engl. J. Med.* *368*, 2059-2074. 10.1056/nejmoa1301689 PMID - 23634996.
6. Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burt, N., Chavez, A., et al. (2014). Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N. Engl. J. Med.* *371*, 2488-2498.
7. Venugopal, K., Feng, Y., Shabashvili, D., and Guryanova, O.A. (2021). Alterations to DNMT3A in Hematologic Malignancies. *Cancer Res* *81*, 254-263. 10.1158/0008-5472.CAN-20-3033.
8. Chaudry, S.F., and Chevassut, T.J. (2017). Epigenetic Guardian: A Review of the DNA Methyltransferase DNMT3A in Acute Myeloid Leukaemia and Clonal Haematopoiesis. *Biomed Res Int* *2017*, 5473197. 10.1155/2017/5473197.
9. Kim, S.J., Zhao, H., Hardikar, S., Singh, A.K., Goodell, M.A., and Chen, T. (2013). A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. *Blood* *122*, 4086-4089. 10.1182/blood-2013-02-483487.
10. Higo, T., Suzuki, Y., Sato, M., Koya, J., Mizuno, H., Miyauchi, M., Masamoto, Y., Kataoka, K., Sumitomo, Y., Tsuruta-Kishino, T., et al. (2022). Heterozygous Dnmt3a R878C induces expansion of quiescent hematopoietic stem cell pool. *Exp Hematol* *109*, 45-54. 10.1016/j.exphem.2022.02.006.
11. Guryanova, O.A., Shank, K., Spitzer, B., Luciani, L., Koche, R.P., Garrett-Bakelman, F.E., Ganzel, C., Durham, B.H., Mohanty, A., Hoermann, G., et al. (2016). DNMT3A mutations promote anthracycline resistance in acute myeloid leukemia via impaired nucleosome remodeling. *Nat. Med.* *22*, 1488-1495. 10.1038/nm.4210 PMID - 27841873.
12. Corces-Zimmerman, M.R., Hong, W.-J., Weissman, I.L., Medeiros, B.C., and Majeti, R. (2014). Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proceedings of the National Academy of Sciences* *111*, 2548-2553. 10.1073/pnas.1324297111 PMID - 24550281.

DNMT3A^{R882H} Is Not Required for Disease Maintenance in Primary Human AML, but Is Associated With Increased Leukemia Stem Cell Frequency

13. Jeong, M., Park, H.J., Celik, H., Ostrander, E.L., Reyes, J.M., Guzman, A., Rodriguez, B., Lei, Y., Lee, Y., Ding, L., et al. (2018). Loss of Dnmt3a Immortalizes Hematopoietic Stem Cells In Vivo. *CellReports* 23, 1-10.
14. Celik, H., Mallaney, C., Kothari, A., Ostrander, E.L., Eultgen, E., Martens, A., Miller, C.A., Hundal, J., Klco, J.M., and Challen, G.A. (2015). Enforced differentiation of Dnmt3a-null bone marrow leads to failure with c-Kit mutations driving leukemic transformation. *Blood* 125, 619-628. 10.1182/blood-2014-08-594564 PMID - 25416276.
15. Nam, A.S., Dusaj, N., Izzo, F., Murali, R., Myers, R.M., Mouhieddine, T.H., Sotelo, J., Benbarche, S., Waarts, M., Gaiti, F., et al. (2022). Single-cell multi-omics of human clonal hematopoiesis reveals that DNMT3A R882 mutations perturb early progenitor states through selective hypomethylation. *Nat Genet* 54, 1514-1526. 10.1038/s41588-022-01179-9.
16. Bak, R.O., Dever, D.P., Reinisch, A., Hernandez, D.C., Majeti, R., and Porteus, M.H. (2017). Multiplexed genetic engineering of human hematopoietic stem and progenitor cells using CRISPR/Cas9 and AAV6. *Elife* 6, 18.
17. Sconocchia, T., Foßelteder, J., Köhnke, T., Majeti, R., and Reinisch, A. (2023). Engineering oncogenic heterozygous gain-of-function mutations in human hematopoietic stem and progenitor cells. *J. Vis. Exp.* 10.3791/64558.
18. Köhnke, T., Nuno, K.A., Alder, C.C., Gars, E.J., Phan, P., Fan, A.C., and Majeti, R. (2024). Human ASXL1-Mutant Hematopoiesis Is Driven by a Truncated Protein Associated with Aberrant Deubiquitination of H2AK119. *Blood Cancer Discov* 5, 202-223. 10.1158/2643-3230.BCD-23-0235.
19. Fan, A.C., Nakauchi, Y., Bai, L., Azizi, A., Nuno, K.A., Zhao, F., Köhnke, T., Karigane, D., Cruz-Hernandez, D., Reinisch, A., et al. (2023). RUNX1 loss renders hematopoietic and leukemic cells dependent on IL-3 and sensitive to JAK inhibition. *J Clin Invest* 133. 10.1172/JCI167053.
20. Nakauchi, Y., Azizi, A., Thomas, D., Corces, M.R., Reinisch, A., Sharma, R., Cruz Hernandez, D., Köhnke, T., Karigane, D., Fan, A., et al. (2022). The Cell Type-Specific 5hmC Landscape and Dynamics of Healthy Human Hematopoiesis and TET2-Mutant Preleukemia. *Blood Cancer Discov* 3, 346-367. 10.1158/2643-3230.BCD-21-0143.
21. Suchy, F.P., Karigane, D., Nakauchi, Y., Higuchi, M., Zhang, J., Pekrun, K., Hsu, I., Fan, A.C., Nishimura, T., Charlesworth, C.T., et al. (2024). Genome engineering with Cas9 and AAV repair templates generates frequent concatemeric insertions of viral vectors. *Nat Biotechnol.* 10.1038/s41587-024-02171-w.
22. Jan, M., Snyder, T.M., Corces-Zimmerman, M.R., Vyas, P., Weissman, I.L., Quake, S.R., and Majeti, R. (2012). Clonal Evolution of Preleukemic Hematopoietic Stem Cells Precedes Human Acute Myeloid Leukemia. *Sci. Transl. Med.* 4, 149ra118-149ra118. 10.1126/scitranslmed.3004315 PMID - 22932223.
23. Izzo, F., Lee, S.C., Poran, A., Chaligne, R., Gaiti, F., Gross, B., Murali, R.R., Deochand, S.D., Ang, C., Jones, P.W., et al. (2020). DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nat Genet* 52, 378-387. 10.1038/s41588-020-0595-4.
24. Heimlich, J.B., Bhat, P., Parker, A.C., Jenkins, M.T., Vlasschaert, C., Ulloa, J., Van Amburg, J.C., Potts, C.R., Olson, S., Silver, A.J., et al. (2024). Multiomic

profiling of human clonal hematopoiesis reveals genotype and cell-specific inflammatory pathway activation. *Blood Adv* 8, 3665-3678. 10.1182/bloodadvances.2023011445.

