

Metformin reduces the competitive advantage of *Dnmt3a*^{R878H} HSPCs

<https://doi.org/10.1038/s41586-025-08871-w>

Received: 18 January 2024

Accepted: 5 March 2025

Published online: 16 April 2025

 Check for updates

Mohsen Hosseini¹, Veronique Voisin^{2,11}, Ali Chegini^{1,3,11}, Angelica Varesi^{1,4,11}, Severine Cathelin^{1,11}, Dhanoop Manikoth Ayyathan¹, Alex C. H. Liu^{1,3}, Yitong Yang^{1,3}, Vivian Wang^{1,3}, Abdula Maher¹, Eric Grignano¹, Julie A. Reisz⁵, Angelo D'Alessandro⁵, Kira Young⁶, Yiyian Wu³, Martina Fiumara^{7,8}, Samuele Ferrari^{7,8}, Luigi Naldini^{7,8}, Federico Gaiti^{1,3}, Shraddha Pai^{2,9}, Grace Egan^{1,10}, Aaron D. Schimmer^{1,3}, Gary D. Bader², John E. Dick^{1,4}, Stephanie Z. Xie¹, Jennifer J. Trowbridge⁶ & Steven M. Chan^{1,3}✉

Clonal haematopoiesis arises when a haematopoietic stem cell (HSC) acquires a mutation that confers a competitive advantage over wild-type HSCs, resulting in its clonal expansion. Individuals with clonal haematopoiesis are at increased risk of developing haematologic neoplasms and other age-related inflammatory illnesses^{1–4}. Suppressing the expansion of mutant HSCs may prevent these outcomes; however, such interventions have not yet been identified. The most common clonal haematopoiesis driver mutations are in the *DNMT3A* gene, with arginine 882 (R882) being a mutation hotspot^{1–3,5–7}. Here we show that mouse haematopoietic stem and progenitor cells (HSPCs) carrying the *Dnmt3a*^{R878H/+} mutation, equivalent to human *DNMT3A*^{R882H/+}, have increased mitochondrial respiration compared with wild-type cells and are dependent on this metabolic reprogramming for their competitive advantage. Treatment with metformin, an anti-diabetic drug that inhibits mitochondrial respiration⁸, reduced the competitive advantage of *Dnmt3a*^{R878H/+} HSCs. Through a multi-omics approach, we found that metformin acts by enhancing methylation potential in *Dnmt3a*^{R878H/+} HSPCs and reversing the aberrant DNA CpG methylation and histone H3 K27 trimethylation profiles in these cells. Metformin also reduced the competitive advantage of human *DNMT3A*^{R882H} HSPCs generated by prime editing. Our findings provide preclinical rationale for investigating metformin as a preventive intervention against *DNMT3A* R882 mutation-driven clonal haematopoiesis in humans.

Mutations in *DNMT3A* are the most common genetic alterations in clonal haematopoiesis and are found in between 50% and 60% of clonal haematopoiesis carriers^{1–3,5,6}. *DNMT3A* encodes a de novo DNA methyltransferase that catalyses transfer of the methyl group from S-adenosylmethionine (SAM) to the C5 position of cytosines in DNA, resulting in 5-methylcytosine (5mC) and production of S-adenosylhomocysteine (SAH). *DNMT3A* mutations are classified into those that affect the mutational hotspot at R882 of the *DNMT3A* protein and those that affect other parts of the gene (non-R882 mutations)^{7,9}. Although both types of mutations are predicted to reduce methyltransferase activity, *DNMT3A* R882 mutations appear to confer a significantly higher risk of progression to acute myeloid leukaemia (AML) compared with non-R882 *DNMT3A* mutations^{10,11}. Thus, *DNMT3A* R882 mutations represent an important target for preventive intervention.

The mutations that affect R882 are almost invariably missense alterations and heterozygous^{12,13}. *DNMT3A* R882 mutations have been shown to both reduce the methyltransferase activity of the mutant protein and decrease the activity of the wild-type protein in a dominant-negative manner^{13,14}. Consistent with these findings, the differentially methylated regions (DMRs) in human AML cells or peripheral blood cells with *DNMT3A* R882 mutations are predominantly hypomethylated compared with their wild-type counterparts^{13,15}.

The effect of *DNMT3A* mutations on cell fate decisions of HSCs has previously been studied using genetically modified mouse models. In the *Dnmt3a*^{R878H/+} mouse model, the mutant HSCs are expanded and have a competitive advantage over wild-type HSCs¹⁶, thus recapitulating a key functional change associated with the mutation in humans. Here we used this model to identify differences in dependencies between

¹Princess Margaret Cancer Centre, Toronto, Ontario, Canada. ²Donnelly Centre for Cellular and Biomolecular Research, Toronto, Ontario, Canada. ³Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. ⁴Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. ⁵Department of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. ⁶The Jackson Laboratory, Bar Harbor, ME, USA. ⁷San Raffaele Telethon Institute for Gene Therapy, IRCCS San Raffaele Scientific Institute, Milan, Italy. ⁸Vita-Salute San Raffaele University, Milan, Italy. ⁹Ontario Institute for Cancer Research, Toronto, Ontario, Canada. ¹⁰Division of Haematology/Oncology, The Hospital for Sick Children, Toronto, Ontario, Canada. ¹¹These authors contributed equally: Veronique Voisin, Ali Chegini, Angelica Varesi, Severine Cathelin. ✉e-mail: steven.chan@uhn.ca

Dnmt3a^{R878H/+} and wild-type HSPCs with the goal of targeting such dependencies to selectively suppress the expansion of mutant HSCs.

Increased OXPHOS in *Dnmt3a*^{R878H/+} HSPCs

Analysis of publicly available RNA-sequencing (RNA-seq) datasets from primary AML samples revealed an increase in the expression of genes involved in oxidative phosphorylation (OXPHOS) in *DNMT3A* R882-mutated patient samples, but not in non-R882 *DNMT3A*-mutated samples, compared with *DNMT3A* wild-type samples (Extended Data Fig. 1a). To validate this finding in the context of clonal haematopoiesis, we analysed a previously published single-cell RNA-sequencing (scRNA-seq) dataset of CD34⁺-enriched HSPCs from individuals with clonal haematopoiesis¹⁷. Pseudobulk differential expression analysis also showed an increase in OXPHOS gene expression in *DNMT3A* R882-mutated cells relative to wild-type cells (Extended Data Fig. 1b).

These findings led us to explore the potential for differences in mitochondrial function between *Dnmt3a*^{R878H/+} and *Dnmt3a*^{+/+} mouse HSPCs. We found that *Dnmt3a*^{R878H/+} lineage-negative, KIT-positive (LK) cells, which are enriched for HSPCs, possessed higher levels of basal and maximal oxygen consumption rates (OCRs) than *Dnmt3a*^{+/+} LK cells, as determined by extracellular flux analysis (Fig. 1a). These differences were also observed in unfractionated whole bone marrow (WBM) cells, albeit by a smaller magnitude (Extended Data Fig. 1c). Furthermore, the level of mitochondrial reactive oxygen species (ROS) and the ratio of mitochondrial transmembrane potential ($\Delta\Psi_m$) to mitochondrial mass were higher in mutant LK cells than in wild-type LK cells (Fig. 1b,c). Together, these findings indicate that the *Dnmt3a*^{R878H} mutation causes metabolic reprogramming in HSPCs, resulting in upregulation of OXPHOS.

Dependency on mitochondrial respiration

We postulated that the increased mitochondrial respiration in *Dnmt3a*^{R878H} HSPCs is required for their competitive advantage over wild-type cells. To test this, we first established an in vitro competition assay in which CD45.2⁺ *Dnmt3a*^{R878H/+} and CD45.1⁺ *Dnmt3a*^{+/+} LK cells were mixed at a ratio of approximately 2:3 and cultured in cytokine-supplemented methylcellulose medium for around 10 days, followed by determination of the proportions of CD45.2⁺ and CD45.1⁺ cells (Extended Data Fig. 1d and Supplementary Fig. 1a). A parallel competition assay between CD45.2⁺ *Dnmt3a*^{+/+} and CD45.1⁺ *Dnmt3a*^{+/+} LK cells mixed at the same starting ratio served as a control. After the culture period, the proportion of CD45.2⁺ *Dnmt3a*^{R878H/+} cells was consistently 20% to 30% higher than that of CD45.2⁺ *Dnmt3a*^{+/+} cells in the first passage and 40% to 50% higher in the second passage (Extended Data Fig. 1e), demonstrating the competitive advantage of the mutant cells. Using this assay, we studied the effect of genetic knockdown of mitochondrial electron transport chain (ETC) subunits on the competitive advantage of *Dnmt3a*^{R878H/+} HSPCs by transducing the mixed CD45.2⁺ mutant–CD45.1⁺ wild-type population with lentiviral vectors expressing short hairpin RNAs (shRNAs) targeting *Ndufv1* and *Cox15* (Extended Data Fig. 2a), which encode critical subunits in complex I and complex IV of the ETC, respectively. Downregulation of these genes reduced the basal OCR and competitive advantage of CD45.2⁺ *Dnmt3a*^{R878H/+} cells (Fig. 1d,e), indicating that mutant HSPCs are dependent on increased OXPHOS to outcompete their wild-type counterparts.

To explore the translational relevance of this finding, we tested the effect of metformin—a commonly used oral anti-diabetic drug and pharmacologic inhibitor of complex I⁸—on *Dnmt3a*^{R878H/+} LK HSPCs. Consistent with the genetic knockdown studies, treatment with metformin at a clinically relevant concentration¹⁸ (50 μM) suppressed the competitive advantage of mutant cells in vitro (Fig. 1f). This effect was rescued by expression of NDI1, a metformin-resistant yeast analogue of complex I^{19,20}, thus confirming that the effect of metformin was due

to on-target complex I inhibition (Fig. 1g). Metformin treatment also selectively reduced the clonogenic potential of *Dnmt3a*^{R878H/+} LK HSPCs over *Dnmt3a*^{+/+} LK HSPCs in standard colony-forming unit (CFU) assays (Extended Data Fig. 2b).

To determine whether the effect of metformin was relevant in vivo and over a longer treatment period, we conducted a competitive repopulation experiment by mixing CD45.2⁺ *Dnmt3a*^{R878H/+} or *Dnmt3a*^{+/+} WBM cells with CD45.1⁺ *Dnmt3a*^{+/+} WBM cells at a 2:3 ratio and transplanting the mixed cells into lethally irradiated recipients (Extended Data Fig. 2c). Five weeks after transplantation, the recipient mice were either left untreated or started on treatment with metformin in their drinking water at 5 mg ml⁻¹, a concentration that has previously been shown to result in blood concentrations similar to those achievable in humans²¹. Peripheral blood chimerism analysis showed a stable ratio of CD45.2⁺ to CD45.1⁺ cells in mice that received CD45.2⁺ *Dnmt3a*^{+/+} control cells, and the ratio was not affected by metformin treatment (Fig. 1h and Supplementary Fig. 1b). By contrast, the ratio of CD45.2⁺ to CD45.1⁺ cells steadily increased over a 7-month period in mice that received CD45.2⁺ *Dnmt3a*^{R878H/+} cells, reflecting their competitive advantage over CD45.1⁺ wild-type cells (Fig. 1h). Notably, metformin decreased this competitive advantage up to 7 months (Fig. 1h). This effect was observed in both the myeloid and lymphoid compartments (Extended Data Fig. 2d and Supplementary Fig. 1b). Our findings collectively indicate that inhibition of mitochondrial respiration is a potential strategy for targeting *DNMT3A* R882 mutation-driven clonal haematopoiesis.

Metformin suppresses *Dnmt3a*^{R878H/+} HSCs

Metformin suppressed the long-term competitive advantage of *Dnmt3a*^{R878H/+} donor cells in vivo (Fig. 1h), reflective of an effect at the HSC level. To provide independent evidence for an effect at the HSC level and gain insights into the mechanism of action of metformin, we performed scRNA-seq analysis on LK-enriched bone marrow cells from the untreated and metformin-treated recipients at the end of the 7-month treatment period (Fig. 1h). The cells were collected from the mice transplanted with CD45.2⁺ *Dnmt3a*^{R878H/+} and CD45.1⁺ *Dnmt3a*^{+/+} competitor cells and stained with antibody–oligonucleotide conjugates specific for CD45.2 or CD45.1 to identify their donor origin. A total of 22,407 cells from untreated control mice ($n = 2$) were sequenced, 84.6% of which were CD45.2⁺ *Dnmt3a*^{R878H/+} cells (Fig. 2a). For comparison, a total of 23,818 cells from metformin-treated mice ($n = 2$) were sequenced, and the proportion of CD45.2⁺ *Dnmt3a*^{R878H/+} cells was significantly less at 56.5% ($P < 0.0001$ by chi-square test) (Fig. 2a). To determine which HSPC subsets were affected, we annotated each cell on the basis of their correlation with reference mouse HSPC gene sets²² and identified 11 different haematopoietic subsets (Fig. 2a). Metformin treatment reduced the ratio of CD45.2⁺ to CD45.1⁺ cells in the HSC cluster as well as myeloid progenitor subsets (Fig. 2b,c), consistent with its effect at the HSC level. Flow cytometric analysis of bone marrow samples collected from mice after 4 months of treatment also showed a reduction in the ratio of CD45.2⁺ to CD45.1⁺ cells in immunophenotypically defined (Lin⁻KIT^{+SCA-1⁺CD150⁺CD48⁻}) HSCs and progenitor subsets (Extended Data Fig. 3a,b).

To corroborate these findings, we transplanted *Dnmt3a*^{R878H/+} or *Dnmt3a*^{+/+} donor WBM cells from sex-matched littermates in a non-competitive manner into lethally irradiated recipients. Five weeks after transplantation, the recipients were either left untreated or started on treatment with metformin for 1 month (Extended Data Fig. 3c). In untreated mice, the number of immunophenotypic HSCs per femur and the proportion of HSCs in the Lin⁻SCA-1⁺KIT (LSK) fraction were higher in *Dnmt3a*^{R878H/+} recipients compared with *Dnmt3a*^{+/+} recipients (Fig. 2d,e). The expansion of mutant HSCs was associated with a trend towards a higher proportion of HSCs in active cycling based on Ki-67 staining (Fig. 2f). Metformin treatment reduced all these parameters in *Dnmt3a*^{R878H/+} recipients to levels comparable to those in untreated *Dnmt3a*^{+/+} recipients (Fig. 2d–f). Our findings demonstrate

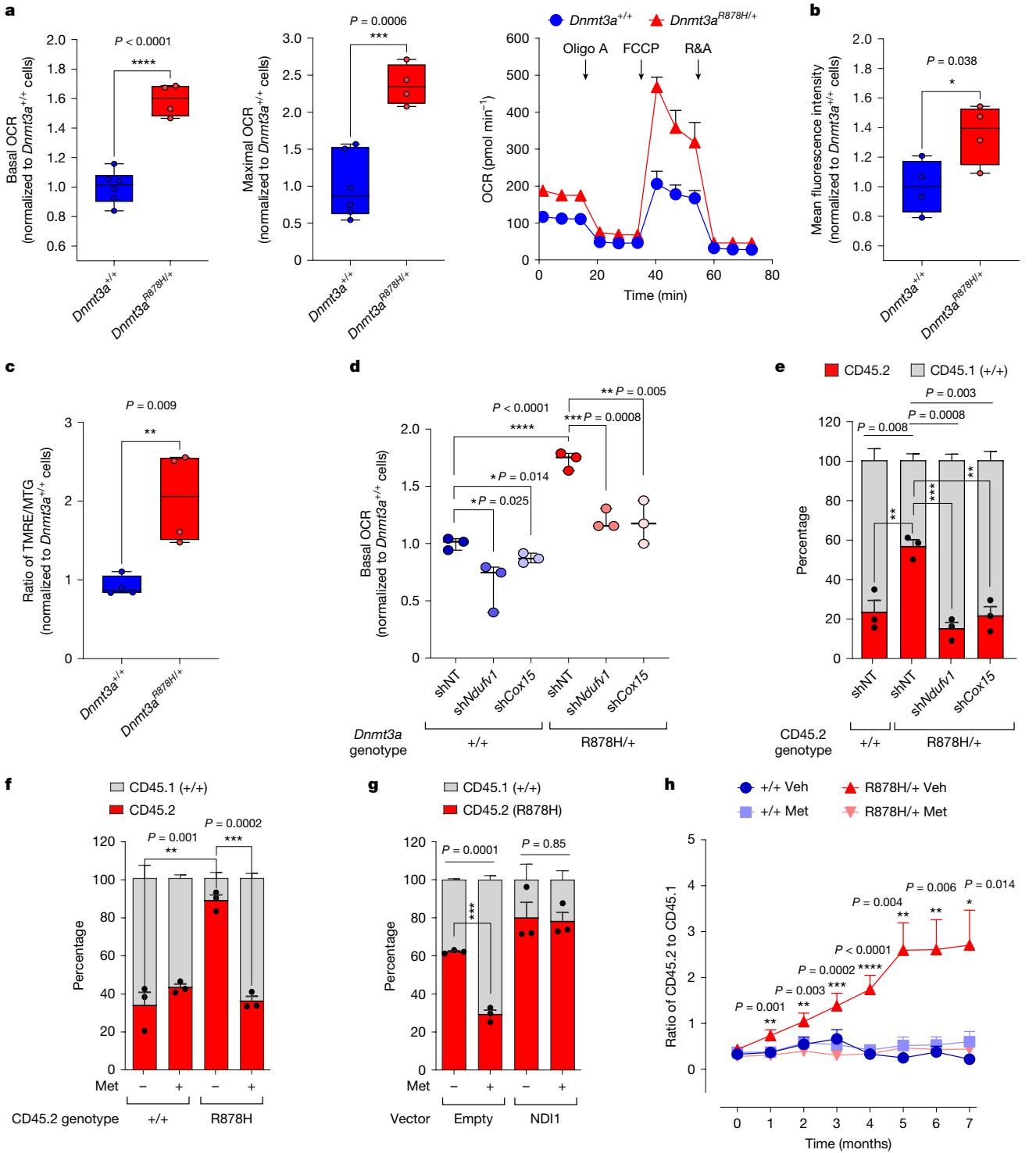


Fig. 1 | $Dnmt3a^{R878H/+}$ HSPCs are dependent on increased mitochondrial respiration. **a**, Basal (left) and maximal (middle) OCRs in LK HSPCs. Right, OCRs of LK cells at baseline and after treatment with the indicated drugs. $n = 4$ biologically independent samples for $Dnmt3a^{R878H/+}$ and $n = 6$ for $Dnmt3a^{+/+}$. Oligo A, oligomycin; R&A, antimycin A + rotenone. **b**, MitoSOX fluorescence in LK HSPCs. $n = 4$ biologically independent samples each. **c**, Ratio of tetramethylrhodamine ethyl ester (TMRE) to MitoTracker Green (MTG) fluorescence in LK HSPCs. $n = 4$ biologically independent samples each. **d**, Basal OCR in LK HSPCs expressing shNT, or shRNA targeting *Ndufv1* (sh*Ndufv1*) or *Cox15* (sh*Cox15*). $n = 3$ biologically independent samples for each condition. **e**, Proportion of CD45.2⁺ and CD45.1⁺ cells transduced with the indicated shRNA vectors in a competition assay. $n = 3$ biologically independent samples each. **f**, Proportion of CD45.2⁺ and CD45.1⁺ cells in a competition assay with or without metformin (Met). $n = 3$ biologically independent samples each.

g, Proportion of CD45.2⁺ and CD45.1⁺ cells in a competition assay with or without metformin. The CD45.2⁺ LK cells were transduced with an empty or NDI1-expressing lentiviral vector. $n = 3$ biologically independent samples each. **h**, Ratio of CD45.2⁺ to CD45.1⁺ in peripheral blood cells collected at the indicated time points after starting treatment with metformin or vehicle (Veh). For months 0–4, data are from 3 independent experiments and 21–24 mice per condition. For months 5–7, data are from one of the 3 experiments and 6–8 mice per condition. Exact n values are provided in the Source Data. Statistical significance was determined in comparison to the R878H/+ Veh group. In box plots in **a–d**, the box represents the interquartile range with the median indicated by the line inside the box; whiskers extend to the minimum and maximum values. In **a**, right, **e–h**, data are mean \pm s.e.m. Statistical significance was calculated using two-sided (**a–c**, **e–h**) or one-sided (**d**) Student's *t*-tests. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$.

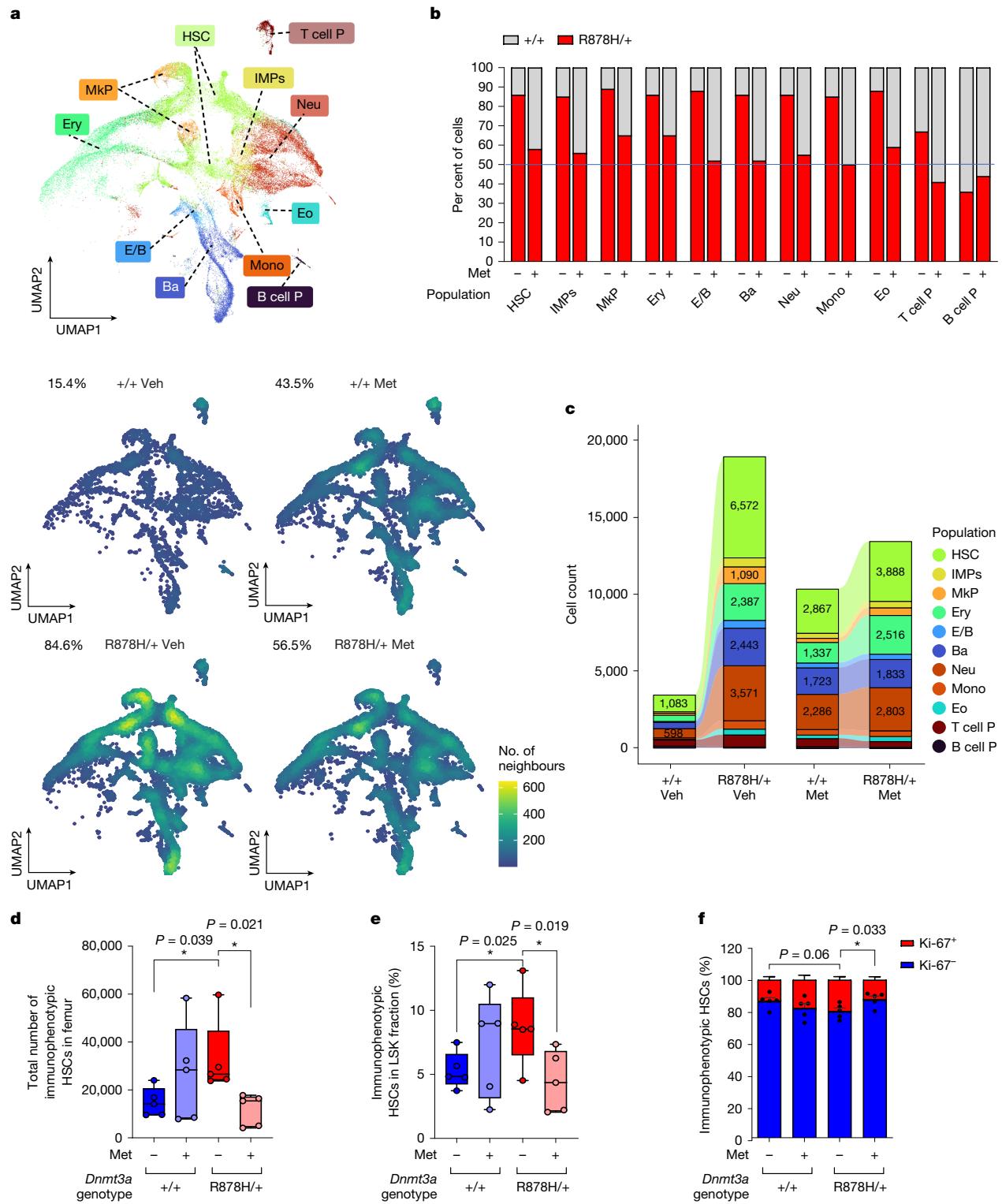


Fig. 2 | Metformin suppresses the competitive advantage of $Dnmt3a^{R878H/+}$ HSCs. **a**, Top, dimensionality reduction using uniform manifold approximation and projection (UMAP) on all sequenced cells ($n = 46,225$ cells). Bottom, UMAP cell density plots of CD45.1 $^{+}$ $Dnmt3a^{+/+}$ cells versus CD45.2 $^{+}$ $Dnmt3a^{R878H/+}$ cells in LK-enriched bone marrow samples collected from mice treated with vehicle or metformin. B cell P, B cell progenitor; Ba, basophil progenitor; E/B, erythroid/basophil progenitor; Eo, eosinophil progenitor; Ery, erythroid progenitor; IMPs, immature myeloid progenitors; MkP, megakaryocyte progenitor; Mono, monocyte progenitor; Neu, neutrophil/granulocyte progenitor; T cell P, T cell progenitor. **b**, Proportion of CD45.1 $^{+}$ $Dnmt3a^{+/+}$ cells versus CD45.2 $^{+}$ $Dnmt3a^{R878H/+}$ cells in each transcriptomically defined HSC subset from untreated and metformin-treated LK samples. **c**, Sankey diagrams showing

the absolute number of sequenced cells in each HSC subset among CD45.1 $^{+}$ $Dnmt3a^{+/+}$ versus CD45.2 $^{+}$ $Dnmt3a^{R878H/+}$ fractions in LK-enriched bone marrow samples collected from mice treated with vehicle or metformin. **d**, Number of immunophenotypic HSCs ($\text{Lin}^{-}\text{KIT}^{+}\text{SCA-1}^{+}\text{CD150}^{+}\text{CD48}^{-}$) in the femur from mice transplanted with WBM cells of the indicated genotype and treated with or without metformin. $n = 5$ biologically independent samples each. **e**, Proportion of immunophenotypic HSCs in the LSK fraction of samples from **d**. **f**, Proportion of immunophenotypic HSCs with positive or negative Ki-67 staining in samples from **d**. In box plots in **d**, **e**, the box represents the interquartile range with the median indicated by the line inside the box; whiskers extend to the minimum and maximum values. In **f**, data are mean \pm s.e.m. Statistical significance was calculated using two-sided (**d**, **f**) or one-sided (**e**) Student's *t*-tests.

that metformin treatment suppresses the competitive advantage of *Dnmt3a*^{R878H/+} HSCs.

Metformin increases methylation capacity

Gene set enrichment analysis (GSEA) of the scRNA-seq data revealed an enrichment of genes associated with OXPHOS in *Dnmt3a*^{R878H/+} HSCs relative to wild-type HSCs and a decrease in expression of these genes with metformin treatment (Extended Data Fig. 4a,b). This unexpected result suggested that metformin could influence mitochondrial respiration not only through a direct inhibition of complex I but also through downregulation of OXPHOS-related genes. To further investigate its mechanism of action, we first studied the effect of in vivo metformin treatment on mitochondrial function of HSPCs. We performed extracellular flux analysis on freshly isolated LK-enriched bone marrow cells from the mice that were untreated or treated with metformin for 1 month (Extended Data Fig. 3c). Metformin treatment reduced the basal and maximal OCRs as well as the $\Delta \Psi_m$ of *Dnmt3a*^{R878H/+} LK cells to levels similar to those of untreated *Dnmt3a*^{+/+} LK cells (Extended Data Fig. 4c,d). To uncover the effect of metformin on specific metabolic pathways, we performed a mass spectrometry-based metabolomic analysis of the untreated and treated LK cells of both *Dnmt3a* genotypes. This analysis, which focused on metabolites that are central to energy and redox metabolism, detected 101 named metabolites, of which 8 were significantly increased in metformin-treated mutant LK cells compared with untreated mutant cells (Supplementary Table 1). Of note, 4 out of the 8 up-regulated metabolites (L-cysteate, dimethylglycine, reduced glutathione (GSH) and taurine) are involved in one-carbon (1C) metabolism through the methionine cycle (Fig. 3a,b). Since the methionine cycle generates SAM, these findings suggest that metformin could potentially affect SAM levels and the ratio of SAM to SAH, which is also known as the methylation index, an indicator of cellular methylation potential. To test this hypothesis, we directly measured the intracellular concentrations of SAM and SAH, which were below the detection threshold of the bulk metabolomic analysis. Consistent with our hypothesis, the methylation index was higher in metformin-treated *Dnmt3a*^{R878H/+} LK cells than in untreated cells (Fig. 3c), indicative of an increase in their cellular methylation potential. The effect of metformin on methylation index was observed only in *Dnmt3a*^{R878H/+} LK HSPCs and not in *Dnmt3a*^{+/+} LK cells (Fig. 3c).

To determine whether the metformin-induced changes in methylation index could be due to alterations in the expression of genes involved in 1C metabolism, we performed bulk RNA-seq analysis of *Dnmt3a*^{+/+} and *Dnmt3a*^{R878H/+} LK-enriched cells that were untreated or treated with metformin for 1 month (Extended Data Fig. 3c). GSEA of the RNA-seq dataset showed that metformin treatment decreased the expression of genes associated with stemness and OXPHOS (Extended Data Fig. 4e,f), consistent with our earlier results. Notably, it also revealed a significant enrichment of genes involved in 1C metabolism in metformin-treated *Dnmt3a*^{R878H/+} LK cells (Fig. 3d). To confirm these findings, we performed quantitative PCR with reverse transcription (RT-qPCR) to measure the expression of eight genes that encode enzymes in the folate and methionine cycles (*Shmt2*, *Mthfd2l*, *Shmt1*, *Mthfd1*, *Mthfr*, *Ahcy*, *Cbs* and *Bhmt*) and found that metformin treatment increased their expression in *Dnmt3a*^{R878H/+} LK cells (Fig. 3e). Although metformin also up-regulated expression of these genes in *Dnmt3a*^{+/+} LK cells, the magnitude of change was much smaller (Fig. 3e). These findings suggest that metformin selectively increases the cellular methylation potential of mutant HSPCs by upregulating the expression of genes involved in 1C metabolism.

On the basis of the above findings, we hypothesized that metformin suppresses the competitive advantage of *Dnmt3a*^{R878H/+} cells by increasing their cellular methylation potential. To test this hypothesis, we investigated the effect of exogenous SAM and SAH on the competitive advantage of mutant LK cells using our in vitro assay (Extended Data Fig. 1d). The addition of exogenous SAM, which increases the

methylation index, was sufficient to reduce the competitive advantage of *Dnmt3a*^{R878H/+} HSPCs (Extended Data Fig. 5a). The suppressive effect of SAM on *Dnmt3a*^{R878H/+} HSPCs was maintained after downregulation of *Dnmt3b* expression (Extended Data Fig. 5b,c), suggesting that this effect was independent of DNMT3B activity. Increasing the concentration of folic acid in the culture medium, which augments the availability of 1C units for SAM synthesis, also reduced the competitive advantage of *Dnmt3a*^{R878H/+} HSPCs (Extended Data Fig. 5d). Conversely, exogenous SAH, which lowers the methylation index, counteracted the suppressive effect of metformin on mutant cells (Extended Data Fig. 5a). To confirm these findings, we inhibited SHMT2 and MAT2A as alternative approaches to lowering the [SAM]/[SAH] ratio. SHMT2 generates a 1C unit (5,10-methylenetetrahydrofolate) that is necessary for SAM synthesis through the folate and methionine cycles, and MAT2A directly catalyses the synthesis of SAM from methionine (Fig. 3a). In line with our hypothesis, pharmacologic inhibition of SHMT with SHIN-1²³ or MAT2A with AG-270²⁴ rescued the suppressive effect of metformin on mutant HSPCs (Extended Data Fig. 5e,f). The addition of exogenous SAM mitigated the rescue effect of SHIN-1 (Extended Data Fig. 5g), indicating that SHIN-1 acts by reducing SAM production. Genetic knockdown of *Shmt2* expression also rendered *Dnmt3a*^{R878H/+} HSPCs resistant to the suppressive effect of metformin (Extended Data Fig. 5h,i). Together, these findings support a mechanism in which metformin selectively reduces the competitive advantage of *Dnmt3a*^{R878H/+} HSPCs by increasing their cellular methylation potential.