25. Abplanalp, W.T., Cremer, S., John, D., Hoffmann, J., Schuhmacher, B., Merten, M., Rieger, M.A., Vasa-Nicotera, M., Zeiher, A.M., and Dimmeler, S. (2021). Clonal Hematopoiesis-Driver DNMT3A Mutations Alter Immune Cells in Heart Failure. *Circ Res* 128, 216-228. 10.1161/CIRCRESAHA.120.317104.
26. Avagyan, S., Henninger, J.E., Mannherz, W.P., Mistry, M., Yoon, J., Yang, S., Weber, M.C., Moore, J.L., and Zon, L.I. (2021). Resistance to inflammation underlies enhanced fitness in clonal hematopoiesis. *Science* 374, 768-772. 10.1126/science.aba9304.
27. Jakobsen, N.A., Turkalj, S., Zeng, A.G.X., Stoilova, B., Metzner, M., Rahmig, S., Nagree, M.S., Shah, S., Moore, R., Usukhbayar, B., et al. (2024). Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging. *Cell Stem Cell* 31, 1127-1144 e1117. 10.1016/j.stem.2024.05.010.
28. Chao, M.P., Gentles, A.J., Chatterjee, S., Lan, F., Reinisch, A., Corces, M.R., Xavy, S., Shen, J., Haag, D., Chanda, S., et al. (2017). Human AML-iPSCs Reacquire Leukemic Properties after Differentiation and Model Clonal Variation of Disease. *Stem Cells* 20, 329-344.e327.
29. Kotini, A.G., Carcamo, S., Cruz-Rodriguez, N., Olszewska, M., Wang, T., Demircioglu, D., Chang, C.J., Bernard, E., Chao, M.P., Majeti, R., et al. (2023). Patient-Derived iPSCs Faithfully Represent the Genetic Diversity and Cellular Architecture of Human Acute Myeloid Leukemia. *Blood Cancer Discov* 4, 318-335. 10.1158/2643-3230.BCD-22-0167.
30. Fowler, J.L., Zheng, S.L., Nguyen, A., Chen, A., Xiong, X., Chai, T., Chen, J.Y., Karigane, D., Banuelos, A.M., Niizuma, K., et al. (2024). Lineage-tracing hematopoietic stem cell origins in vivo to efficiently make human HLF+ HOXA+ hematopoietic progenitors from pluripotent stem cells. *Dev Cell* 59, 1110-1131 e1122. 10.1016/j.devcel.2024.03.003.
31. Thomas, D., and Majeti, R. (2017). Biology and relevance of human acute myeloid leukemia stem cells. *Blood* 129, 1577-1585.
32. Anteneh, H., Fang, J., and Song, J. (2020). Structural basis for impairment of DNA methylation by the DNMT3A R882H mutation. *Nat Commun* 11, 2294. 10.1038/s41467-020-16213-9.
33. Li, Y., Abel, H.J., Cai, M., LaValle, T.A., Yin, T., Helton, N.M., Smith, A.M., Miller, C.A., and Ley, T.J. (2024). Rapid and accurate remethylation of DNA in Dnmt3a-deficient hematopoietic cells with restoration of DNMT3A activity. *Sci Adv* 10, eadk8598. 10.1126/sciadv.adk8598.
34. Smith, A.M., Verdoni, A.M., Abel, H.J., Chen, D.Y., Ketkar, S., Leight, E.R., Miller, C.A., and Ley, T.J. (2022). Somatic Dnmt3a inactivation leads to slow, canonical DNA methylation loss in murine hematopoietic cells. *iScience* 25, 104004. 10.1016/j.isci.2022.104004.
35. Korthauer, K., Chakraborty, S., Benjamini, Y., and Irizarry, R.A. (2019). Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. *Biostatistics* 20, 367-383. 10.1093/biostatistics/kxy007.

36. Torchia, E.C., Boyd, K., Rehg, J.E., Qu, C., and Baker, S.J. (2007). EWS/FLI-1 induces rapid onset of myeloid/erythroid leukemia in mice. *Mol Cell Biol* 27, 7918-7934. 10.1128/MCB.00099-07.
37. Bill, M., Pathmanathan, A., Karunasiri, M., Shen, C., Burke, M.H., Ranganathan, P., Papaioannou, D., Zitzer, N.C., Snyder, K., LaRocco, A., et al. (2020). EGFL7 Antagonizes NOTCH Signaling and Represents a Novel Therapeutic Target in Acute Myeloid Leukemia. *Clin Cancer Res* 26, 669-678. 10.1158/1078-0432.CCR-19-2479.
38. Rodriguez-Fraticelli, A.E., Weinreb, C., Wang, S.W., Migueles, R.P., Jankovic, M., Usart, M., Klein, A.M., Lowell, S., and Camargo, F.D. (2020). Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. *Nature* 583, 585-589. 10.1038/s41586-020-2503-6.
39. Davies, O.R., Lin, C.Y., Radzisheuskaya, A., Zhou, X., Taube, J., Blin, G., Waterhouse, A., Smith, A.J., and Lowell, S. (2013). Tcf15 primes pluripotent cells for differentiation. *Cell Rep* 3, 472-484. 10.1016/j.celrep.2013.01.017.
40. Koldobskiy, M.A., Abante, J., Jenkinson, G., Pujadas, E., Tetens, A., Zhao, F., Tryggvadottir, R., Idrizi, A., Reinisch, A., Majeti, R., et al. (2020). A Dysregulated DNA Methylation Landscape Linked to Gene Expression in MLL-Rearranged AML. *Epigenetics* 15, 841-858. 10.1080/15592294.2020.1734149.
41. Jeong, M., Sun, D., Luo, M., Huang, Y., Challen, G.A., Rodriguez, B., Zhang, X., Chavez, L., Wang, H., Hannah, R., et al. (2014). Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat Genet* 46, 17-23. 10.1038/ng.2836.
42. Timp, W., Bravo, H.C., McDonald, O.G., Goggins, M., Umbricht, C., Zeiger, M., Feinberg, A.P., and Irizarry, R.A. (2014). Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Med* 6, 61. 10.1186/s13073-014-0061-y.
43. Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D., et al. (2011). Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 43, 768-775. 10.1038/ng.865.
44. Hansen, K.D., Sabunciyan, S., Langmead, B., Nagy, N., Curley, R., Klein, G., Klein, E., Salamon, D., and Feinberg, A.P. (2014). Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization. *Genome Res* 24, 177-184. 10.1101/gr.157743.113.
45. Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. *Science* 339, 1546-1558. 10.1126/science.1235122.
46. Curtius, K., Wright, N.A., and Graham, T.A. (2017). Evolution of Premalignant Disease. *Cold Spring Harb Perspect Med* 7. 10.1101/cshperspect.a026542.
47. Shibata, H., Toyama, K., Shioya, H., Ito, M., Hirota, M., Hasegawa, S., Matsumoto, H., Takano, H., Akiyama, T., Toyoshima, K., et al. (1997). Rapid colorectal adenoma formation initiated by conditional targeting of the Apc gene. *Science* 278, 120-123. 10.1126/science.278.5335.120.
48. Barrett, M.T., Sanchez, C.A., Prevo, L.J., Wong, D.J., Galipeau, P.C., Paulson, T.G., Rabinovitch, P.S., and Reid, B.J. (1999). Evolution of neoplastic cell lineages in Barrett oesophagus. *Nat Genet* 22, 106-109. 10.1038/8816.

49. Kimura, H., Klein, A.P., Hruban, R.H., and Roberts, N.J. (2021). The Role of Inherited Pathogenic CDKN2A Variants in Susceptibility to Pancreatic Cancer. *Pancreas* *50*, 1123-1130. 10.1097/MPA.0000000000001888.
50. Welcsh, P.L., and King, M.C. (2001). BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum Mol Genet* *10*, 705-713. 10.1093/hmg/10.7.705.
51. Desai, P., Mencia-Trinchant, N., Savenkov, O., Simon, M.S., Cheang, G., Lee, S., Samuel, M., Ritchie, E.K., Guzman, M.L., Ballman, K.V., et al. (2018). Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nat. Med.* *24*, 1015-1023. 10.1038/s41591-018-0081-z PMID - 29988143.
52. Abelson, S., Collord, G., Ng, S.W.K., Weissbrod, O., Cohen, N.M., Niemeyer, E., Barda, N., Zuzarte, P.C., Heisler, L., Sundaravadanam, Y., et al. (2018). Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature* *559*, 400-404. 10.1038/s41586-018-0317-6 PMID - 29988082.
53. Fabre, M.A., de Almeida, J.G., Fiorillo, E., Mitchell, E., Damaskou, A., Rak, J., Orrù, V., Marongiu, M., Chapman, M.S., Vijayabaskar, M.S., et al. (2022). The longitudinal dynamics and natural history of clonal haematopoiesis. *Nature* *606*, 335-342. 10.1038/s41586-022-04785-z.
54. Brunetti, L., Gundry, M.C., and Goodell, M.A. (2017). DNMT3A in Leukemia. *Cold Spring Harb Perspect Med* *7*. 10.1101/cshperspect.a030320.
55. Xie, M., Lu, C., Wang, J., McLellan, M.D., Johnson, K.J., Wendl, M.C., McMichael, J.F., Schmidt, H.K., Yellapantula, V., Miller, C.A., et al. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* *20*, 1472-1478.
56. Genovese, G., Kähler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M., et al. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* *371*, 2477-2487.
57. Ribeiro, A.F., Pratorcorona, M., Erpelinck-Verschueren, C., Rockova, V., Sanders, M., Abbas, S., Figueroa, M.E., Zeilemaker, A., Melnick, A., Lowenberg, B., et al. (2012). Mutant DNMT3A: a marker of poor prognosis in acute myeloid leukemia. *Blood* *119*, 5824-5831. 10.1182/blood-2011-07-367961.
58. Dohner, H., Wei, A.H., Appelbaum, F.R., Craddock, C., DiNardo, C.D., Dombret, H., Ebert, B.L., Fenaux, P., Godley, L.A., Hasserjian, R.P., et al. (2022). Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood* *140*, 1345-1377. 10.1182/blood.2022016867.
59. Nishimura, K., Ohtaka, M., Takada, H., Kurisaki, A., Tran, N.V.K., Tran, Y.T.H., Hisatake, K., Sano, M., and Nakanishi, M. (2017). Simple and effective generation of transgene-free induced pluripotent stem cells using an auto-erasable Sendai virus vector responding to microRNA-302. *Stem Cell Res* *23*, 13-19. 10.1016/j.scr.2017.06.011.
60. Miyazaki, T., Isobe, T., Nakatsuji, N., and Suemori, H. (2017). Efficient Adhesion Culture of Human Pluripotent Stem Cells Using Laminin Fragments in an Uncoated Manner. *Sci Rep* *7*, 41165. 10.1038/srep41165.
61. Luca, V.C., Jude, K.M., Pierce, N.W., Nachury, M.V., Fischer, S., and Garcia, K.C. (2015). Structural biology. Structural basis for Notch1 engagement of Delta-like 4. *Science* *347*, 847-853. 10.1126/science.1261093.