Reversal of aberrant epigenetic profiles

The mechanism by which *DNMT3A* mutations confer a competitive advantage to mutant HSCs is believed to be mediated through focal DNA hypomethylation secondary to reduced de novo DNA methylation activity^{13,17}. In the context of *DNMT3A* R882 mutants in which a copy of the wild-type gene remains, the residual DNA methylation activity is estimated to be around 20% of normal, but is not absent¹³. We hypothesized that the metformin-induced increase in methylation index could augment this activity, resulting in a reversal of the aberrant DNA CpG hypomethylation pattern in mutant cells and consequent decrease in their competitive advantage. A prediction based on this hypothesis is that elimination of residual DNMT3A activity by inactivating the remaining wild-type *Dnmt3a* allele in mutant cells should render them resistant to the effect of metformin. To test this prediction, we used CRISPR-Cas9 to knock out *Dnmt3a* in *Dnmt3a*^{+/+} and *Dnmt3a*^{R878H/+} LK cells. The knockout efficiency was approximately 98%, indicative of biallelic inactivation in both cell types (Extended Data Fig. 6a). Using our in vitro competition assay, we found that *Dnmt3a* knockout significantly increased the competitive advantage of LK HSPCs, regardless of their initial *Dnmt3a* genotype (Extended Data Fig. 6b). Metformin did not suppress the competitive advantage of *Dnmt3a*^{R878H/+} cells following *Dnmt3a* knockout, supporting the idea that residual DNMT3A activity is required for the metformin effect (Extended Data Fig. 6b). Another prediction is that metformin treatment should increase the level of methylation at CpG sites that are differentially hypomethylated in *Dnmt3a*^{R878H/+} cells. To test this, we performed reduced representation bisulfite sequencing (RRBS) analysis of LK-enriched bone marrow cells from recipient mice that received *Dnmt3a*^{R878H/+} or *Dnmt3a*^{+/+} WBM cells from sex-matched littermate donors and were either untreated or treated with metformin for 1 month (Extended Data Fig. 3c). The RRBS technique, which enriches for CpG-rich regions, was chosen because DNMT3A preferentially catalyses DNA methylation at CpG dinucleotides. We identified 5,430 DMRs in the comparison between untreated *Dnmt3a*^{R878H/+} ($n = 4$) and untreated *Dnmt3a*^{+/+} ($n = 3$) samples. Consistent with prior reports^{13,17}, the majority of the DMRs ($n = 4,649$; 85.6%) were hypomethylated in the untreated *Dnmt3a*^{R878H/+} samples (Fig. 4a). In the comparison between metformin-treated *Dnmt3a*^{R878H/+} ($n = 3$) and untreated *Dnmt3a*^{R878H/+} ($n = 4$) samples, we identified 3,285 DMRs, 1,923

Article

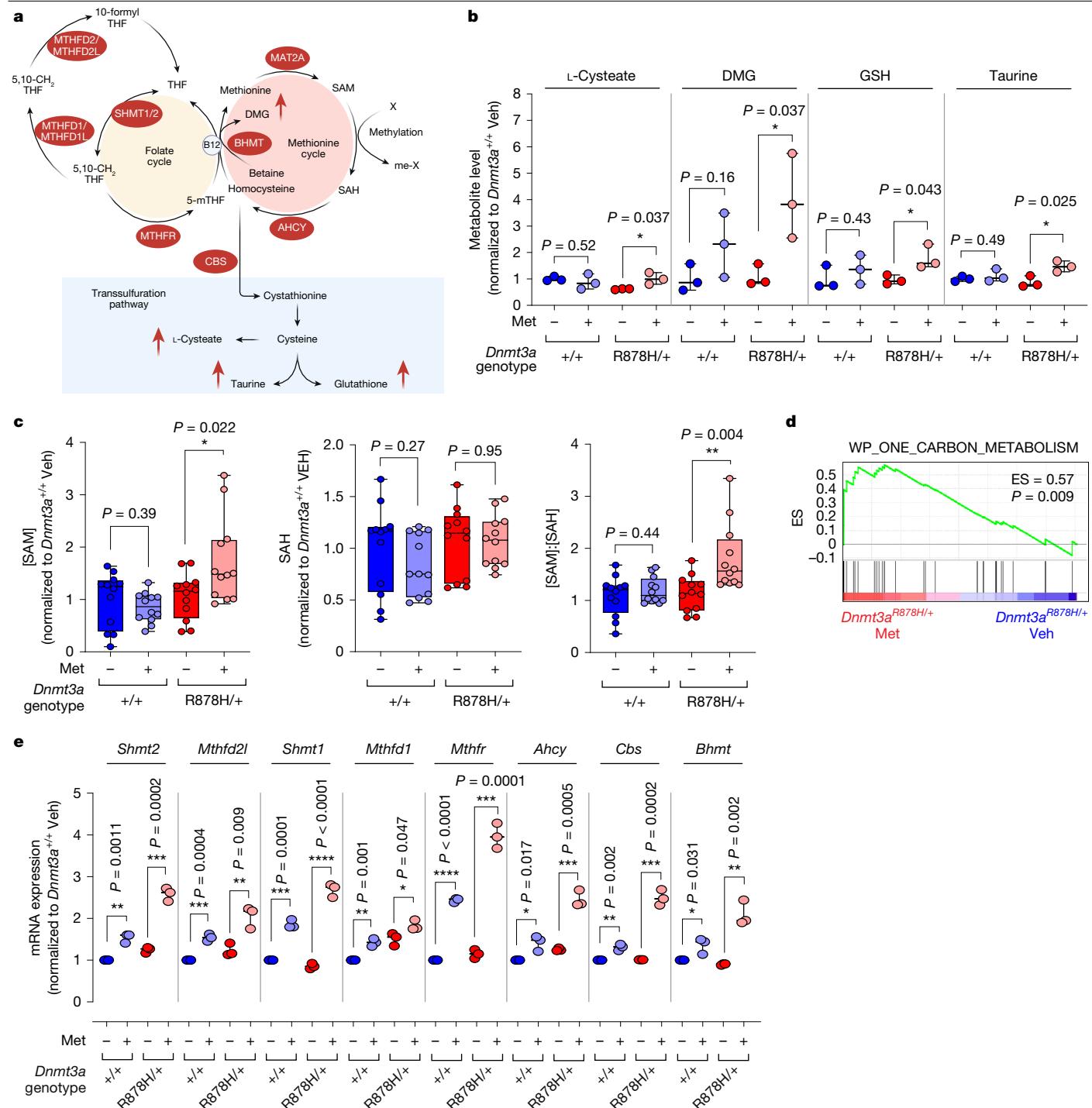


Fig. 3 | Metformin suppresses the competitive advantage of *Dnmt3a*^{R878H/+} HSPCs by enhancing their methylation potential. **a**, Schematic of the metabolic pathways involved in 1C metabolism. B12, vitamin B12; DMG, dimethylglycine; me, methyl; mTHF, methyltetrahydrofolate; THF, tetrahydrofolate. Created in BioRender. Chan, S. (2025) <https://BioRender.com/a57k225>. **b**, Quantification of the indicated metabolites in LK cells isolated from mice transplanted with bone marrow cells of the indicated genotype. The mice were either left untreated or treated with metformin for 1 month. $n=3$ biologically independent samples per condition. **c**, Levels of SAM and SAH and the SAM:SAH ratio in LK cells isolated from mice transplanted with bone marrow cells of the indicated genotype. The mice were either left untreated or treated with metformin for 1 month. $n=12$ biologically independent samples.

for each condition. **d**, Gene set enrichment plot of bulk RNA-seq data comparing metformin-treated *Dnmt3a*^{R878H/+} LK cells ($n=3$ biologically independent samples) versus vehicle-treated *Dnmt3a*^{R878H/+} LK cells ($n=3$ biologically independent samples) using the indicated gene set (WP435). ES, enrichment score. **e**, Expression level of the indicated genes by RT-qPCR in LK cells isolated from mice transplanted with bone marrow cells of the indicated genotype. The mice were either left untreated or treated with metformin for 1 month. $n=3$ biologically independent samples for each condition. In **b,c,e**, the box represents the interquartile range with the median indicated by the line inside the box; whiskers extend to the minimum and maximum values. Statistical significance was calculated using two-sided Student's *t*-tests.

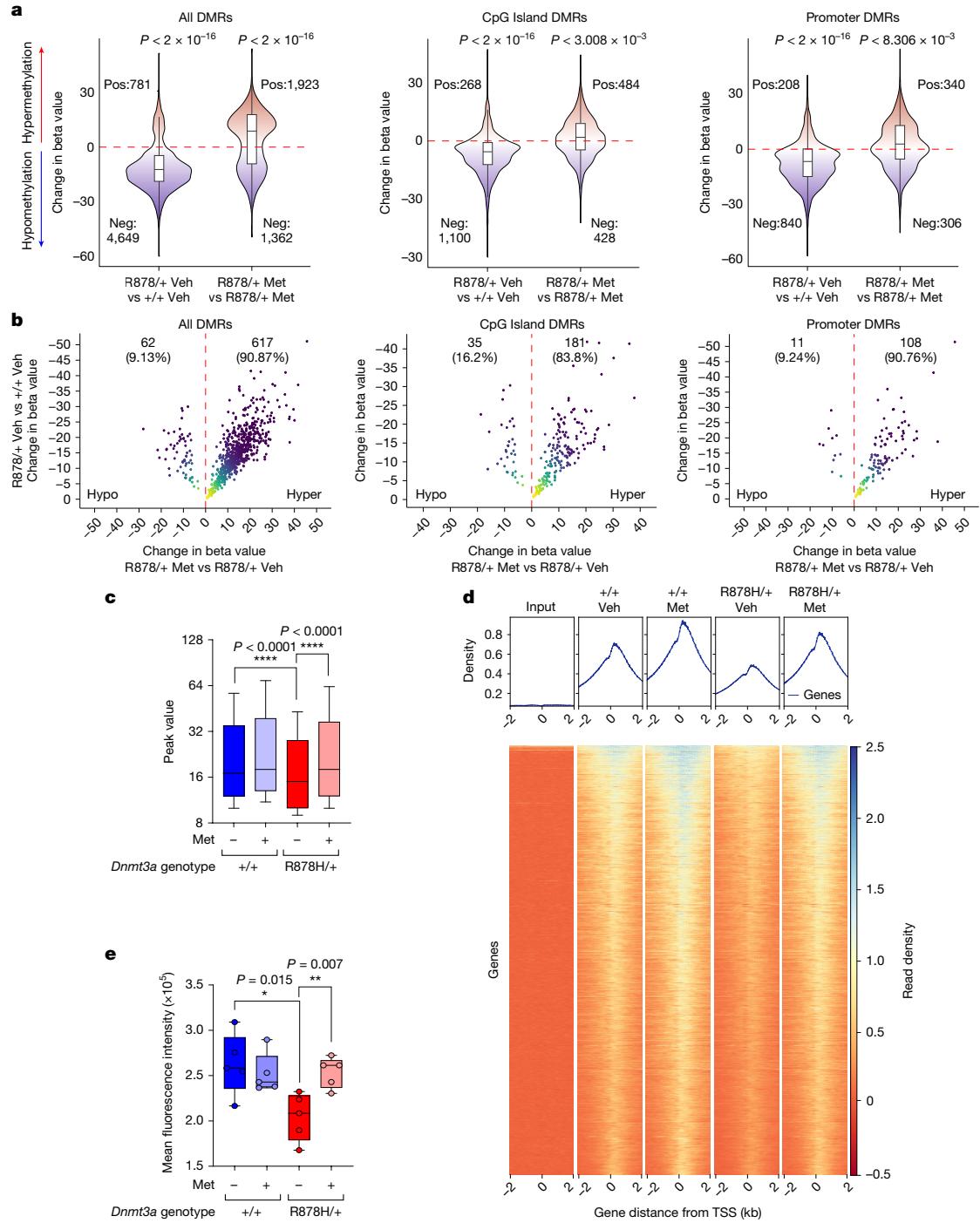


Fig. 4 | Metformin reverses the aberrant epigenetic profiles in *Dnmt3a*^{R878H/+} HSPCs. **a**, Violin plots of the difference in beta values at all DMRs, CpG island-associated DMRs or promoter-associated DMRs in the comparison between untreated *Dnmt3a*^{R878H/+} LK samples versus untreated *Dnmt3a*^{+/+} LK samples and between metformin-treated *Dnmt3a*^{R878H/+} LK samples versus untreated *Dnmt3a*^{R878H/+} LK samples. $n = 3$ biologically independent samples for untreated *Dnmt3a*^{+/+} and metformin-treated *Dnmt3a*^{R878H/+}. $n = 4$ biologically independent samples for metformin-treated *Dnmt3a*^{+/+} and untreated *Dnmt3a*^{R878H/+} samples. P values are calculated using the two-tailed one-sample Wilcoxon signed-rank test to determine whether the median was significantly different from 0. Pos, positive change in methylation state; Neg, negative change in methylation state. **b**, Plot showing the change in beta values at overlapping DMRs between metformin-treated *Dnmt3a*^{R878H/+} samples versus untreated *Dnmt3a*^{R878H/+} samples on the x axis and between untreated *Dnmt3a*^{R878H/+} samples versus untreated *Dnmt3a*^{+/+} samples on the reverse y axis. **c**, Peak values from

H3K27me3 ChIP-seq analysis of LK HSPC samples collected from mice transplanted with bone marrow cells of the indicated genotype and treated with or without metformin. The box represents the 10th–90th percentile range with the median indicated by the line inside the box and whiskers extend to the minimum and maximum values. Statistical significance was calculated using the Mann–Whitney test. **d**, Distribution of H3K27me3 signals surrounding (± 2 kb) the TSS regions with the strongest signals ($n = 10,622$) in the indicated samples. Darker blue indicates higher read densities, while red corresponds to lower read densities. **e**, Intracellular H3K27me3 staining of LK HSPCs collected from mice transplanted with bone marrow cells of the indicated genotype and treated with or without metformin. $n = 5$ biologically independent samples for each condition. The box represents the interquartile range with the median indicated by the line inside the box; whiskers extend to the minimum and maximum values. Statistical significance was calculated using two-sided Student's *t*-tests.

Article

(58.5%) of which were hypermethylated in the treated samples (Fig. 4a). We found 870 overlapping DMRs at the intersection between these two sets (Extended Data Fig. 6c). In line with our hypothesis, metformin treatment increased the methylation level at 617 (90.9%) of the 679 hypomethylated DMRs in *Dnmt3a^{R878H/+}* samples, bringing their average methylation levels close to those of untreated *Dnmt3a^{+/+}* samples (Fig. 4b and Extended Data Fig. 6d). Similar findings were observed in the subsets of DMRs associated with CpG islands and gene promoter regions (Fig. 4a,b and Extended Data Fig. 6c,d). To explore the potential effect of these changes in DNA methylation on gene expression, we conducted GSEA on genes ranked by differential promoter methylation and mRNA expression from our bulk RNA-seq analysis. This analysis revealed that *Dnmt3a^{R878H/+}* cells, compared with *Dnmt3a^{+/+}* cells, exhibited decreased promoter methylation and increased expression of genes involved in mitochondrial respiration (Extended Data Fig. 7a). Re-analysis of a previously published combined transcriptome and methylome dataset of single progenitor cells from individuals with *DNMT3A* R882-mutated clonal haematopoiesis also revealed similar patterns in changes in the methylation and expression of genes associated with OXPHOS¹⁷ (Extended Data Fig. 7b). Analysis of our bulk RRBS and RNA-seq datasets also showed promoter hypermethylation and decreased expression of genes associated with 1C metabolism in *Dnmt3a^{R878H/+}* cells relative to *Dnmt3a^{+/+}* cells (Extended Data Fig. 7a). Notably, metformin treatment reversed these aberrant changes in DNA methylation and gene expression in mutant cells (Extended Data Fig. 7a).

In human clonal haematopoiesis, the *DNMT3A* R882 mutation has previously been reported to result in preferential DNA hypomethylation of targets of the polycomb repressive complex 2 (PRC2)¹⁷, which catalyses the methylation of histone H3K27 (H3K27). In addition, the *Dnmt3a^{R878H}* mutation was previously found to be associated with a reduction in H3K27 trimethylation (H3K27me3)²⁵, indicating its potential influence on another layer of epigenetic regulation. Given that PRC2-mediated methylation activity is also regulated by the [SAM]/[SAH] ratio, we hypothesized that metformin could reverse the aberrant H3K27 hypomethylation profile in *Dnmt3a^{R878H/+}* HSPCs. To test this, we performed H3K27me3 chromatin immunoprecipitation followed by sequencing (ChIP-seq) analysis of *Dnmt3a^{R878H/+}* and *Dnmt3a^{+/+}* LK-enriched bone marrow cells from the mice that were either untreated or treated with metformin for 1 month (Extended Data Fig. 3c). This analysis revealed a reduction in H3K27me3 levels globally and in the regions surrounding transcription start sites (TSSs) in the untreated *Dnmt3a^{R878H/+}* samples relative to the untreated *Dnmt3a^{+/+}* samples (Fig. 4c,d). Consistent with our hypothesis, metformin treatment restored H3K27me3 levels in *Dnmt3a^{R878H/+}* samples to a level resembling that of untreated *Dnmt3a^{+/+}* samples (Fig. 4c,d). To confirm these findings using an orthogonal approach, we measured the amount of H3K27me3 by intracellular flow cytometry. Similar to the ChIP-seq results, we found that *Dnmt3a^{R878H/+}* LK cells had lower levels of H3K27me3 than *Dnmt3a^{+/+}* LK cells and metformin treatment restored H3K27me3 levels in mutant cells to levels resembling those of wild-type cells (Fig. 4e). Together, the above findings demonstrate that metformin treatment can reverse the aberrant epigenetic landscape in *Dnmt3a^{R878H/+}* HSPCs.

Metformin suppresses human mutant HSPCs

To explore the relevance of our findings in human clonal haematopoiesis, we modelled the *DNMT3A* R882 mutation in human HSPCs by transducing purified CD34⁺ HSPCs from cord blood with lentiviral vectors expressing a *DNMT3A* shRNA and BFP. Since *DNMT3A* R882-mutated cells retain about 20% of wild-type *DNMT3A* activity¹³, knockdown of *DNMT3A* expression should phenocopy the effects of *DNMT3A* R882 mutations. Using this approach and an *in vitro* competition assay with human cytokines (Fig. 5a and Supplementary Fig. 1c), we found that *DNMT3A*-knockdown (*DNMT3A*-KD) HSPCs displayed a competitive advantage over control HSPCs expressing a non-targeting shRNA

(shNT) (Fig. 5b and Extended Data Fig. 8a,b). Consistent with our mouse studies, metformin treatment reduced the competitive advantage of *DNMT3A*-KD HSPCs (Fig. 5b). The effects of *DNMT3A* knockdown and metformin treatment were observed in both the CD34⁺ and CD34⁻ cell fractions following ex vivo culture (Extended Data Fig. 8c). The suppressive effect of metformin was mitigated by the co-expression of NDI1 (Fig. 5b), indicating that the effect was mediated through inhibition of complex I. *DNMT3A*-KD HSPCs exhibited higher basal OCRs compared with control HSPCs, a difference that was eliminated by metformin treatment (Fig. 5c). Similar to mouse *Dnmt3a^{R878H}* HSPCs, the levels of 5mC and H3K27me3 were lower in *DNMT3A*-KD cells than in control cells (Fig. 5d,e). Metformin treatment augmented the methylation index (Fig. 5f and Extended Data Fig. 8d) and increased 5mC and H3K27me3 in *DNMT3A*-KD cells (Fig. 5d,e), mirroring the mechanistic findings observed in mouse *Dnmt3a^{R878H}* HSPCs.

To further evaluate the effect of metformin on *DNMT3A* R882-mutated human HSPCs, we designed and optimized a prime editing strategy to introduce the R882H mutation into the *DNMT3A* gene in purified CD34⁺ HSPCs from cord blood samples. The Cas nickase (nCas)-based prime editing technique has been shown to cause less cytotoxic and genotoxic stress and edit with higher precision and efficiency in long-term repopulating HSPCs than conventional homology-directed repair-based CRISPR–Cas9 editing strategies²⁶. Using the optimized prime editing strategy, we introduced the *DNMT3A*^{R882H} mutation in five HSPC samples from independent donors. As a negative control, we introduced a T>G single nucleotide variant in exon 1 of the *B2M* gene that causes a premature stop codon²⁶. After prime editing, the baseline (day 0) mean variant allele frequency (VAF) was 20% for *DNMT3A*^{R882H} (Fig. 5g and Extended Data Fig. 8e) and 49.8% for the *B2M* single nucleotide variant (Fig. 5h). The edited cell pools were plated in methylcellulose medium to assess the relative competitive advantage of the *DNMT3A*^{R882H} versus *DNMT3A* wild-type cells. After an additional 14 days in culture, the mean VAF of *DNMT3A*^{R882H} increased to 42.4% in untreated cells, consistent with a relative expansion of the mutant population (Fig. 5g). Metformin treatment suppressed the expansion of *DNMT3A*-mutated cells with a mean VAF of 26.2% (Fig. 5g). By contrast, the mean VAFs of *B2M*-edited cells did not significantly change after 14 days in culture with or without metformin treatment (Fig. 5h), indicating that the observed effects on *DNMT3A*^{R882H} HSPCs were not an artefact of prime editing. Together, these results support the idea that metformin has the potential to suppress the competitive advantage of *DNMT3A* R882-mutated clones in human clonal haematopoiesis.

Discussion

Targeting the cell-intrinsic mechanisms that are critical for the selective advantage of mutant HSPCs in clonal haematopoiesis is a potential strategy for suppressing clonal expansion and lowering the risk of developing disease related to clonal haematopoiesis. Here we found that upregulation of mitochondrial respiration is a key functional consequence of the *Dnmt3a* R878H mutation and mutant HSPCs are dependent on this metabolic reprogramming to outcompete their wild-type counterparts. Notably, this dependency was evident at the level of HSCs. Thus, our findings provide evidence that mitochondrial metabolism is a critical cell-intrinsic regulator of the clonal advantage of *DNMT3A* R882-mutated clones in clonal haematopoiesis. This notion is consistent with the growing body of evidence that demonstrates a role for mitochondrial bioenergetics and dynamics in the regulation of stem cell fate.

Our finding that *Dnmt3a^{R878H/+}* HSPCs are dependent on increased mitochondrial respiration has important therapeutic implications because many components of the ETC are druggable cellular targets. In this study, we focused on the therapeutic potential of metformin, a biguanide that is widely used in the treatment of diabetes. Although biguanides have been reported to target many cellular proteins, their inhibitory effect on complex I (NADH dehydrogenase) activity is the

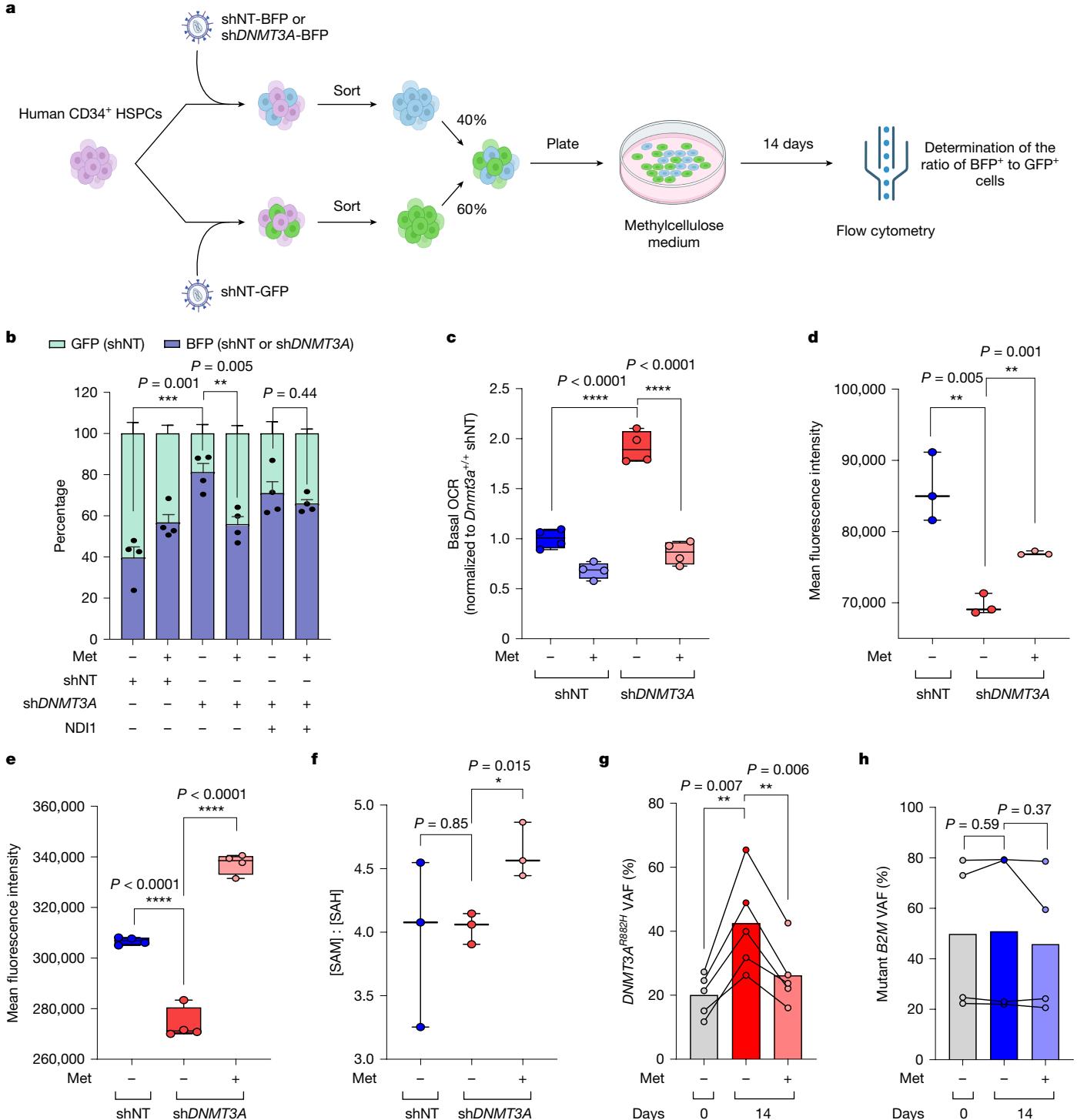


Fig. 5 | Metformin suppresses human DNMT3A^{R882H} HSPCs. **a**, Schematic diagram of the in vitro competition assay of human CD34⁺-enriched HSPCs. Created in BioRender. Chan, S. (2025) <https://BioRender.com/w95m335>.

b, Proportion of BFP⁺ and GFP⁺ cells in a competition assay between BFP⁺ HSPCs expressing shNT, shRNA targeting DNMT3A (shDNMT3A) alone or shDNMT3A plus NDI1 and GFP⁺ HSPCs expressing shNT in the absence or presence of metformin. $n = 4$ biologically independent samples each. **c**, Basal OCR in human HSPCs expressing shNT or shDNMT3A and treated with or without metformin. $n = 4$ biologically independent samples each. **d**, Intracellular 5mC staining in human HSPCs expressing the indicated shRNA and treated with or without metformin. $n = 3$ biologically independent samples each. **e**, Intracellular H3K27me3 staining in human HSPCs expressing the indicated shRNA and treated with or without metformin. $n = 4$ biologically independent samples

each. **f**, SAM:SAH ratio in human HSPCs expressing the indicated shRNA and treated with or without metformin. $n = 3$ biologically independent samples each. **g**, DNMT3A^{R882H} VAFs of prime-edited human HSPCs at baseline (day 0) and after 14 days in culture in the presence or absence of metformin. $n = 5$ biologically independent samples. **h**, Mutant B2M VAFs of prime-edited human HSPCs at baseline (day 0) and after 14 days in culture in the presence or absence of metformin. $n = 4$ biologically independent samples. Lines connect VAFs from the same sample in **g,h**. In **c-f**, the box represents the interquartile range with the median indicated by the line inside the box; whiskers extend to the minimum and maximum values. In **b**, data are mean \pm s.e.m. In **g,h**, the bar plot represents the mean. Statistical significance was calculated using two-sided unpaired (**b-f**) or paired (**g,h**) Student's *t*-tests.

most well established and supported by structural evidence⁸. Indeed, our finding that ectopic expression of NDI1, a metformin-resistant yeast analogue of complex I, rendered *Dnmt3a*^{R878H/+} HSPCs insensitive to effects of metformin strongly supports complex I as the main protein target. However, the observed reduction in mitochondrial respiration in metformin-treated *Dnmt3a*^{R878H/+} HSPCs was not due to complex I inhibition alone but also occurred through the downstream downregulation of genes involved in OXPHOS. Results from our multi-omics studies suggest that metformin exerts its downstream effects on gene expression, at least in part, by increasing the methylation potential and consequently augmenting the activity of DNMT3A, the PRC2 complex and possibly other SAM-dependent methyltransferases in *Dnmt3a*^{R878H/+} HSPCs. This proposed mechanism is consistent with prior studies that demonstrate an association between metformin exposure and an increase in 5mC and H3K27me3 levels in various cellular contexts^{27–29}. It is noteworthy that metformin appears to preferentially increase the expression of genes involved in 1C metabolism and methylation index in *Dnmt3a*^{R878H/+} HSPCs over wild-type cells, indicating a degree of selectivity in its effects. Whether this selectivity is specific for metformin or common across other ETC inhibitors is unclear and warrants further investigations.

To explore the translational relevance of our findings, we utilized two distinct approaches to model the *DNMT3A* R882 mutation in human cells. In the first approach, we down-regulated *DNMT3A* expression using RNA interference to mimic the functional effect of the *DNMT3A* R882 mutation, which has been shown to reduce DNMT3A enzymatic activity to around 20% that of the wild type¹³. Metformin treatment effectively suppressed the competitive advantage of these *DNMT3A*-KD HSPCs, aligning with our proposed mechanistic model in which metformin acts by augmenting the residual DNMT3A activity in the cell. In the second approach, we optimized a prime editing strategy to introduce the *DNMT3A*^{R882H} mutation into human HSPCs with high editing efficiencies. The prime editing technique has important advantages over Cas9 nuclease-based genome editing strategies that depend on the generation of DNA double-strand breaks, which are highly toxic to HSCs. Although prime editing can still induce a small amount of double-strand breaks, it is less genotoxic and can achieve high editing efficiencies in long-term repopulating HSPCs²⁶. Our reported methodology represents an important technical resource for the study of *DNMT3A* R882 mutations in human HSPCs.