62. Gregorevic, P., Blankinship, M.J., Allen, J.M., Crawford, R.W., Meuse, L., Miller, D.G., Russell, D.W., and Chamberlain, J.S. (2004). Systemic delivery of genes to striated muscles using adeno-associated viral vectors. *Nat Med* *10*, 828-834. 10.1038/nm1085.
- 5 63. Aurnhammer, C., Haase, M., Muether, N., Hausl, M., Rauschhuber, C., Huber, I., Nitschko, H., Busch, U., Sing, A., Ehrhardt, A., and Baiker, A. (2012). Universal Real-Time PCR for the Detection and Quantification of Adeno-Associated Virus Serotype 2-Derived Inverted Terminal Repeat Sequences. *Human Gene Therapy, Part B: Methods* *23*, 18-28. 10.1089/hgtb.2011.034 PMID - 22428977.
- 10 64. Lehnertz, B., Chagraoui, J., MacRae, T., Tomellini, E., Corneau, S., Mayotte, N., Boivin, I., Durand, A., Gracias, D., and Sauvageau, G. (2021). HLF expression defines the human hematopoietic stem cell state. *Blood* *138*, 2642-2654. 10.1182/blood.2021010745.
- 15 65. Selvaraj, S., Feist, W.N., Viel, S., Vaidyanathan, S., Dudek, A.M., Gastou, M., Rockwood, S.J., Ekman, F.K., Oseghale, A.R., Xu, L., et al. (2024). High-efficiency transgene integration by homology-directed repair in human primary cells using DNA-PKcs inhibition. *Nat Biotechnol* *42*, 731-744. 10.1038/s41587-023-01888-4.
- 20 66. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* *27*, 1571-1572. 10.1093/bioinformatics/btr167.
67. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* *25*, 1754-1760. 10.1093/bioinformatics/btp324 PMID - 19451168.
- 25 68. Hansen, K.D., Langmead, B., and Irizarry, R.A. (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* *13*, R83. 10.1186/gb-2012-13-10-r83.
69. Cavalcante, R.G., and Sartor, M.A. (2017). annotatr: genomic regions in context. *Bioinformatics* *33*, 2381-2383. 10.1093/bioinformatics/btx183.
- 30 70. Akalin, A., Franke, V., Vlahovicek, K., Mason, C.E., and Schubeler, D. (2015). Genomation: a toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics* *31*, 1127-1129. 10.1093/bioinformatics/btu775.
71. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* *32*, 2847-2849. 10.1093/bioinformatics/btw313.
- 35 72. Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* *9*, 215-216. 10.1038/nmeth.1906.
73. Wu, H., Caffo, B., Jaffee, H.A., Irizarry, R.A., and Feinberg, A.P. (2010). Redefining CpG islands using hidden Markov models. *Biostatistics* *11*, 499-514. 10.1093/biostatistics/kxq005.
- 40 74. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* *102*, 15545-15550. 10.1073/pnas.0506580102.
- 45 75. Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., et al. (2003). PGC-

- 1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273. 10.1038/ng1180.
- 5 76. Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., et al. (2006). The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* **34**, D590-598. 10.1093/nar/gkj144.
- 10 77. Melsted, P., Boeshaghi, A.S., Liu, L., Gao, F., Lu, L., Min, K.H., Beltrame, E.d.V., Hjörleifsson, K.E., Gehring, J., and Pachter, L. (2021). Modular, efficient and constant-memory single-cell RNA-seq preprocessing. *Nat. Biotechnol.* **39**, 813-818. 10.1038/s41587-021-00870-2 PMID - 33795888.
78. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525-527. 10.1038/nbt.3519 PMID - 27043002.
- 15 79. Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **4**, 1521. 10.12688/f1000research.7563.1 PMID - 26925227.
80. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. 10.1186/s13059-014-0550-8 PMID - 25516281.
- 20 81. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., and Satija, R. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* **42**, 293-304. 10.1038/s41587-023-01767-y.
- 25 82. McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337 e324. 10.1016/j.cels.2019.03.003.
83. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296. 10.1038/s41592-019-0619-0.
- 30 84. Busarello, E., Biancon, G., Lauria, F., Ibnat, Z., Ramirez, C., Tome, G., Aass, K.R., VanOudenhove, J., Standal, T., Viero, G., et al. (2024). Interpreting single-cell messages in normal and aberrant hematopoiesis with the Cell Marker Accordion. *bioRxiv*. 10.1101/2024.03.08.584053.
- 35 85. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2021). Fast gene set enrichment analysis. *bioRxiv*, 060012. 10.1101/060012.

METHODS

Primary Human Samples

AML patient samples were obtained from patients with AML, with informed consent according to institutional guidelines (Stanford University Institutional Review Board No. 6453). Mononuclear cells were isolated by Ficoll-Paque PLUS (GE Healthcare) density gradient centrifugation, red blood cells were removed by ACK lysis (ACK Lysing Buffer, Thermo Fisher Scientific). Samples were resuspended in 90% FBS + 10% DMSO (Sigma-Aldrich) and cryopreserved in liquid nitrogen for future use.

Purification of cell populations from primary human AML samples

Purification of residual HSCs from AML specimens was performed as previously described.¹² Briefly, AML specimens were thawed, washed, stained and sorted on a FACSAria II SORP (BD). The residual HSC fraction was defined as CD3-/CD19-/CD34+/CD38-/CD99-/TIM3- and AML blasts were sorted as CD3-/CD19-/CD14-. FACS staining panels can be found in Supplementary Table 1.

iPSC culture and derivation

The iSU444 iPSC line was established from purified AML patient blasts as described previously,²⁸ with some modifications. AML blasts were cultured at 37°C, 5% CO₂ and 5% O₂ in AML pre-culture media comprised of MyeloCult H5100 (STEMCELL Technologies) base media supplemented with 20 ng/mL each of TPO, SCF, FLT3L, G-CSF, GM-CSF, IL-3 and IL-6 (Peprotech) for three days. Cell viability was confirmed to be >70% at the time of transduction with reprogramming factors. AML blast transgene-free reprogramming was performed using Sendai virus containing SeVdp(KOSM)302L as previously described,⁵⁹ with some modifications. Cells were cultured in AML pre-culture media at a density of 0.5×10⁶ cells/mL (0.5×10⁵ cells per 100 μl) in a 96 well plate. On day 0, SeVdp(KOSM)302L Sendai virus was added at an MOI of 10 with cells kept at 37°C at 5% O₂ for 12 hr. After infection, the cell suspensions were collected, diluted with 900 μl of pre-culture medium, and centrifuged at 300 xg for 5 min at room temperature. Cells were then resuspended in a mixture of 375 μl AML pre-culture media and 375 μl of StemFit Basic04 Complete Type medium iPSCs media (Ajinomoto Co.) in a 12-well plate coated with iMatrix-511 basement membrane matrix (ReproCell). On day 3, 375 μl AML pre-culture media and 375 μl iPSCs media were gently added. On day 5, 1500 μl iPSCs media was gently added. From day 7, half media changes with 1 ml of fresh iPSCs media were performed until formation of colonies with iPSC morphology could be observed. iPSC-like colonies were manually picked and expanded individually for further analyses. iPSCs and their derivative subclones were propagated feeder-free in StemFit Basic04 Complete Type medium (Ajinomoto Co.) on cell culture plastics with iMatrix-511 basement membrane matrix (ReproCell).⁶⁰ HSPC Differentiation from iPSCs was performed using a non-transgene 10-day-culture differentiation method as previously described³⁰, with slight modifications. Specifically, on day 4, artery endothelial cells were dissociated into a single-cell suspension (Accutase) and densely re-seeded at 500,000 cells/cm² onto plates precoated with 10 μg/mL Vitronectin and 20 nM of high-affinity NOTCH agonist DLL4-E12 (Vincent Luca's laboratory, Moffitt Cancer Center and Chris Garcia's laboratory, Stanford).⁶¹ Cells were then further differentiated towards hemogenic endothelium in CDM3 media supplemented with Forskolin (10 μM), LIF (20 ng/mL [R&D Systems, 7734-LF-025]), OSM (10 ng/mL [R&D Systems, 295-OM-010]) and SB505124

(2 μ M) for 72 hours. Hemogenic endothelium induction media was refreshed every 24 hours with a complete media change. Day 7 hemogenic endothelium cells were differentiated towards HLF+ HOXA+ hematopoietic progenitors in CDM3 media supplemented with Forskolin (10 μ M), SB505124 (2 μ M), OSM (10 ng/mL [R&D Systems, 295-OM-010]), LIF (20 ng/mL [R&D Systems, 7734-LF-025]), SB505124 (2 μ M), SR1 (750 nM [Cellagen, C7710-5]), UM171 (75 nM [ApexBio, A8950]) and IL-1 β (5 ng/mL [PeproTech, 200-01B]) for 72 hours. Media changes were performed gently to avoid disturbing the emerging semi-adherent cells. Hematopoietic progenitor induction media was refreshed every 24 hours with a complete media change on days 8 and 9; on day 10, fresh media was added without media change. On day 11, CD45+ cells were isolated by FACS (CD45 Mouse anti-Human, APC, Clone: 2D1 [BD, 340943]).

AAV vector production

AAV vector plasmids were cloned in the pAAV-MCS plasmid containing ITRs from AAV serotype 2 (AAV2).¹⁶ Between the ITRs, we inserted two homology arms of 400 bp flanking the remaining, codon-optimized corrected (R882R) or re-mutated (R882H) coding sequence of the final DNMT3A exon, as well as a cassette containing the UBC promoter and either turboGFP (for DNMT3A correction) or mTagBFP2 (for DNMT3A re-mutation). For experiments with iPSC-AML, we increased the length of two homology arms to 600-729 bp and inserted a loxP site 81bp upstream of the final, mutant DNMT3A exon, followed by a UBC-Hygromycin selection cassette, the second loxP site, followed by the final, corrected DNMT3A exon. Finally, all vectors were engineered to contain distal cut sites for ITR removal contained the same sequences as the genomic sgRNA target sequences, with the PAM sequence facing inward between the ITR and homology arms.²¹ AAV6 particles were produced in 293FT cells (Thermo Fisher Scientific) transfected using standard PEI transfection with ITR-containing plasmids and pDGM6 containing the AAV6 cap genes, AAV2 rep genes and adenovirus serotype 5 helper genes (gift from David Russell - Addgene plasmid #110660).⁶² Particles were harvested after 72 h, purified using the AAVpro Purification Kit (Takara Bio) according to the manufacturer's instructions and then stored at -80 °C until further use. Viral vector titers measured as vector genomes per microliter (vg/ μ l) were determined by ddPCR analysis of serially diluted vector samples using primers and probes that target the vector ITR as previously described.⁶³

CRISPR/Cas9 Editing

CRISPR/Cas9 nucleofection and rAAV6-mediated HDR were performed as previously described,¹⁶ with some modifications. For experiments with residual HSPCs, cells were cultured at 37°C, 5% CO₂ and 5% O₂ in in growth media comprised of StemSpan SFEM II (STEMCELL Technologies) base media supplemented with 20 ng/mL of TPO, SCF, FLT3L, and IL-6 (PeproTech) and 35nM of UM-171 (Selleckchem) for 48h. Synthetic, chemically modified sgRNAs targeting DNMT3A:R882H (5'-CGTCTCCAACATGAGCCACT-3') or DNMT3A:R882C (5'-CGTCTCCAACATGAGCTGCT-3') and Cas9 protein (Alt-R HiFi CRISPR-Cas9) were purchased from IDT. The appropriate sgRNA was precomplexed with Cas9 protein at a molar ratio of 1:2.5 at 25°C for 10 min immediately prior to electroporation. Cells were electroporated using the Lonza Nucleofector 4D (program DZ-100) using 150 μ g/mL Cas9 protein, with the addition of 20 pmol siTP53⁶⁴ in P3 Primary Cell Nucleofector Solution (Lonza). Immediately after electroporation, cells were plated with rAAV6 particles at a multiplicity of infection of 5,000 for 8h in growth media as above, supplemented with 0.5

5 μ M AZD7648.⁶⁵ Cells were washed and cultured in growth media for a total of 3 days after electroporation before isolation of fluorescent protein positive populations by FACS. For experiments with purified AML blasts, cells were cultured at 37°C, 5% CO₂ and 21% O₂ in AML media comprised of StemSpan SFEM II (STEMCELL Technologies) base media supplemented with 20 ng/mL of TPO, SCF, FLT3L, GM-CSF and IL-6 (Peprotech) and editing was performed as described above. For experiments with iPSC, editing was performed immediately after dissociation using program CB-150 (Lonza Nucleofector 4D), but otherwise as described above.

10 Genotyping

15 Droplet Digital PCR (ddPCR) assays were performed using custom TaqMan SNP Genotyping Assays (Thermo Fisher) and ddPCR Supermix for Probes (Bio-Rad). Droplet generation and analysis were performed on the Bio-Rad QX200 Droplet Digital PCR System (Bio-Rad). FLT3-ITD and In-Out-PCR were performed using target-specific primers and SeqAmp DNA Polymerase (Takara). For RT-PCR, cDNA was generated using SuperScript VILO cDNA Synthesis Kit (Thermo Fisher), and PCR was performed using cDNA-specific primers and SeqAmp DNA Polymerase (Takara). Primer sequences can be found in Supplementary Table 2.

20 Colony Formation Assays

25 For colony formation assays, 200-500, edited residual HSCs were added to 4 mL of MethoCult H4435 (Stemcell Technologies) and plated in triplicate. The number of colonies in each sample was scored at 12-16 days. For replating assays, cells were washed and replated 100,000 cells/plate.

30 Animal Studies

All mouse experiments were conducted in accordance with a protocol by the Institutional Animal Care and Use Committee (Stanford Administrative Panel on Laboratory Animal Care #22264) and in adherence with the U.S. National Institutes of Health's Guide for the Care and Use of Laboratory Animals.

35 NOD.Cg-Prkdcscid Il2rgtm1WjITg(CMV-IL3,CSF2,KITLG)1Eav/MloySzJ (NSGS) mice were purchased from Jackson laboratory and bred in-house. Mice were housed in specific-pathogen-free animal facilities in microisolator cages. Human cells were engrafted into 6 to 8 week old mice 2 to 24h after sublethal irradiation (200 rad) by intravenous (iPSC-AML studies) or intrafemoral (primary AML studies) injection of 10-50,000 cells. Bone marrow aspirates were obtained at the indicated time points from the right femur. For terminal engraftment analysis, mice were humanely euthanized, and femurs, tibias, hip bones, sternum, and spine were harvested and crushed. Mononuclear cells were isolated by Ficoll-Paque PLUS (GE Healthcare) density gradient centrifugation. 40 For both bone marrow aspirates and terminal engraftment analysis, red blood cells were removed with ACK lysis (ACK Lysing Buffer, ThermoFisher Scientific), and stained for flow cytometry. For secondary transplantation, cells were sorted for human CD45 and transplanted into recipients as indicated.