The presence of clonal haematopoiesis has been shown to be associated with an increased risk of developing not only haematologic malignancies but also a growing list of age-related inflammatory illnesses. Interventions that effectively lower the risk of these adverse outcomes in clonal haematopoiesis carriers have the potential to positively affect the health of a large segment of the aging population. However, this goal is not yet possible owing to the lack of known interventions that effectively suppress the expansion of mutant clones in clonal haematopoiesis. The ideal preventive intervention should be not only effective, but also easy to administer and safe for long-term use. Metformin fulfills these criteria and can be readily repurposed as a preventive treatment for *DNMT3A* R882-mutated clonal haematopoiesis carriers, especially those at high risk of malignant transformation or other clonal haematopoiesis-related illness. A recent retrospective case-control study bolsters this concept, revealing that metformin use correlated with a decreased risk of myeloproliferative neoplasm development³⁰. Alternative interventions that augment the availability of 1C metabolites for SAM production, such as folic acid, may also be effective in suppressing the competitive advantage of *DNMT3A* R882-mutated clones. Our research provides the preclinical rationale for investigating these approaches in future clinical trials.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-025-08871-w>.

- Jaiswal, S. et al. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N. Engl. J. Med.* **377**, 111–121 (2017).
- Jaiswal, S. et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
- Genovese, G. et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
- Weeks, L. D. & Ebert, B. L. Causes and consequences of clonal hematopoiesis. *Blood* **142**, 2235–2246 (2023).
- Bick, A. G. et al. Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* **586**, 763–768 (2020).
- Buscarlet, M. et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood* **130**, 753–762 (2017).
- Ley, T. J. et al. *DNMT3A* mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363**, 2424–2433 (2010).
- Bridges, H. R. et al. Structural basis of mammalian respiratory complex I inhibition by medicinal biguanides. *Science* **379**, 351–357 (2023).
- Venugopal, K., Feng, Y., Shabashvili, D. & Guryanova, O. A. Alterations to DNMT3A in hematologic malignancies. *Cancer Res.* **81**, 254–263 (2021).
- Young, A. L., Tong, R. S., Birnbaum, B. M. & Druley, T. E. Clonal hematopoiesis and risk of acute myeloid leukemia. *Haematologica* **104**, 2410–2417 (2019).
- Jawad, M. et al. DNMT3A R882 mutations confer unique clinicopathologic features in MDS including a high risk of AML transformation. *Front. Oncol.* **12**, 849376 (2022).
- Larsson, C. A., Cote, G. & Quintas-Cardama, A. The changing mutational landscape of acute myeloid leukemia and myelodysplastic syndrome. *Mol. Cancer Res.* **11**, 815–827 (2013).
- Russell-Germain, D. A. et al. The R882H *DNMT3A* mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer Cell* **25**, 442–454 (2014).
- Kim, S. J. et al. A *DNMT3A* mutation common in AML exhibits dominant-negative effects in murine ES cells. *Blood* **122**, 4086–4089 (2013).
- Smith, A. M. et al. Functional and epigenetic phenotypes of humans and mice with *DNMT3A* overgrowth syndrome. *Nat. Commun.* **12**, 4549 (2021).
- Loberg, M. A. et al. Sequentially inducible mouse models reveal that Npm1 mutation causes malignant transformation of *Dnmt3a*-mutant clonal hematopoiesis. *Leukemia* **33**, 1635–1649 (2019).
- Nam, A. S. et al. 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 (2022).
- LaMoia, T. E. & Shulman, G. I. Cellular and molecular mechanisms of metformin action. *Endocr. Rev.* **42**, 77–96 (2021).
- Seo, B. B. et al. Molecular remedy of complex I defects: rotenone-insensitive internal NADH-quinone oxidoreductase of *Saccharomyces cerevisiae* mitochondria restores the NADH oxidase activity of complex I-deficient mammalian cells. *Proc. Natl. Acad. Sci. USA* **95**, 9167–9171 (1998).
- Wheaton, W. W. et al. Metformin inhibits mitochondrial complex I of cancer cells to reduce tumorigenesis. *eLife* **3**, e02242 (2014).
- Dowling, R. J. et al. Metformin pharmacokinetics in mouse tumors: implications for human therapy. *Cell Metab.* **23**, 567–568 (2016).
- Izzo, F. et al. DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nat. Genet.* **52**, 378–387 (2020).
- Ducker, G. S. et al. Human SHMT inhibitors reveal defective glycine import as a targetable metabolic vulnerability of diffuse large B-cell lymphoma. *Proc. Natl. Acad. Sci. USA* **114**, 11404–11409 (2017).
- Kontestatis, Z. et al. Discovery of AG-270, a first-in-class oral MAT2A inhibitor for the treatment of tumors with homozygous MTAP deletion. *J. Med. Chem.* **64**, 4430–4449 (2021).
- Dai, Y. J. et al. Conditional knockin of *Dnmt3a* R878H initiates acute myeloid leukemia with mTOR pathway involvement. *Proc. Natl. Acad. Sci. USA* **114**, 5237–5242 (2017).
- Fiumara, M. et al. Genotoxic effects of base and prime editing in human hematopoietic stem cells. *Nat. Biotechnol.* **42**, 877–891 (2024).
- Cuyas, E. et al. Metformin regulates global DNA methylation via mitochondrial one-carbon metabolism. *Oncogene* **37**, 963–970 (2018).
- Cuyas, E. et al. Metformin directly targets the H3K27me3 demethylase KDM6A/UTX. *Aging Cell* **17**, e12772 (2018).
- Garcia-Calzon, S. et al. DNA methylation partially mediates antidiabetic effects of metformin on HbA1c levels in individuals with type 2 diabetes. *Diabetes Res. Clin. Pract.* **202**, 110807 (2023).
- Kristensen, D. T. et al. Metformin use and risk of myeloproliferative neoplasms: a Danish population-based case-control study. *Blood Adv.* **8**, 4478–4485 (2024).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025

Methods

Human HSPC isolation

Cord blood samples were obtained with informed consent from Trillium Health, Credit Valley and William Osler Hospitals according to procedures approved by the University Health Network (UHN) Research Ethics Board. Mononuclear cells were isolated by centrifugation over Ficoll-Paque. HSPCs were enriched using the Human Progenitor Cell Enrichment Kit with Platelet Depletion kit according to manufacturer's instructions (Stemcell Technologies, 19356). For prime-editing experiments, CD34⁺ cells in the enriched population were further purified using the human CD34 MicroBead Ultrapure kit according to manufacturer's instructions (Miltenyi Biotec 130-100-453). The purified cells were used immediately for experiments or viably frozen in 90% (v/v) FBS with 10% (v/v) DMSO and stored in liquid nitrogen.

mRNA in vitro transcription

The PEmax plasmid (Addgene, #204472) was used to synthesize the mRNA encoding nCas9-RT by in vitro transcription as described²⁶. In brief, the plasmid was linearized with SpeI and purified by phenol-chloroform extraction. mRNAs were transcribed in vitro using the MEGAscript T7 Transcription kit (ThermoFisher), capped with 8 mM CleanCap Reagent AG (TriLink), purified using the RNeasy Plus Mini Kit (Qiagen). The quality of the transcribed mRNA was assessed by capillary electrophoresis. mRNAs were resuspended in RNase free water and stored at -80 °C.

Prime editing of human HSPCs

A total of 1 × 10⁵ to 5 × 10⁵ CD34⁺ HSPCs were rinsed with PBS and subjected to electroporation using the P3 Primary Cell 4D-Nucleofector X Kit and Nucleofector 4D device (Lonza) with program EO-100. The electroporation mixture contained 180 pmol nicking single guide RNA (sgRNA) from Synthego, 270 pmol prime editing guide RNA (pegRNA) from IDT, and 12 µg PEmax mRNA. Following electroporation, cells were allowed to recover for 3 min at room temperature and subsequently maintained in culture. Three to four days after electroporation, cells were harvested to obtain genomic DNA for molecular studies. The nicking sgRNA and pegRNA sequences can be found in Supplementary Table 2. The pegRNA for *B2M* editing was described by Fiumara et al.²⁶. The engineered pegRNA (epegRNA) for *DNMT3A* editing consisting of a protective linker and motif at the 3' end for guide degradation was designed with pegFinder (<http://pegfinder.sidichenlab.org/>) and pegLIT (<https://peglit.liugroup.us/>)³¹.

Human competition assay of prime-edited HSPCs

Three days after nucleofection, edited cells were cultured at a cell density of 1,000 cells per ml in methylcellulose-based medium (MethoCult H4330, Stemcell Technologies) supplemented with BIT9500 serum substitute at 10% (v/v) and StemSpan CC100 at 1×. Cells were cultured in a 5% CO₂ humidified atmosphere at 37 °C. After 2 weeks in culture with or without metformin at 50 µM, cells were collected for molecular analysis to determine the VAF of either the *DNMT3A*^{R882H} or *B2M* mutation.

Molecular analysis of prime-edited cells

Genomic DNA (gDNA) was extracted from prime-edited HSPCs using the QIAamp DNA Micro Kit (Qiagen) or QuickExtract DNA Extraction Solution (Lucigen), following manufacturer's instructions. The efficiency of *B2M* editing was assessed using Sanger sequencing and the EditR software (<http://baseeditr.com>). To adapt EditR for assessing *B2M* editing, we used the input sequence TGGCCTAGCTGTGCTCGC and selected the reverse complement orientation option as described²⁶. The efficiency of *DNMT3A*^{R882H} editing was determined using the QX200 Droplet Digital PCR System and a TaqMan-based genotyping assay. The primer and probe sequences can be found in Supplementary

Table 3. The VAF was calculated as the number of fluorescein amidite (FAM)-positive droplets divided by total droplets containing a target.

Lentivirus production and transduction

HEK293T cells obtained from ATCC, short tandem repeat verified, and regularly tested for mycoplasma contamination. The cells were grown in DMEM supplemented with 10% (v/v) FBS and 2 mM Gluta-Plus (Wisent, 609-066 EL). Cells were seeded in 15-cm tissue culture plates at a density of 7 × 10⁶ cells per plate 1 day before transfection. On the day of transfection, cells were co-transfected with lentiviral plasmids, psPAX2, and pMD2.G using the jetPRIME transfection reagent (Polyplus) according to manufacturer's protocol. Supernatant containing viral particles was collected at 48 and 72 h post-transfection and filtered through a 0.45-µm PVDF filter. Viral particles were precipitated in 40% (w/v) polyethylene glycol overnight. The next day, viral particles were collected by centrifugation at 3,700 rpm for 30 min at 4 °C. The pellet was resuspended in HBSS with 25 mM HEPES and stored at -80 °C for long-term storage.

For lentiviral transductions, non-tissue culture-treated 24-well plates were coated with 20 µg ml⁻¹ of Retronectin (Takara, T100B) for 2 h at room temperature, followed by aspiration, and blocking with PBS containing 2% (w/v) BSA for 30 min at room temperature. After aspiration of the blocking buffer, the concentrated virus suspension was added to wells. The plates were centrifuged at 3,700 rpm for 2 h at 4 °C to promote virus binding. Following centrifugation, unbound virus was aspirated, and cells were added. The plates were then transferred to a 37 °C incubator to initiate lentiviral infection.

Human competition assay of *DNMT3A*-KD HSPCs

Freshly isolated human HSPCs were transduced with lentiviral vectors using the procedures described above. After transduction, cells were cultured in StemSpan SFEM II medium (Stemcell Technologies, 09655) supplemented with L-glutamine at 4 mM, human SCF at 100 ng ml⁻¹, human FLT3-L at 100 ng ml⁻¹, human TPO at 20 ng ml⁻¹, SR1 at 1 µM and UM171 at 50 nM. Three days after transduction, the transduced CD34⁺ cells marked by BFP or GFP expression were sorted by fluorescence-activated cell sorting (FACS). The sorted cells were mixed with appropriate competitor cells and cultured at a density of 2,000 cells per ml in methylcellulose-based medium (MethoCult H4330, Stemcell Technologies) supplemented with BIT9500 serum substitute at 10% (v/v) and StemSpan CC100 at 1×. Cells were cultured in a 5% CO₂ humidified atmosphere at 37 °C. After 14 days in culture, cells were collected and analysed by flow cytometry for the proportion of BFP⁺ and GFP⁺ cells.

Generation of *Dnmt3a*^{-/-} cells using CRISPR-Cas9

Pooled synthetic sgRNAs targeting *Dnmt3a* were ordered from Synthego. The sequences of the three sgRNAs can be found in Supplementary Table 4. The sgRNAs were reconstituted in TE buffer to achieve a concentration of 0.1 mM. 2.5 µl of the sgRNAs was mixed with 1.5 µl recombinant SpCas9 (IDT, 1081058) at a concentration of 10 µg µl⁻¹ in Lonza P3 nucleofection solution (Lonza, V4XP-3024) to achieve a final volume of 25 µl. The mixture was incubated at room temperature for 15 min to allow formation of the sgRNA–Cas9 ribonucleoprotein (RNP) complex. A total of 2 × 10⁵ *Dnmt3a*^{R878H/+} or *Dnmt3a*^{+/+} LK cells were resuspended in 75 µl of the P3 nucleofection solution, combined with the 25 µl of the sgRNA–Cas9 RNP complex, and transferred to a Lonza nucleofection microcuvette. Nucleofection was conducted using an Amaxa 4D-Nucleofector with program CA137. The efficiency of *Dnmt3a* knockout was assessed by Sanger sequencing and ICE analysis four days after nucleofection. The primers used are listed in Supplementary Table 3.

Mouse models and in vivo repopulation experiments

All animal experiments were performed in accordance with institutional guidelines approved by the University Health Network Animal Care

Article

Committee. CD45.2⁺ C57BL/6J and CD45.1⁺ B6.SJL-Ptprc^aPepc^b/Boy (Pep-Boy) mice were obtained from The Jackson Laboratory and maintained in the same animal facility for the duration of the study. Mice were maintained on a 12-hour light/dark cycle at a temperature of 21–22 °C and relative humidity of 45–60%. *Dnmt3a*^{R878H/+} mice (Jackson Laboratory, 032289) were crossbred with Mx-Cre mice (Jackson Laboratory, 003556) and genotyped using protocols described by the Jackson Laboratory. Female *Dnmt3a*^{R878H/+};Mx1^{cre/+} and *Dnmt3a*^{+/+};Mx1^{cre/+} mice were used in our studies. Experiments were initiated at 8–12 weeks of age. To induce Cre recombinase expression, mice were administered polyinosinic-polycytidyllic acid (pIpc) at a dose of 15 mg kg⁻¹ via intraperitoneal injection once every other day for a total of 5 doses. After pIpc administration, genomic DNA was extracted from peripheral blood cells to confirm recombination using a PCR-based assay. In addition, cDNA synthesized from total RNA extracted from WBM cells was sequenced to verify mutant allele expression. Primers used for genotyping and detection of recombination can be found in Supplementary Table 3. In all experiments involving competition with CD45.1⁺ cells, the donor Pep-Boy mice were also treated with pIpc as above prior to cell harvesting to control for any potential effects of pIpc.

For non-competitive repopulation experiments, CD45.2⁺ *Dnmt3a*^{R878H/+} or *Dnmt3a*^{+/+} bone marrow cells were resuspended in Opti-MEM medium and transplanted by tail vein injection (1 × 10⁶ cells per mouse) into 10-week-old female CD45.2⁺ C57BL/6J recipient mice conditioned with 12 Gy of irradiation. For competitive repopulation experiments, CD45.2⁺ *Dnmt3a*^{R878H/+} or *Dnmt3a*^{+/+} bone marrow cells were mixed with CD45.1⁺ *Dnmt3a*^{+/+} bone marrow cells from sex- and age-matched Pep-Boy donors at a 2:3 ratio prior to transplantation. Five weeks after transplantation, mice were block-randomized to receive either metformin treatment (5 mg ml⁻¹ in drinking water) or no treatment. Metformin in the drinking water was replaced twice a week during the treatment period. The investigators were not blinded to the treatment groups during the experiment.

Isolation of mouse HSPCs from bone marrow

Mouse HSPCs were isolated from bone marrow using a two-step enrichment process. First, lineage-negative (Lin⁻) cells were negatively selected using the EasySep Mouse Hematopoietic Progenitor Isolation Kit (Stemcell Technologies, 19856). These cells were further enriched for KIT expression using the EasySep Mouse CD117 (c-KIT) Positive Selection Kit (Stemcell Technologies, 18757) to obtain LK HSPCs.

Mouse in vitro competition assays

In vitro competition experiments were performed in 96-well flat-bottom tissue culture plates. CD45.2⁺ *Dnmt3a*^{R878H/+} or *Dnmt3a*^{+/+} LK cells were mixed with CD45.1⁺ *Dnmt3a*^{+/+} LK cells from Pep-Boy mice at a ratio of 2:3. The cell mixture was plated in MethoCult GF M3434 medium (Stemcell Technologies, 03434) at a density of 200 cells per well, with treatments added as indicated. Cells were cultured in a 5% CO₂ humidified atmosphere at 37 °C for 10 days.

CFU assays

Mouse LK cells (3 × 10³) were suspended in 1.1 ml of MethoCult GF M3434 medium and plated in six-well tissue culture plates with or without metformin. Cultures were maintained at 37 °C in a humidified atmosphere with 5% CO₂ for 10 days. Colonies were counted, and cells were re-plated to assess serial colony-forming capacity.

Flow cytometry

Cells were stained with fluorochrome-conjugated antibodies (Supplementary Table 5) at the indicated dilutions in FACS buffer (HBSS containing 2% FBS and 0.1% sodium azide) for 30 min at 4 °C. After washing in FACS buffer, cells were analysed using a Beckman Coulter CytoFLEX flow cytometer. Data analysis was performed using FlowJo V10 software.

Intracellular protein staining

Cells were first stained with fluorochrome-conjugated antibodies against cell surface proteins for 30 min at 4 °C. Cells were then fixed and permeabilized using a fixation/permeabilization solution (BD Biosciences, 554714) for 20 min at 4 °C. Fixed cells were then incubated with antibodies targeting intracellular proteins (Ki-67, H3K27me3, or DNMT3A) or isotype control antibodies for 1 h at room temperature in 1×Perm/Wash buffer (BD Biosciences, 554714). For unconjugated primary antibodies, cells were stained appropriate fluorophore-conjugated secondary antibodies. After staining, cells were washed twice with 1×Perm/Wash buffer and resuspended in FACS buffer for flow cytometric analysis.

Analysis of mitochondrial parameters

Mitochondrial ROS, ΔΨ_m, and mass were measured in freshly isolated mouse bone marrow cells or enriched LK cells by flow cytometry. Cells were incubated in FACS buffer containing MitoSOX Red (5 μM), TMRE (100 nM), or MitoTracker Green (100 nM) for 20 min at 37 °C. Reagents are listed in Supplementary Table 6. After staining, cells were washed twice and resuspended in FACS buffer with a live/dead stain before flow cytometric analysis.

Mitochondrial function analysis

Cellular bioenergetics were assessed using a Seahorse XFe96 Extracellular Flux Analyzer. One day before the assay, the sensor cartridge was hydrated in calibration buffer at 37 °C in a non-CO₂ incubator. XFe96 microplate wells were coated with Cell-Tak (Corning, 354240) at a concentration of 22.4 μg ml⁻¹ at 4 °C and rinsed with sterile water. Cells were resuspended in XF base minimal DMEM media containing 11 mM glucose, 2 mM glutamine, and 1 mM sodium pyruvate. Approximately 3 × 10⁵ cells in 180 μl of medium were added per well in the microplate. The plates were centrifuged at 100g for 5 min to promote cell adhesion and incubated for 1 h at 37 °C in a non-CO₂ incubator. OCR and extracellular acidification rate were measured at baseline and following sequential addition of oligomycin (1 μM), FCCP (1 μM), and antimycin A and rotenone (1 μM each) using the XFe96 Analyzer.

RNA extraction and RT-qPCR analysis

Total RNA was isolated using the RNeasy Plus Kit (Qiagen) and quantified using a Nanodrop spectrophotometer. RT-qPCR analysis was performed using the Luna Universal One-Step RT-qPCR kit (NEB, E3005S) on a Bio-Rad CFX Touch Real-Time PCR Detection System. Gene expression was normalized to mouse *Actb* or human *ACTB* expression. The primer sequences used for RT-qPCR analysis can be found in Supplementary Table 3.

Plasmid vector construction

Synthetic oligonucleotides were annealed and ligated into the BbsI site of the shRNA lentiviral expression vector pRSI9-U6-(sh)-UbiC-TagRFP-2A-Puro (Addgene, 28289). The TagRFP sequence was replaced with the BFP sequence between the XbaI and BamHI sites. The NDII coding sequence was amplified by PCR from the PMXS-NDII plasmid (Addgene, 72876) and cloned into the pLVX-EF1α-IRES-ZsGreen1 vector (Clontech, 631982) between the EcoRI and SpeI sites. Oligonucleotide and primer sequences are listed in Supplementary Table 3.

Bulk RNA-seq analysis

Total RNA was isolated from enriched mouse LK HSPCs using the RNeasy Plus Mini Kit (QIAGEN, 74136), following manufacturer's instructions. Library preparation was performed using the NEBNext Ultra RNA Library Prep Kit for Illumina (Novogene Corporation). Libraries were sequenced on a NovaSeq 6000 S4 platform using paired-end 150 bp reads. Reads were aligned to the mm10 reference genome using STAR (v2.5), and gene-level quantification was performed using HTSeq v0.6.1.

CITE-seq analysis

Single-cell RNA and protein expression were simultaneously measured using CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing). Enriched mouse LK HSPCs were labelled with anti-CD45.2 and anti-CD45.1 TotalSeq-B antibodies (Supplementary Table 5) to distinguish between *Dnmt3a*^{R878H/+} cells and *Dnmt3a*^{+/+} cells, respectively. CITE-seq libraries were generated using the 10x Genomics Chromium Single Cell 3' v3.1_CellSurfaceProtein_RevD kit and sequenced on a Novaseq 6000 system (20,000 cells per sample). RNA reads were aligned to mm10 reference using Cell Ranger v6.1.2. The filtered Cell Ranger output was then analysed using Seurat package v4 with the following quality control criteria: 500–8,000 genes per cell and <15% mitochondrial genes. Fast integration using reciprocal principal components analysis (RPCA) was used to find anchors across datasets to integrate the four samples. Data normalization, variance stabilization and selection of the top 3,000 variable features of the molecular count data were performed using SCTransform, followed by dimension reduction by PCA and UMAP embedding using the top 30 principal components.

Haematopoietic subtypes were assigned to each cell using the AddModuleScore function with previously defined mouse haematopoietic population gene sets²². The highest gene set enrichment score determined the cell type annotation for each single cell. As an additional filtering step, the maturing erythroblastic cells expressing low levels of *Kit* (CD117) and *Ptprc* (CD45) were removed from the analysis. scDblFinder 1.16.0 was run on the Cell Ranger raw output of each individual sample to identify potential cell doublets. A total of 46,225 cells were kept for downstream analyses.

The CD45.1 and CD45.2 sequencing antibody-derived tags (ADTs) were log normalized. ADTs with a normalized value greater than 6 were identified as outlier points and removed from the analysis. Cells were classified as wild-type (CD45.1 > CD45.2) or mutant (CD45.2 > CD45.1) based on normalized ADT ratios. To perform differential gene expression analysis, RNA counts from each defined haematopoietic cell population labelled as mutant and wild type were transformed in a single-cell experiment object and aggregated for each of the metformin or vehicle samples using the sum of the counts in the R package scuttle (v1.0.4).

RRBS analysis

Genomic DNA from LK-enriched HSPCs was analysed by RRBS (Novogene). In brief, DNA was digested with MsP1, end-repaired, A-tailed, and ligated with methylated adapters. Size-selected fragments were bisulfite-converted using the EZ DNA Methylation-Gold Kit (Zymo Research) and amplified by PCR to enrich for adapter-ligated fragments. RRBS libraries were quality-checked and sequenced on an Illumina HiSeq/NovaSeq platform, generating paired-end reads of 150 bp nucleotides.

Raw data were trimmed to remove adapter and low-quality bases using Trimmomatic-0.36, followed by quality control assessment with FastQC (v0.11.5). Trimmed reads were aligned to mouse reference genome from Ensembl (GRCm38/mm10) and duplicated reads were removed. DNA methylation calls were extracted from the aligned reads as CpG coverage files. DMRs were identified using the open-source R package methylKit (v1.26.0)^{32,33}. CpG sites on unmapped genome assembly contigs were removed, and remaining CpG sites were filtered to exclude CpGs with <10× coverage PCA analysis in R. We used methylKit to perform pairwise comparisons to identify DMRs between untreated *Dnmt3a*^{R878H/+} versus untreated *Dnmt3a*^{+/+} samples, and between metformin-treated *Dnmt3a*^{R878H/+} and untreated *Dnmt3a*^{R878H/+} samples. To this end, the genome was tiled into 500-bp non-overlapping bins. To calculate DMR *P* values, a logistic regression test was used methylKit. *P* values were adjusted for multiple testing (that is, *q*-value) via the SLIM method³⁴. DMRs with *P* value < 0.01 were used for downstream analysis. CpG islands were annotated by using the University of California Santa Cruz (<https://genome.ucsc.edu/index.html>) database

with using plyranges R package (v1.20.0)³⁵. Promoters were defined as 1 kb upstream and 150 bp downstream around the TSS and annotated by ChIPseeker R package (v1.36.0)^{36,37}.

Gene set enrichment analysis

Differential expression analysis was performed using DESeq2 v1.40.1 for bulk RNA-seq and edgeR v3.36.0 for scRNA-seq data. Genes were ranked from the top up-regulated to down-regulated using the formula: rank score = $-\log_{10}(P \text{ value}) \times \text{sign}(\log \text{ fold change})$. For methylation analysis, RRBS data were processed in 10-kb segments with 25% overlap. Methylation levels were calculated for each tile. Promoter regions were defined as ± 3 kb from TSSs. Gene promoters were ranked from hypermethylated to hypomethylated using the same ranking formula. GSEA was performed using ClusterProfiler v4.2.2 with the fgsea algorithm against selected gene sets from Gene Ontology (GO) Biological Process, KEGG and WikiPathways databases.

H3K27me3 ChIP-seq analysis

LK-enriched bone marrow cells were fixed with 1% formaldehyde for 15 min following the Active Motif chromatin immunoprecipitation protocol. ChIP-seq of the fixed cell pellets was performed by Active Motif using 15 µg chromatin and anti-H3K27me3 antibody (Active Motif, 39155). Sequencing data were processed and demultiplexed using bcl2fastq2 v2.20, with low-quality bases (Phred<33) trimmed. Single-end 75 bp sequence reads were mapped to the genome through BWA v0.7.12 algorithm with default settings.

Low-quality reads were filtered out and PCR duplicates were removed. Aligned sequencing reads, or tags, were extended to 200 bp from the 3' end, followed by dividing the genome into 32 bp bins and counting the number of fragments in each bin. The resulting histograms (genomic 'signal maps') were stored in bigWig files. Peak locations were determined using the MACS algorithm (v2.1.0) with a cutoff of *P* value = 10^{-7} . Peaks that were on the ENCODE blacklist of known false ChIP-seq peaks were removed. A total of 18,556 peaks that were identified in at least one sample with a cutoff *P* value of 10^{-7} were all merged in a common matrix. The total number of present peaks as well as averaged peak values were calculated and plotted for each sample and condition using the R package ggplot2. ChIP-seq profiles were generated ± 2 kb around TSSs of 10,622 unique genes using DeepTools v3.5.1 plotHeatmap function.

scRNA-seq data processing

The 10x scRNA data of CD34⁺ enriched cells from GCSF-mobilized samples (GSE158067; *n* = 4) were processed using Seurat v5.1, following the method and parameters described¹⁷. In brief, each sample was preprocessed individually by removing cells with fewer than 200 RNA features, more than 3 median absolute deviations from the median UMI, and more than 20% mitochondrial genes, before being log-normalized. A total of 2,000 variable features were selected. RPCA was then used to integrate the 4 samples, utilizing 30 principal components and 2,000 anchors. The integrated data was log-normalized, scaled, and centred using sex and mitochondrial percentage as covariates. The top 30 statistically significant principal components were used as inputs to the UMAP algorithm for cluster visualization. Clusters were defined using the Louvain algorithm and annotated based on the maximum average enrichment score obtained from Landau group's list of cell population markers²². Differential expression between the *DNMT3A* R882 mutant and wild-type phenotype, determined via targeted genotyping, was assessed using the Wilcoxon rank sum test with the limma method. Genes were ranked based on the $-\log_{10} P$ value and the sign of the log fold change. The enrichment of hsa00190 (OXPHOS) was evaluated using the GSEA method in ClusterProfiler 4.8.

Single-cell DNA methylation data and scRNA-seq data processing

Single-cell DNA methylation data, combined with scRNA-seq (SmartSeq2) data from CD34⁺ enriched cells in GCSF-mobilized bone marrow

Article

samples ($n = 2$), were retrieved from GSE158067 (GSE158067_gene_exp mtx.txt and supplementary table 5 in ref. 17). Seurat V5.1 was used to process the scRNA data. The counts were log-normalized, scaled, and centred using the number of RNA features, mitochondrial percentage and plates as covariates. Differential expression between the *DNMT3A* R882 mutant and wild-type phenotype, determined through targeted genotyping, was assessed using the Wilcoxon rank sum test via the limma method. Genes were ranked based on the $-\log_{10} P$ value and the sign of the log fold change. To perform GSEA on the DNA methylation data, 12,131 genes were ranked by the DMR value within the ± 1 kb TSS region using the $-\log_{10} P$ value and fold change sign. GSEA was performed using all GO gene sets. The enrichment of GO gene sets was evaluated using the GSEA method in ClusterProfiler 4.8 for both RNA and DNA methylation data.

Metabolomics analysis

LK-enriched mouse HSPCs were washed in PBS, pelleted, and snap-frozen for metabolomic analysis (University of Colorado School of Medicine Metabolomics Core). Metabolites were extracted from cell pellets (2×10^6 cells per ml) using a cold methanol:acetonitrile:water solution (5:3:2). Analysis was performed on a Thermo Vanquish UHPLC coupled to a Thermo Q Exactive mass spectrometer in positive and negative ion modes (separate runs) exactly as described previously^{38,39}. Signals were annotated and integrated using Maven in conjunction with the KEGG database and an in-house standard library as reported⁴⁰.

SAM and SAH quantification

SAM and SAH levels were measured using the SAM and SAH Combo ELISA Kit (Cell Biolabs, STA-671-C). In brief, mouse LK-enriched HSPCs (3×10^6 cells) were sonicated in 1 ml cold PBS on ice. The homogenized samples were centrifuged at 10,000g for 15 min at 4 °C, and the supernatant was collected. Samples were either analysed immediately or stored at -80 °C. SAM and SAH concentrations were determined following the manufacturer's protocol.

Statistical analysis

Data were analysed using GraphPad Prism v10 (GraphPad Software). Sample sizes and the specific statistical tests used are specified in figure legends. Sample size determination for animal experiments was not based on statistical power calculations. Statistical significance was defined as $P < 0.05$. For GSEA, results were considered significant if the nominal P value was <0.05 for single gene set testing and false discovery rate < 0.05 for testing against multiple gene sets.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw and processed data from the bulk RNA-seq, scRNA-seq, RRBS and ChIP-seq experiments are deposited in a publicly accessible Gene Expression Omnibus (GEO) repository under the accession number GSE255101. Source data are provided with this paper.

Code availability

All code used for bioinformatics analysis is publicly available via GitHub at <https://github.com/veroniquevoisin/DNMT3A>.