45 Whole-genome bisulfite sequencing

Cells were sorted from crushed whole bone marrow for Whole Genome Bisulfite Sequencing (WGBS) at the indicate timepoints. Genomic DNA was extracted from >100,000 cells using DNA/RNA microprep plus kit (Zymo) following manufacturer's

protocol. The generation of dual-indexed libraries for WGBS followed a modified protocol based on the NEBNext Ultra DNA Library Prep Kit for Illumina (New England BioLabs), in accordance with the manufacturer's instructions. Library input was 200 ng genomic DNA, quantified by the Qubit BR assay (Invitrogen), and 1% unmethylated Lambda DNA (Promega) was added as a control for bisulfite conversion efficiency. Fragmentation of input gDNA to achieve an average insert size of 350 bp was done with a Covaris S220 Focused-ultrasonicator. Size selection was performed using SPRIselect beads, isolating fragments within the 300-400 bp range. Bisulfite conversion post-size selection was done with the EZ DNA Methylation-Lightning Kit (Zymo), following the manufacturer's guidelines. For amplification post-bisulfite conversion, Kapa HiFi Uracil+ (Kapa Biosystems) polymerase was used under the following cycling conditions: 98°C for 45 s, followed by 9 cycles of 98°C for 15 s, 65°C for 30 s, and 72°C for 30 s, with a final extension step at 72°C for 1 min. WGBS libraries were sequenced on an Illumina NovaSeq 6000 instrument using an S4 flowcell and a 150 bp paired-end dual indexed run, with 5% PhiX library control (Illumina). Autosomal ranged from 23X to 29X. The bisulfite conversion rate for unmethylated Lambda DNA averaged 99.8%.

Sequencing reads were trimmed and quality control was performed using TrimGalore v.0.6.6 (<https://doi.org/10.5281/zenodo.5127899>) with default parameters. Alignment of reads to the hg19 genome was performed using Bismark v.0.23.0⁶⁶ with default parameters. Resulting BAM files were processed with samtools v.1.18⁶⁷ for merging, sorting, deduplication, and indexing. Finally, Bismark's methylation_extractor was used to extract methylation calls and generate CpG_report files. Bismark CpG_report files were imported into bsseq v.1.36.0⁶⁸ for downstream analysis.

Methylation analysis

Global methylation levels were assessed using raw CpG methylation over chromosomes 1-22. Principal component analysis (PCA) was performed on these raw methylation values using the prcomp function from the R stats package. Feature-level CpG methylation was calculated over genomic annotations generated by the Bioconductor package annotatr v.1.22.0.⁶⁹ Differentially methylated regions (DMRs) were identified between re-mutated and corrected samples within each patient. CpGs were filtered to include only autosomal CpG sites with at least 1x coverage in at least one sample per group. DMRs were called using Bioconductor package dmrseq v.1.16.0³⁵ with cutoff set to 0.05, and significant DMRs were determined with a Benjamini-Hochberg q-value ($q \leq 0.2$). Block DMRs were identified using the dmrseq() function with minInSpan=500, bpSpan=2000, and block=TRUE; blockSize was set to 5000 for identification of blocks as originally defined (5 kb and up), or set to 3500 to include detection of canyon regions (3.5 kb and up). Meta-region plots over DMRs were generated using the Bioconductor package genomation v.1.28.0,⁷⁰ and heatmaps were generated using the ComplexHeatmap package v.2.15.4⁷¹ with default clustering.

For genomic annotation of DMR regions, promoters were defined as ± 2 kb from transcriptional start sites generated by the Bioconductor package TxDb.Hsapiens.UCSC.hg19.knownGene. DMRs were assigned to genes based on their presence within annotated promoter regions. Annotations for K562 chromatin states from chromHMM⁷² and genic features were generated using annotatr's built-in hg19_K562-chromatin and hg19_basicgenes. Annotations for hg19 CpG islands were obtained from Wu et al.⁷³ CpG island shores were defined as 2 kb regions flanking islands, CpG island shelves were defined as 2 kb regions flanking shores, and open seas were defined as

everything else. Enrichment analysis of DMRs overlapping genomic features was performed using the odds ratio statistic and Fisher's two-sided exact test on a 2 x 2 contingency table containing the number of CpGs within and outside of DMRs and the genomic feature (such as inside DMR and inside CpG island, inside DMR and outside CpG island, and so on). Gene set enrichment analysis (GSEA) of genes associated with DMRs was performed using gene sets from the GSEA database.^{74,75}

For analysis of methylation over canyon regions, the UCSC LiftOver tool⁷⁶ was used to identify hg19 genomic regions orthologous to murine canyons identified in DNMT3A-null HSCs obtained from Jeong et al.⁴¹ WGBS data from a healthy CD34+ HSPC control was included for visualization of a normal comparison. Meta-region plots were generated using genomation, and dmrseq's plotDMRs was used for visualization of individual canyon regions. Genomic coordinates for blocks identified in EBV-transformed B-cells were obtained from Hansen et al.⁴⁴ For enrichment analysis of canyons overlapping block edges, block edges were defined as 1 kb regions flanking blocks. Enrichment analysis was performed as above, with the 2 x 2 contingency table containing the number of CpGs inside canyons and inside block edges, inside canyons and outside block edges, and so on.

RNA-Sequencing

Cells were sorted from crushed whole bone marrow for RNA-seq at the indicate timepoints. Total RNA was extracted from >100,000 cells using DNA/RNA microprep plus kit (Zymo) following manufacturer's protocol. RNA quality control, SMART-Seq v4 library preparation, Illumina Next Generation Sequencing, and data quality control were performed at the MedGenome commercial sequencing facility. Sequencing reads were trimmed and quality control was performed using TrimGalore v.0.6.6 (<https://doi.org/10.5281/zenodo.5127899>). The cDNA index was built using Kallisto Bustools v.0.40.0⁷⁷ using gene annotations downloaded from Ensembl (release 110) and the GRCh37 reference genome. Pseudo-alignment was performed using Kallisto v0.44.0⁷⁸. Data was imported into R using tximport v.1.30.0⁷⁹ using gene mode. Differential testing was performed using DESeq2 v1.42.1.⁸⁰ Gene set enrichment analysis (GSEA) was performed using gene sets from the GSEA database.^{74,75} Methylation and expression were correlated by associating the methylation difference of putative DMRs in promoter regions to the fold change in expression of the corresponding genes. Pearson's correlation coefficient was calculated using cor.test in R.

Single-cell RNA-Sequencing and data processing

200-500 edited residual HSCs were added to 4 mL of MethoCult H4435 (Stemcell Technologies) and differentiated for 14 days. Cells were washed and live cells were sorted by FACS. Samples were then fixed using the Parse Biosciences Cell Fixation kit (Parse Biosciences); barcoding and library prep was then performed using the the Evercode Whole Transcriptome (v2) kit (Parse Biosciences). Resulting libraries were sequenced on a NovaSeq Illumina instrument using 150 cycle kits. Read trimming, quality control, alignment and count matrix generation was performed using the Parse Pipeline v.1.1.2. Data was processed using Seurat v5.0.3,⁸¹ pre-processing and normalization was performed using the SCTransformv2, putative doublets were removed using DoubletFinder v2.0.4⁸² and samples were integrated using HarmonyIntegration.⁸³ Cell types were assigned using Cell Marker Accordion⁸⁴ and GSEA was performed using fgsea.⁸⁵

Statistical analysis

Statistical analyses were performed in R version 4.3.1 or Prism10 (GraphPad Software). Paired/unpaired t-test was used to define statistical significance (*P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001). Enrichment analyses of genomic regions were performed as described above.

Acknowledgments: We would like to thank all patients and their families for providing AML specimens for this study. We thank the Stanford Veterinary Service Center for animal husbandry services. We thank the Flow Cytometry core at the Stanford Institute for Stem Cell Biology and Regenerative Medicine for providing flow cytometry training, equipment, and support. Some of the computing for this project was performed on the Sherlock cluster, and we thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results. We thank Kasper Hansen for his thoughtful review and suggestions. We also thank all members of the Majeti and Feinberg labs for helpful input, comments, and discussion.