31. Nelson, J. W. et al. Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**, 402–410 (2022).
32. Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
33. Akalin, A. et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet.* **8**, e1002781 (2012).
34. Wang, H. Q., Tuominen, L. K. & Tsai, C. J. SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. *Bioinformatics* **27**, 225–231 (2011).
35. Lee, S., Cook, D. & Lawrence, M. plyranges: a grammar of genomic data transformation. *Genome Biol.* **20**, 4 (2019).
36. Wang, Q. et al. Exploring epigenomic datasets by ChIPseeker. *Curr. Protoc.* **2**, e585 (2022).
37. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Biocconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
38. Gehrk, S. et al. Red blood cell metabolic responses to torpor and arousal in the hibernator Arctic ground squirrel. *J. Proteome Res.* **18**, 1827–1841 (2019).
39. Nemkov, T., Reisz, J. A., Gehrk, S., Hansen, K. C. & D'Alessandro, A. High-throughput metabolomics: isocratic and gradient mass spectrometry-based methods. *Methods Mol. Biol.* **1978**, 13–26 (2019).
40. Nemkov, T., Hansen, K. C. & D'Alessandro, A. A three-minute method for high-throughput quantitative metabolomics and quantitative tracing experiments of central carbon and nitrogen pathways. *Rapid Commun. Mass Spectrom.* **31**, 663–673 (2017).

Acknowledgements This project was supported by funds from an Invest in Research Award from the Princess Margaret Cancer Foundation, a Medicine by Design (MbD) Award from the University of Toronto, a New Investigator Award from the Leukemia Research Foundation, and a Project Grant (PJT-175186) from the Canadian Institutes of Health Research. J.E.D. is supported by funds from the Princess Margaret Cancer Centre Foundation, Canadian Institutes for Health Research (RN380110–409786), Canadian Cancer Society (grants 703212 (end date 2019) and 706662 (end date 2025)), Terry Fox New Frontiers Program Project Grant (project 1106), a Canada Research Chair, Princess Margaret Cancer Centre, The Princess Margaret Cancer Foundation and Ontario Ministry of Health. The authors thank R. Culp-Hill for metabolomics data acquisition. J.A.R. and A.D. acknowledge support from the University of Colorado Cancer Center Support Grant (P30CA046934).

Author contributions M.H. and S.M.C. conceived the study and designed the experiments. M.H. performed and analysed most experiments. V.V., A.C. and S.P. performed the bioinformatics analyses with assistance from Y.W. and A.C.H.L. F.G., A.D.S., G.D.B. and G.E. provided input and supervised the bioinformatics analyses. S.C., D.M.A., A.C.H.L., Y.Y., V.W., A.M. and E.G. performed experiments. J.A.R. and A.D. performed and analysed the metabolomics experiments. A.V. performed the prime editing experiments under the supervision of S.Z.X. and J.E.D. M.F. designed the prime editing strategy under the supervision of L.N. and S.F. M.H. and S.M.C. interpreted the data and wrote the manuscript. K.Y. and J.J.T. reviewed and edited the manuscript. All authors provided input in the preparation of the final manuscript.

Competing interests S.M.C. has received research funding from the Centre for Oncology and Immunology in Hong Kong, Celgene/BMS, AbbVie Pharmaceuticals, Agios Pharmaceuticals and Servier Laboratories. F.G. serves as a consultant for S2 Genomics Inc. A.D.S. has received research funding from Takeda Pharmaceuticals, BMS and Medivir AB, and consulting fees/honorarium from Takeda, Novartis, Jazz and Otsuka Pharmaceuticals. A.D.S. is named on a patent application for the use of DNT cells to treat AML. A.D.S. is a member of the Medical and Scientific Advisory Board of the Leukemia and Lymphoma Society of Canada. A.D.S. holds the Ronald N. Buick Chair in Oncology Research. J.E.D. has received research funding from Celgene/BMS, and has patents licensed to Trillium Therapeutics/Pfizer. J.J.T. has received research funding from H3 Biomedicine, Inc. and patent royalties from Fate Therapeutics. The other authors declare no competing interests.

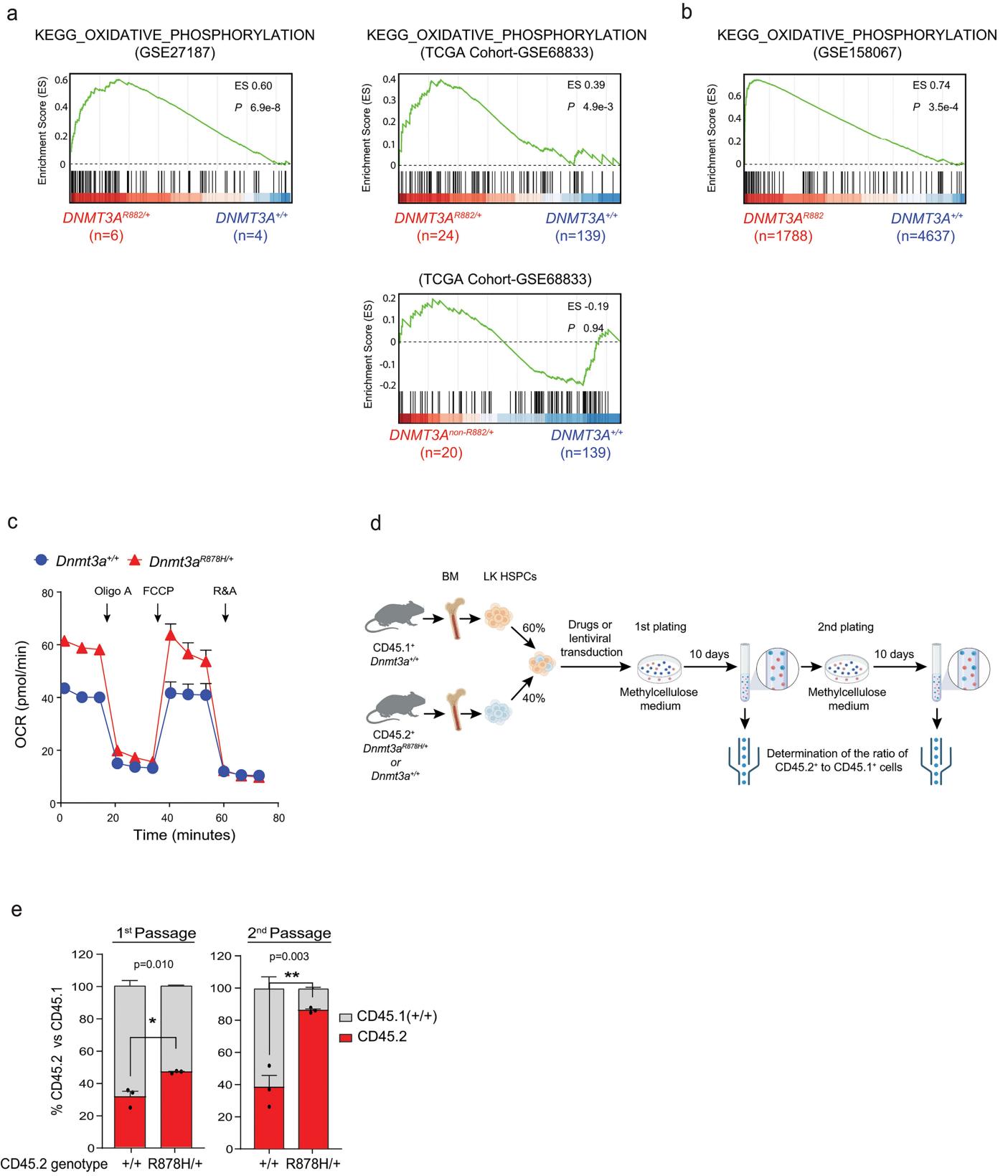
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-025-08871-w>.

Correspondence and requests for materials should be addressed to Steven M. Chan.

Peer review information *Nature* thanks Navdeep Chandel, Erik Eldering and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

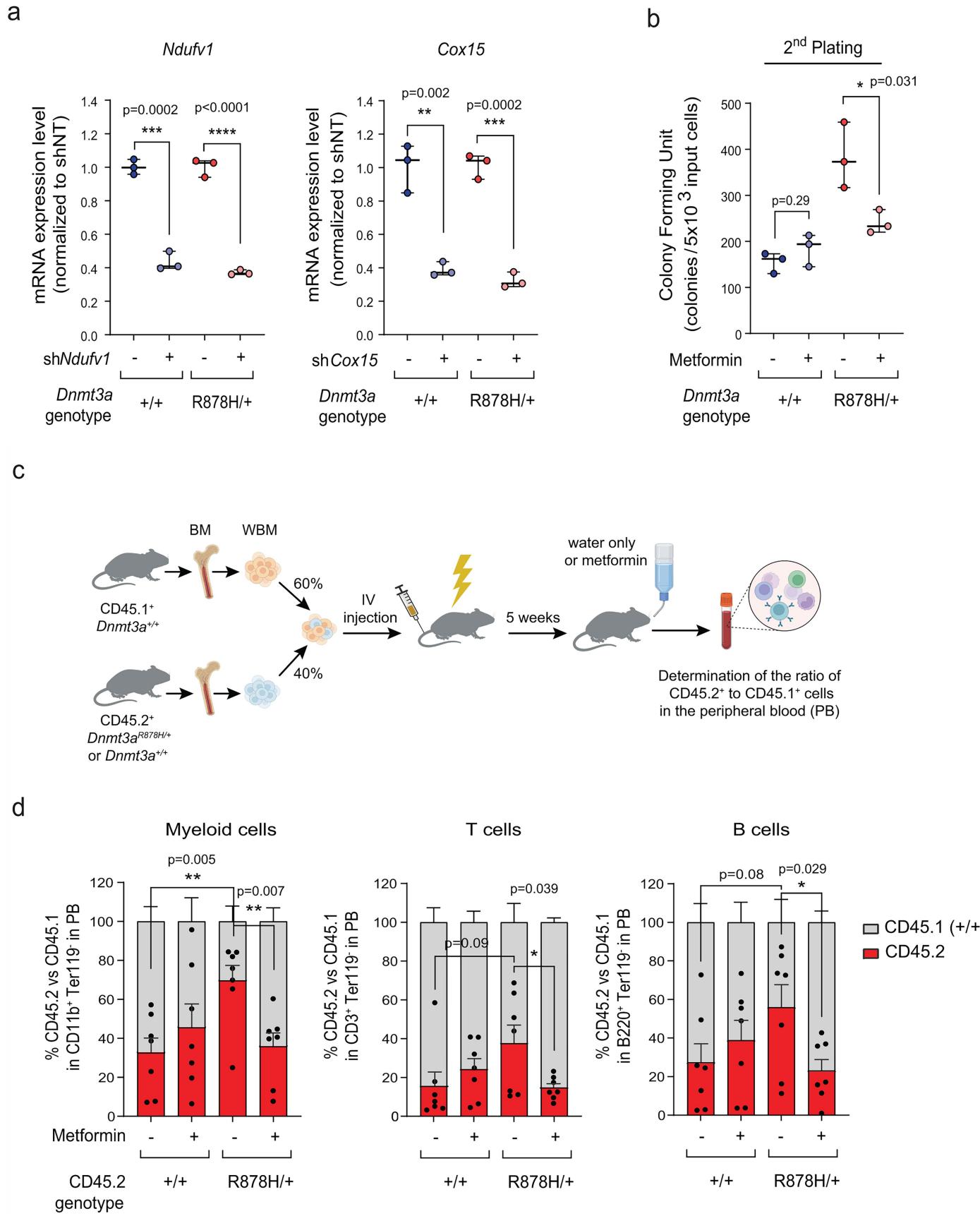


Extended Data Fig. 1 | See next page for caption.

Article

Extended Data Fig. 1 | *Dnmt3a*^{R878H/+} HSPCs have increased mitochondrial respiration compared with *Dnmt3a*^{+/+} cells and are dependent on this metabolic reprogramming for their competitive advantage. **a**, Gene set enrichment plots comparing *DNMT3A*^{RS82} or *DNMT3A*^{non-RS82} mutated AML samples versus *DNMT3A*^{WT} AML samples using two publicly available gene expression datasets (GSE27187 and GSE68833). Number of patients in each cohort is shown at the bottom of the plot. **b**, Gene set enrichment plots comparing *DNMT3A*^{RS82} versus *DNMT3A*^{WT} CD34⁺ HSPCs from individuals with CH using the publicly available GSE158067 gene expression dataset and analyzed as pseudobulk populations. Number of cells in each pseudobulk population is shown at the bottom of the plot. **c**, OCRs of whole bone marrow

cells of the indicated genotype at baseline and at different time points following treatment with oligomycin A (Oligo A), FCCP, and rotenone plus antimycin A (R&A). n = 3 biologically independent samples for each condition. **d**, Schematic diagram showing the design of the in vitro competition assay. Created in BioRender. Chan, S. (2025) <https://BioRender.com/ks3i0wn>. **e**, Proportion of CD45.2⁺ and CD45.1⁺ cells in a competition assay between CD45.2⁺ LK cells of the indicated genotype and CD45.1⁺ *Dnmt3a*^{+/+} LK cells after the 1st passage (n = 3 biologically independent samples each) and 2nd passage (n = 3 biologically independent samples each). In **c**, **e**, data are presented as mean values ± SEM. Statistical significance was calculated using two-sided Student's t-test. *P < 0.05, **P < 0.01.



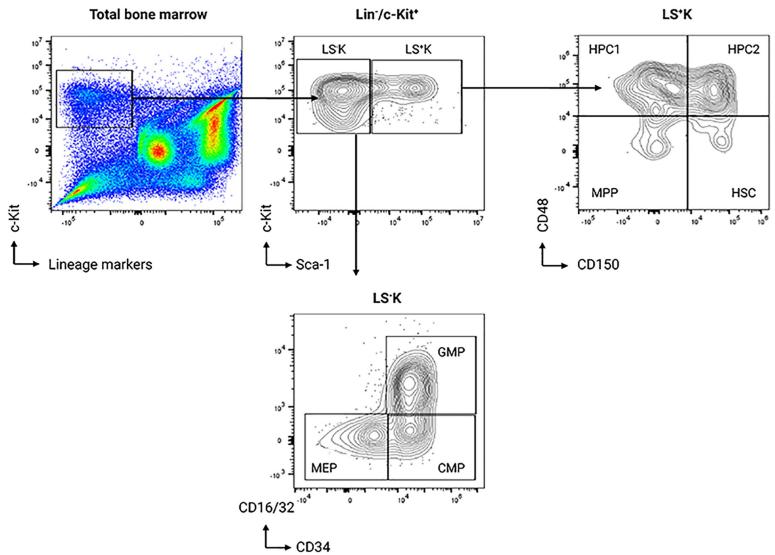
Extended Data Fig. 2 | See next page for caption.