Funding: German Research Foundation grant KO 5509/1-1, American Society of Hematology Research Restart Award, Leukemia and Lymphoma Special Fellow Career Development Award grant 3406-21 (TK); Japan Society for the Promotion of Science grant JP21J01690 (DK); Stanford Graduate Fellowship, NSF Graduate Research Fellowship Program, Stanford Lieberman Fellowship (ACF); Nakayama Foundation for Human Science, Stanford University School of Medicine Dean's Postdoctoral Fellowship (YN); NDSEG and Stanford Bio-X Fellowships (JLF); NIH Director's Early Independence Award (DP5OD024558), Packard Foundation Fellowship, Pew Scholar Award, Human Frontier Science Program Young Investigatorship, Baxter Foundation Faculty Scholar Award, and The Anthony DiGenova Endowed Faculty Scholar Fund (KML), National Institutes of Health grant 5R01CA054358 (AF); National Institutes of Health grant 1R01CA251331, Stanford Ludwig Center for Cancer Stem Cell Research and Medicine, Blood Cancer Discoveries Grant Program through The Leukemia & Lymphoma Society, The Mark Foundation for Cancer Research, The Paul G. Allen Frontiers Group (RM).

Author contributions: Conceived and performed experiments and/or analyzed the data: T.K., D.K., E.H., A.C.F., K.K., M.M., C.T.C., F.P.S., A.R., Y.F., Y.N., E.M.M., M.K., J.L.F., K.M.L., A.P.F., and R.M.; contributed reagents, materials, and analysis tools, J.L.F., and K.M.L.; wrote the paper, T.K., E.H., M.K., A.P.F., and R.M.

Declaration of interests: R.M. is on the Advisory Boards of Kodikaz Therapeutic Solutions, Orbital Therapeutics, Pheast Therapeutics, 858 Therapeutics, Prelude Therapeutics, Mubadala Capital, and Aculeus Therapeutics. R.M. is a co-founder and equity holder of Pheast Therapeutics, MyeloGene, and Orbital Therapeutics.

Data and materials availability: All unique reagents generated in this study are available from the lead contact without restriction. Sequencing data will be uploaded to GEO. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

5

Supplementary Materials

Figs. S1 to S6

Tables S1 to S14

Data S1 to S7

10

DNMT3A^{R882H} Is Not Required for Disease Maintenance in Primary Human AML, but Is Associated With Increased Leukemia Stem Cell Frequency

Fig. 1

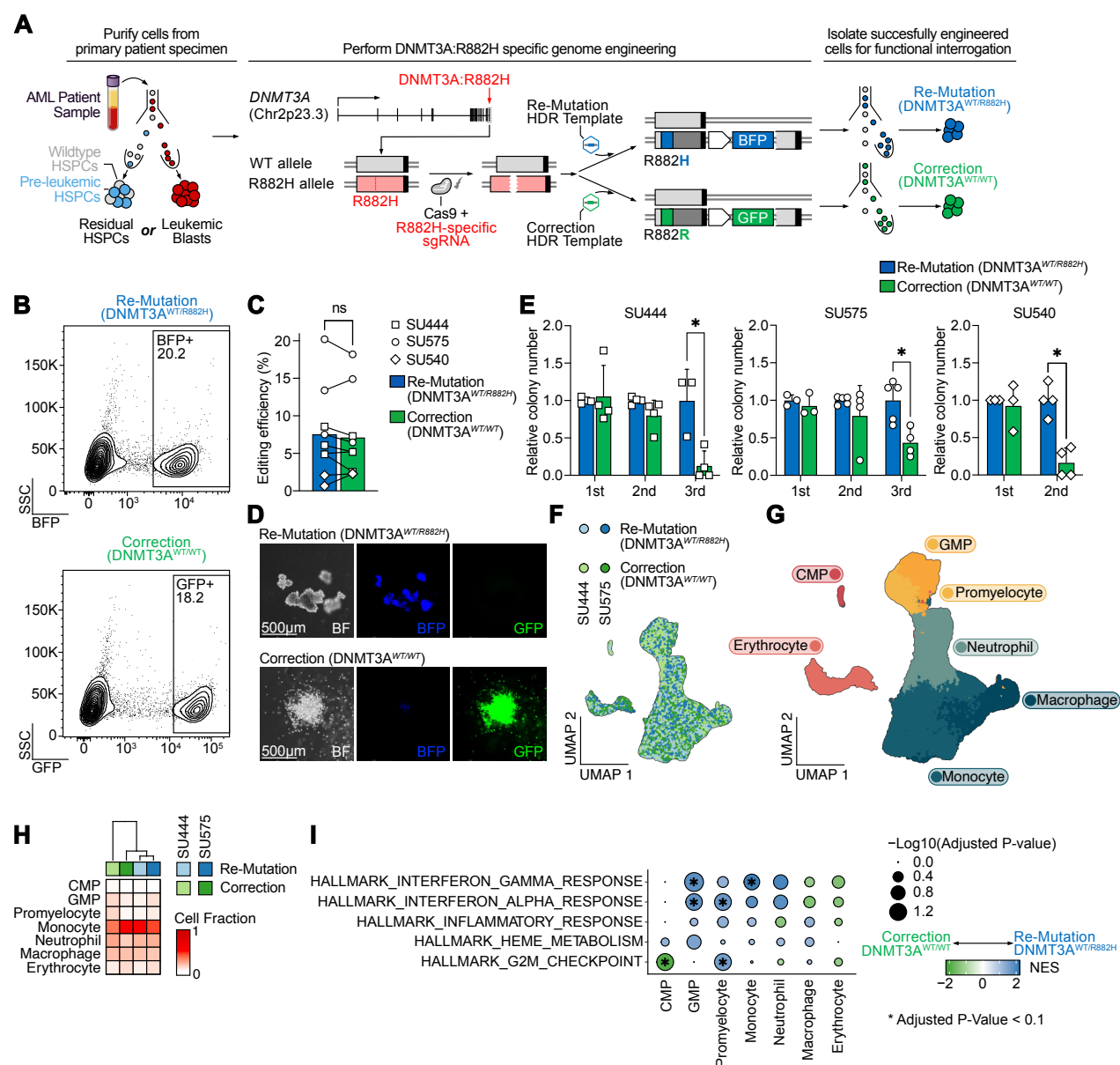


Fig. 1: Correction of *DNMT3A*^{R882H} in pre-leukemic HSCs results in normalization of aberrant self-renewal.

A: Schematic of *DNMT3A*^{R882H} correction approach. Cell populations of interest (either residual HSPCs containing pre-leukemic HSPCs or leukemic blasts) are purified by flow cytometry. Subsequently, pre-complexed Cas9 and sgRNA targeting the mutant, R882H allele, are electroporated into the target cells and cells are transduced with rAAV6 containing either the codon-optimized mutant (R882H) or wildtype (R882R) sequence, and a selectable fluorescent marker (BFP or GFP). Successfully engineered cells are then isolated for functional experiments. B: Isolation by FACS of successfully edited residual HSPCs (from patient SU575) for either re-mutation (top panel) or correction (bottom panel) of *DNMT3A*^{R882H}. C: Editing efficiency across all experiments in residual HSPCs as percentage of fluorescent protein positive cells (n=3).

DNMT3A^{R882H} Is Not Required for Disease Maintenance in Primary Human AML, but Is Associated With Increased Leukemia Stem Cell Frequency

independent patient specimens). D: Representative hematopoietic colonies 14 days after plating into semisolid media (Methocult). E: Serial replating assay of gene-edited residual HSPCs derived from three independent patient specimens (n=3-4 technical replicates per patient), * P < 0.05. F: UMAP embedding of single-cell RNA-Seq from gene-edited residual HSPCs differentiated in vitro for 14 days after editing colored by genotype and donor. G: UMAP embedding as in F colored by assigned cell type. H: Relative abundance of cell types within each sample. Sample identities and *DNMT3A* genotypes are indicated at the top. I: Gene set enrichment analysis of Hallmark gene sets between re-mutation (*DNMT3A*^{WT/R882H}, blue) and correction (*DNMT3A*^{WT/WT}, green) across cell types. * P < 0.1.

DNMT3A^{R882H} Is Not Required for Disease Maintenance in Primary Human AML, but Is Associated With Increased Leukemia Stem Cell Frequency

Fig. 2

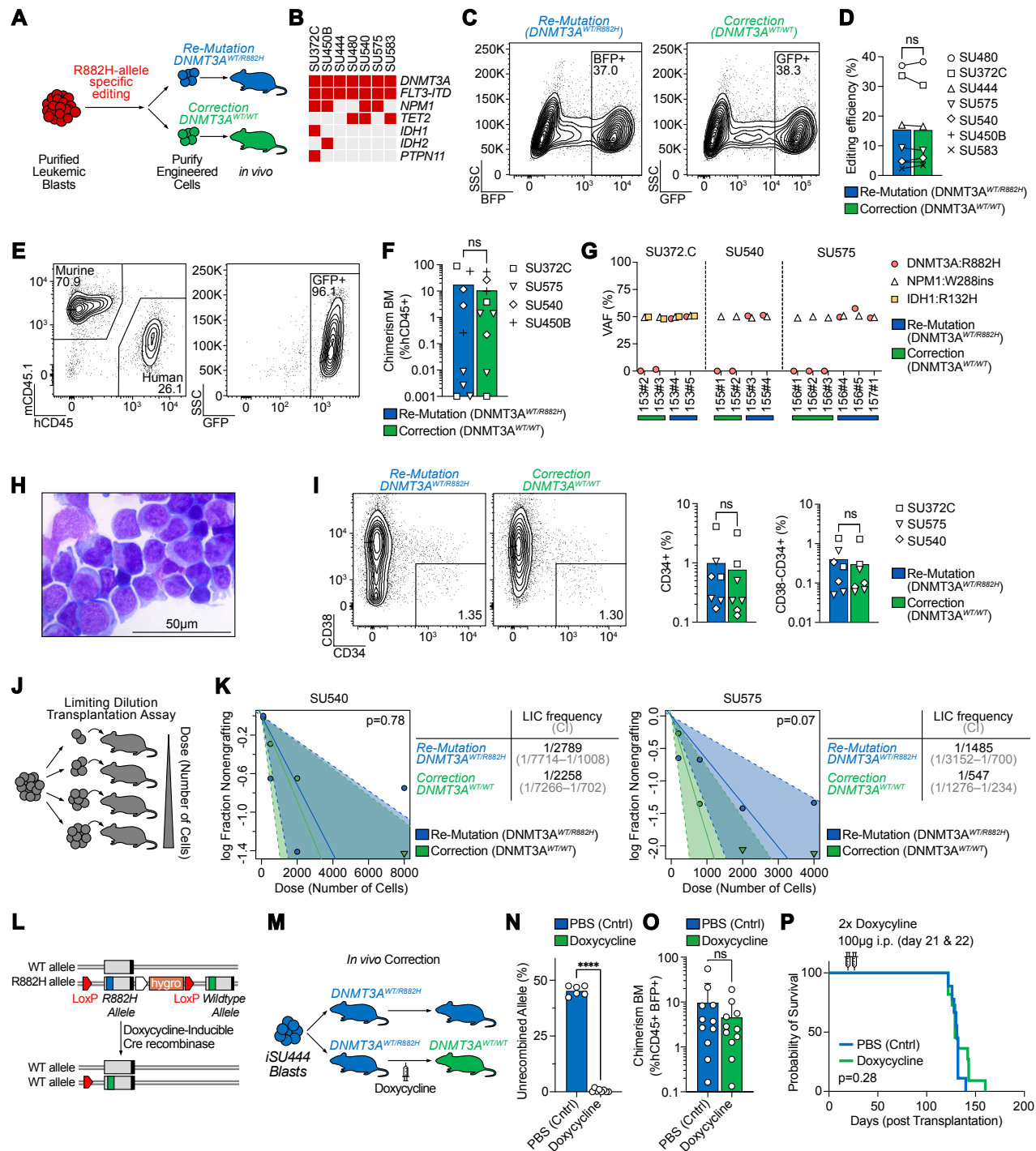


Fig. 2: Correction of *DNMT3A*^{R882H} in AML blasts has no effect on leukemogenicity or leukemia disease maintenance. A: Schematic of leukemic blast editing and xenotransplantation. B: Oncoprint of samples included in this study. C: Isolation of successfully edited leukemic blasts for either re-mutation (left panel) or correction (right panel) of *DNMT3A*^{R882H} from patient SU480. D: Editing efficiency across patient specimens quantified as percentage of fluorescent protein positive cells 3 days after

DNMT3A^{R882H} Is Not Required for Disease Maintenance in Primary Human AML, but Is Associated With Increased Leukemia Stem Cell Frequency

gene-editing. E: Representative flow cytometry analysis of bone marrow aspirate 12 weeks after injection of human, gene corrected (*DNMT3A*^{WT/WT}) leukemic blasts showing human engraftment with retention of the GFP fluorescent protein. F: Human chimerism in bone marrow aspirates at 12 weeks after injection of the indicated patient sample with either *DNMT3A* re-mutation or correction. G: Variant allele frequency (VAF) of somatic mutations found in the indicated patient specimens in human genomic DNA extracted from mice 4 months after engraftment indicating clonal retention of all somatic variants, except for *DNMT3A* in *DNMT3A*-corrected specimens. H: Representative Giemsa staining of cytopsin of human leukemic cells isolated 4 months after *DNMT3A*^{R882H} correction. I: Representative flow cytometry plots (left) and quantification (right) of CD34+ and CD34+CD38- leukemic subpopulations 4 months after re-mutation or correction of *DNMT3A*^{R882H}. J: Schematic of limiting dilution transplantation assay. K: Quantification of leukemia initiating cell (LIC) frequency for patient specimens SU540 (left) and SU575 (right) after re-mutation or correction of *DNMT3A*^{R882H}. Estimated frequency of LIC and test statistics are indicated. L: Schematic of doxycycline-inducible Cre recombinase driven correction of the endogenous *DNMT3A*^{R882H} allele in iPSC-derived AML blasts (iSU444). M: Schematic of *in vivo* correction experiment. Mice are engrafted with mutant (*DNMT3A*^{WT/R882H}) leukemic blasts, and administration of doxycycline 3 weeks after engraftment leads to correction of *DNMT3A*^{WT/R882H} to *DNMT3A*^{WT/WT} *in vivo*. N: Quantification of successful *in vivo* recombination of *DNMT3A*^{WT/R882H}, P < 0.0001. O: Human Chimerism 5 weeks after *in vivo* *DNMT3A*^{R882H} correction. P: Kaplan-Meier survival curve of mice with or without *in vivo* *DNMT3A*^{R882H} correction.

DNMT3A^{R882H} Is Not Required for Disease Maintenance in Primary Human AML, but Is Associated With Increased Leukemia Stem Cell Frequency

Fig. 3

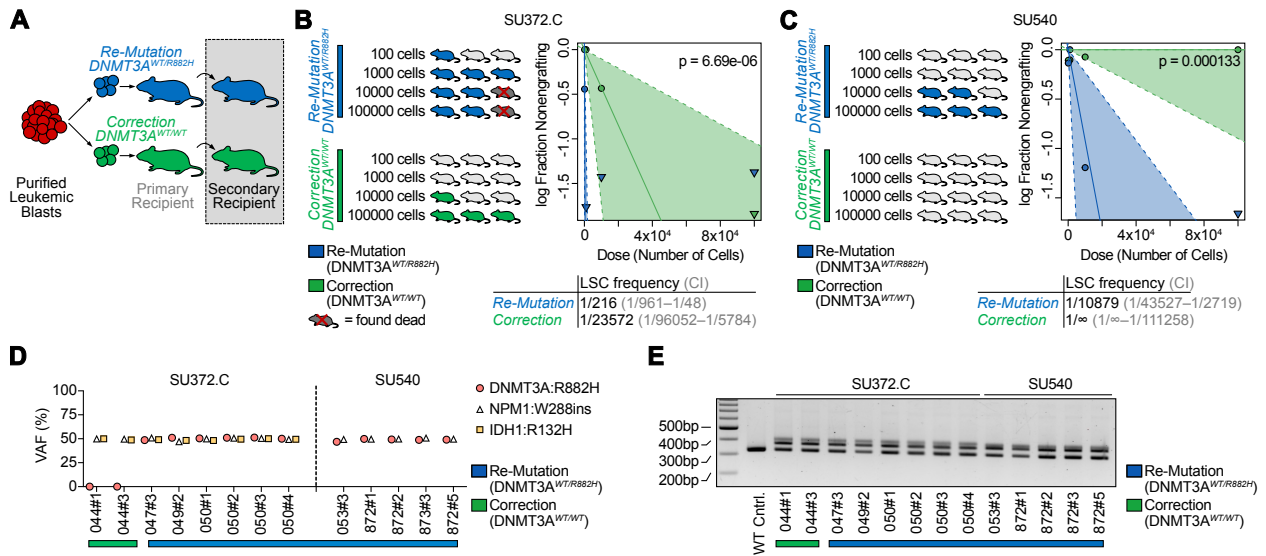


Fig. 3: Secondary transplantation reveals a reduction of leukemic stem frequency upon $DNMT3A^{R882H}$ correction. A: Schematic of secondary transplantation experiments. Human leukemic cells are harvested from primary recipients 4 months after gene-editing and then transplanted in limiting dilution into secondary host animals. B: Quantification of secondary limiting dilution assay results for patient specimen SU372.C. Estimated LSC frequencies and test statistics are indicated. C: Quantification of secondary limiting dilution assay results for patient specimen SU540. None of the mice receiving corrected ($DNMT3A^{WT/WT}$) leukemic cells showed engraftment at the doses indicated. Estimated LSC frequencies and test statistics are shown. D: VAF in human genomic DNA isolated from secondary recipients for recurrent somatic mutations found in patient samples utilized for secondary transplantation studies. E: *FLT3-ITD* PCR from human genomic DNA isolated from secondary recipients.

DNMT3A^{R882H} Is Not Required for Disease Maintenance in Primary Human AML, but Is Associated With Increased Leukemia Stem Cell Frequency

Fig. 4

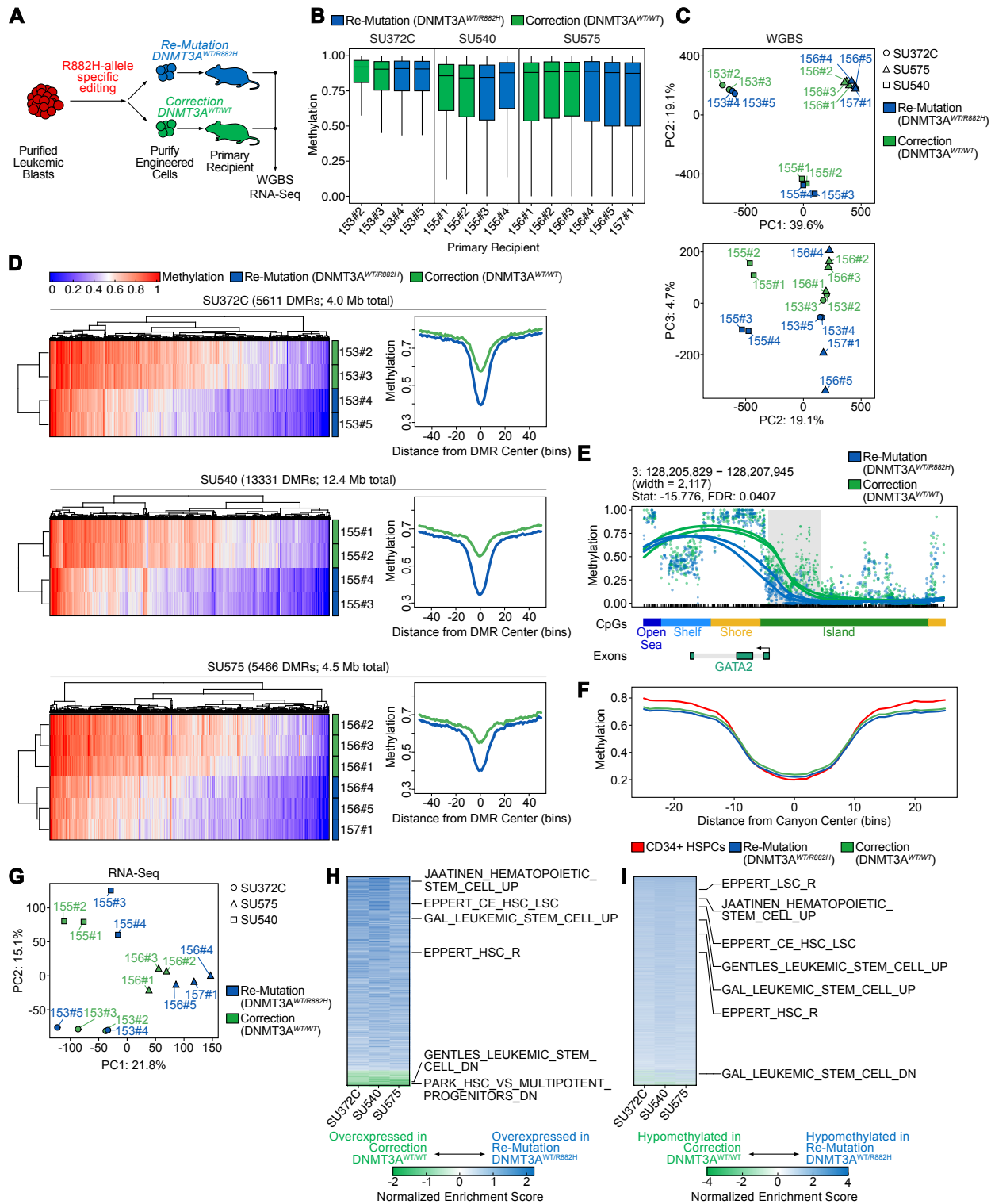


Fig. 4: Epigenetic and transcriptional identity is highly patient-specific with only minimal changes upon correction of *DNMT3A*^{R882H}. A: Schematic of sample generation for whole-genome bisulfite (WGBS) and RNA-Sequencing (RNA-Seq).

5

DNMT3A^{R882H} Is Not Required for Disease Maintenance in Primary Human AML, but Is Associated With Increased Leukemia Stem Cell Frequency

Human cells were harvested from recipient mice 4 months after gene-editing. B: Average CpG methylation across the human genome for three patient specimens with either correction (green) or re-mutation (blue). Each bar represents an individual sample harvested from a primary recipient mouse (mouse ID indicated on x-axis). C: Principal component analysis of global CpG methylation from engrafted mice. Patient ID is indicated by shape and *DNMT3A* genotype is indicated by color. D: Methylation at differentially methylated loci (heatmaps, left) for all DMRs identified within each patient sample as well as a summary meta-region plot of methylation at DMRs normalized for width (right). DMRs in meta-region plots are extended by 2 kb for visualization of flanking regions. E: Methylation at a representative DMR (adjacent to *GATA2*) for patient sample SU372.C. Individual lines indicate cells harvested from separate mice, color indicates the *DNMT3A* genotype. F: Summary meta-region plot of methylation at methylation canyon loci from Jeong et al.⁴¹ for re-mutated (blue) or corrected (green) patient samples, as well as healthy CD34+ HSPC control (red). G: Principal component analysis of global RNA-Seq patterns from engrafted mice. Patient ID is indicated by shape and *DNMT3A* genotype is indicated by color. H: Gene set enrichment analysis of differential gene expression within each patient plotted as a heatmap of normalized enrichment scores (NES) sorted by descending average. Gene sets enriched in re-mutated (*DNMT3A*^{WT/R882H}) are colored in blue, gene sets enriched in corrected (*DNMT3A*^{WT/WT}) are colored in green. Representative gene sets relating to hematopoietic or leukemic stem cell function are highlighted. I: Gene set enrichment analysis of differential promoter methylation within each patient as a heatmap of NES sorted by descending average. Gene sets enriched in hypomethylated DMRs in re-mutated (*DNMT3A*^{WT/R882H}) samples are colored in blue, and gene sets enriched in hypermethylated DMRs in re-mutated samples are colored in red. Representative gene sets relating to hematopoietic or leukemic stem cell function are highlighted.