Article

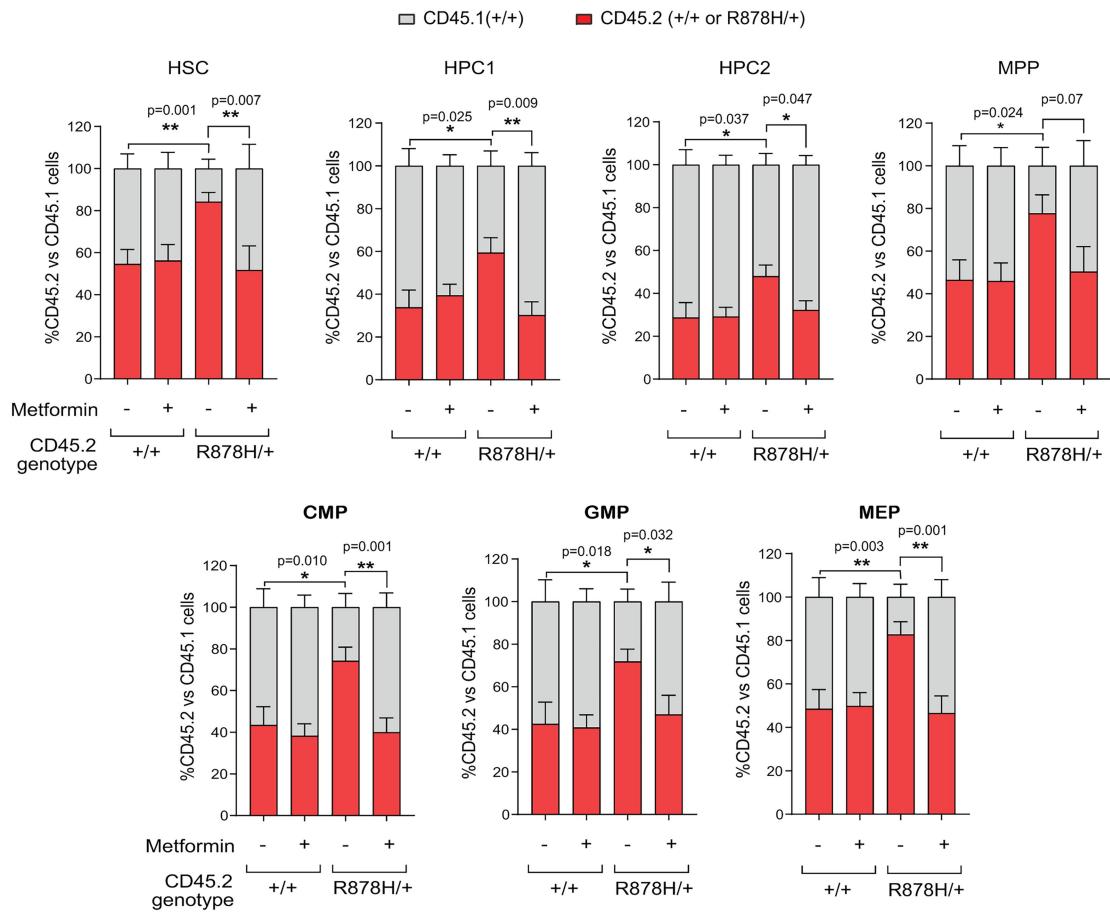
Extended Data Fig. 2 | *Dnmt3a*^{R878H/+} HSPCs have increased mitochondrial respiration compared with *Dnmt3a*^{+/+} cells and are dependent on this metabolic reprogramming for their competitive advantage. **a**, Expression of the indicated genes in *Dnmt3a*^{+/+} or *Dnmt3a*^{R878H/+} LK cells transduced with a lentiviral vector expressing a non-targeting shRNA or a shRNA targeting *Ndufv1* (sh*Ndufv1*) or *Cox15* (sh*Cox15*). n = 3 technical replicates for 1 biological sample in each condition. **b**, Number of colony forming units in the second plating from *Dnmt3a*^{+/+} or *Dnmt3a*^{R878H/+} LK HSPCs in the absence or presence of metformin. n = 3 biologically independent samples per condition. **c**, Schematic diagram showing the design of the in vivo competitive repopulation experiment.

Created in BioRender. Chan, S. (2025) <https://BioRender.com/wkp2h9m>.
d, Proportion of CD45.2⁺ vs. CD45.1⁺ cells in the myeloid (CD11b⁺), T (CD3⁺), and B (B220⁺) cell compartments in peripheral blood cells collected from mice after 4 months of treatment with metformin or vehicle in the experiment shown in Fig. 1h. n = 7 biologically independent samples per condition. In **a,b**, the box represents the interquartile range with the median indicated by the line inside the box. Whiskers extend to the minimum and maximum values. In **d**, data are presented as mean values ± SEM. Statistical significance was calculated using two-sided Student's t-test for all comparisons. *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001.

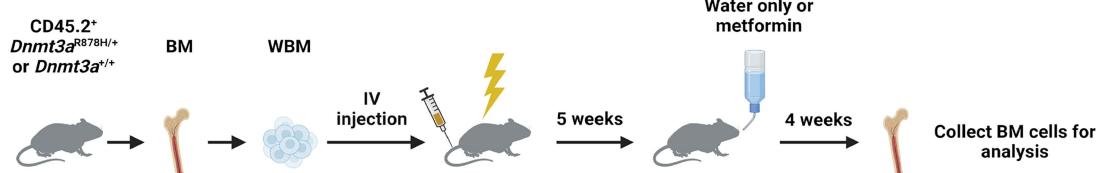
a



b



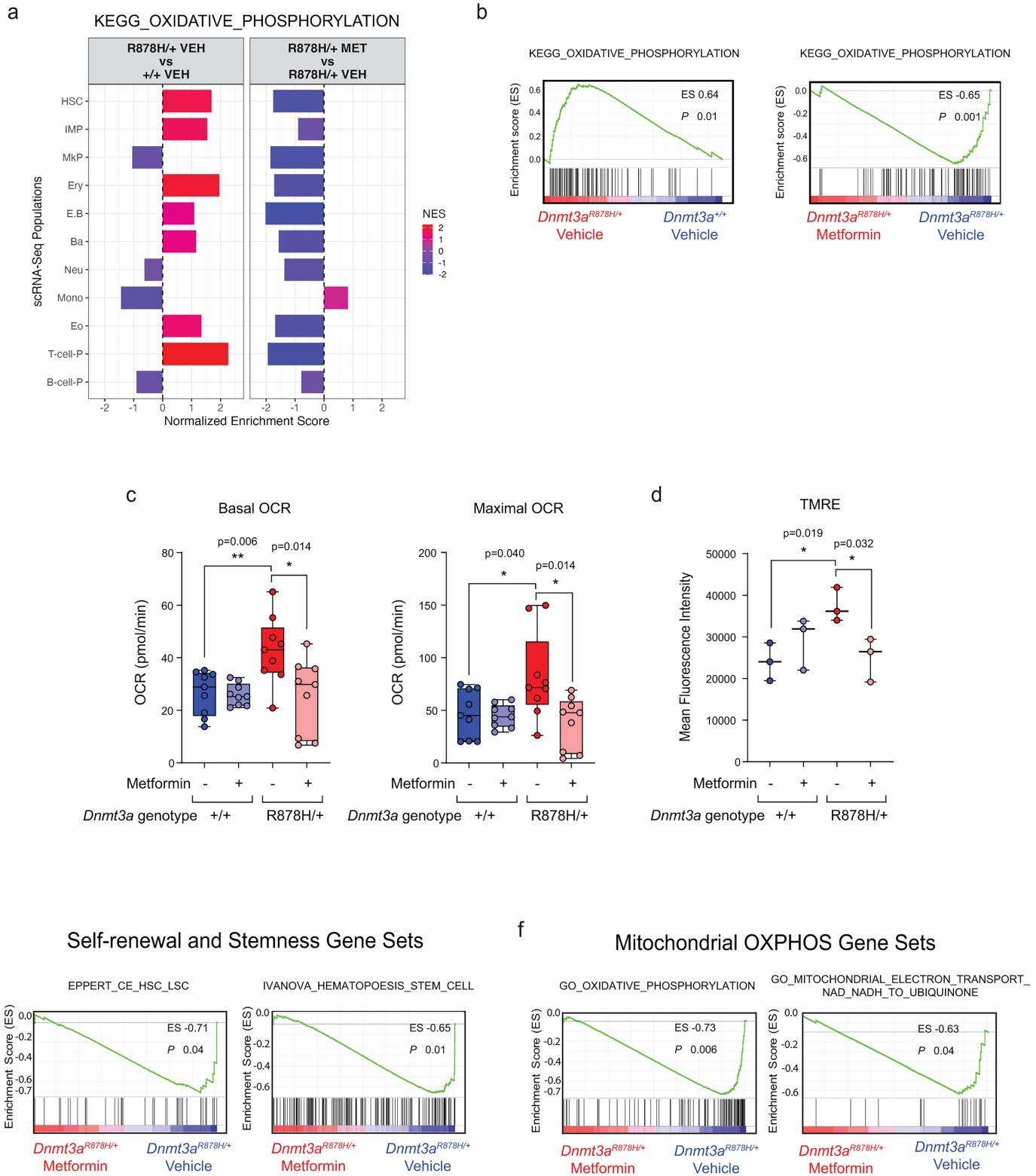
c

**Extended Data Fig. 3** | See next page for caption.

Article

Extended Data Fig. 3 | Metformin suppresses the competitive advantage of *Dnmt3a*^{R878H/+} HSCs. **a**, Gating strategy for the identification of murine HSPCs. **b**, Proportion of CD45.1⁺ *Dnmt3a*^{+/+} cells vs. CD45.2⁺ *Dnmt3a*^{R878H/+} or *Dnmt3a*^{+/+} cells in each immunophenotypically-defined HSPC subset. The bone marrow samples were collected from untreated and metformin-treated mice in the experiment shown in Fig. 1h after 4 months of treatment. The subsets were defined as: HSC (CD150⁺CD48⁻LS⁺K), MPP (CD150⁺CD48⁻LS⁺K), HPC1 (CD150⁻CD48⁺LS⁺K), HPC2 (CD150⁺CD48⁺LS⁺K), CMP (CD34⁺CD16/32^{Lo}LSK), GMP (CD34⁺CD16/32^{Hi}LS⁺K), and MEP (CD34⁺CD16/32^{Hi}LS⁺K). For HSC, MPP, HPC1 and HPC2, the number of biologically independent samples in the *Dnmt3a*^{+/+} VEH, *Dnmt3a*^{+/+} MET, *Dnmt3a*^{R878H/+} VEH, and *Dnmt3a*^{R878H/+} MET groups was 10, 12, 12, and 8, respectively. For CMP, GMP, and MEP, the number of biologically independent

samples in the *Dnmt3a*^{+/+} VEH, *Dnmt3a*^{+/+} MET, *Dnmt3a*^{R878H/+} VEH, and *Dnmt3a*^{R878H/+} MET groups was 9, 17, 11, and 11, respectively. Data are presented as mean values \pm SEM. Statistical significance was calculated using two-sided Student's t-test for all comparisons. *P < 0.05, **P < 0.01. **c**, Schematic diagram showing the design of the non-competitive repopulation experiment in which lethally-irradiated recipient mice were transplanted with CD45.2⁺ *Dnmt3a*^{+/+} or *Dnmt3a*^{R878H/+} whole bone marrow (WBM) cells. Five weeks after transplantation, the mice were either left untreated or treated with metformin in the drinking water. After 4 weeks of treatment, BM cells were collected and used for analysis. Created in BioRender. Chan, S. (2025) <https://BioRender.com/2k9dcho>.



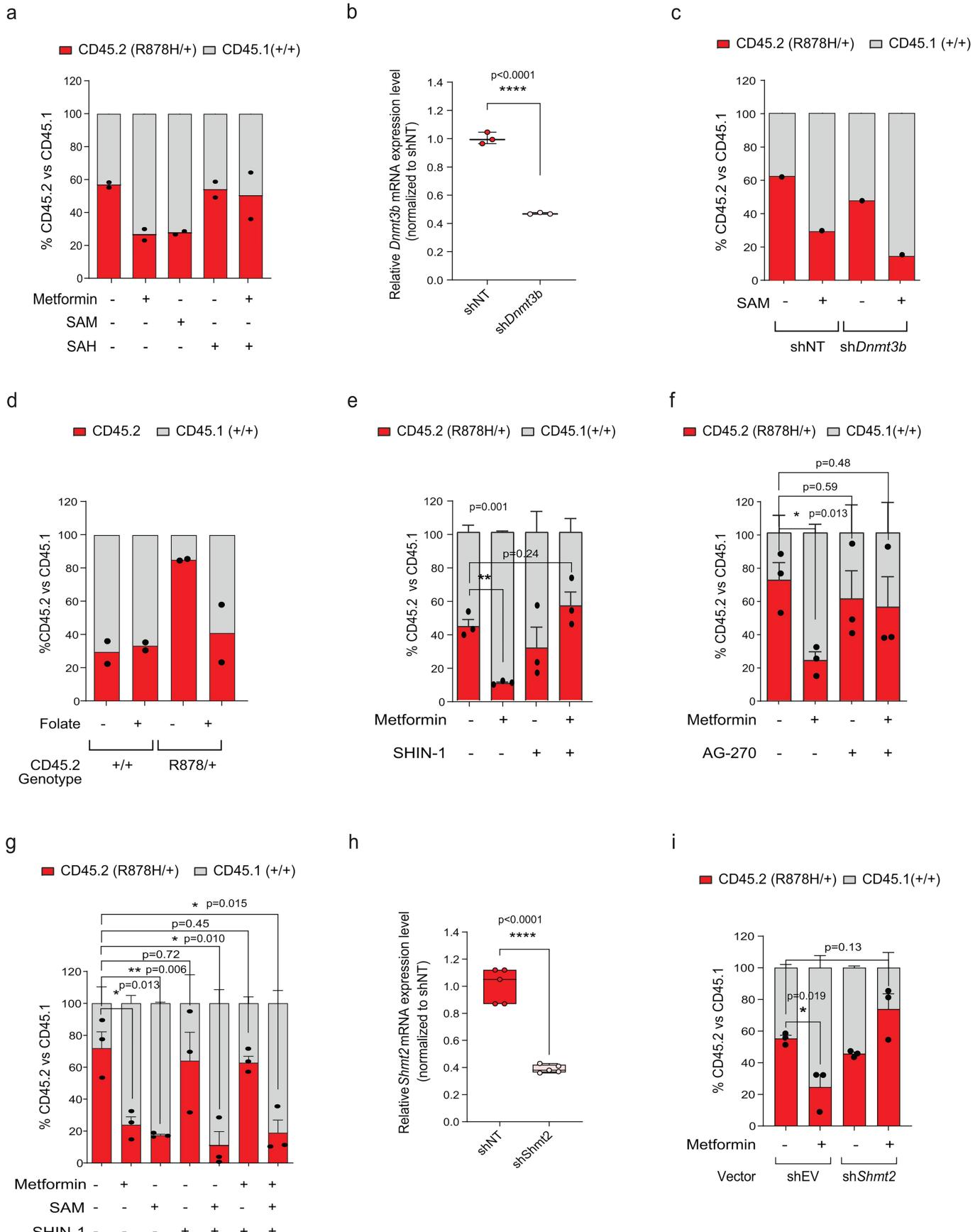
Extended Data Fig. 4 | See next page for caption.

Article

Extended Data Fig. 4 | Metformin suppresses OXPHOS in *Dnmt3a*^{R878H/+}

HSPCs. **a**, Normalized enrichment scores (NES) for the indicated gene set in HSPC subsets using scRNA-seq gene expression data and analyzed as pseudobulk populations. **b**, Gene set enrichment plots for the indicated gene set comparing *Dnmt3a*^{R878H/+} cells versus *Dnmt3a*^{+/+} cells and metformin-treated *Dnmt3a*^{R878H/+} versus vehicle-treated *Dnmt3a*^{R878H/+} cells in the HSC fraction using scRNA-seq gene expression data and analyzed as pseudobulk populations. **c**, Basal and maximal OCRs in LK HSPCs collected from mice transplanted with WBM cells of the indicated genotype and treated with or without metformin for 1 month. n = 9 biologically independent samples for each condition. **d**, Mean fluorescence intensity of TMRE staining in LK HSPCs collected from mice transplanted with WBM cells of the indicated genotype and treated with or without metformin for 1 month. n = 3 biologically independent samples for each condition. **e**, Gene set

enrichment plots of bulk RNA-seq data comparing metformin-treated *Dnmt3a*^{R878H/+} LK cells (n = 3 biologically independent samples) versus vehicle-treated *Dnmt3a*^{R878H/+} LK cells (n = 3 biologically independent samples) using the indicated self-renewal and stemness-related gene sets. **f**, Gene set enrichment plots of bulk RNA-seq data comparing metformin-treated *Dnmt3a*^{R878H/+} LK cells (n = 3 biologically independent samples) versus vehicle-treated *Dnmt3a*^{R878H/+} LK cells (n = 3 biologically independent samples) using the indicated mitochondrial OXPHOS-related gene sets. In **c**, **d**, the box represents the interquartile range with the median indicated by the line inside the box. Whiskers extend to the minimum and maximum values. Statistical significance was calculated using two-sided Student's t-test for all comparisons. *P < 0.05, **P < 0.01.



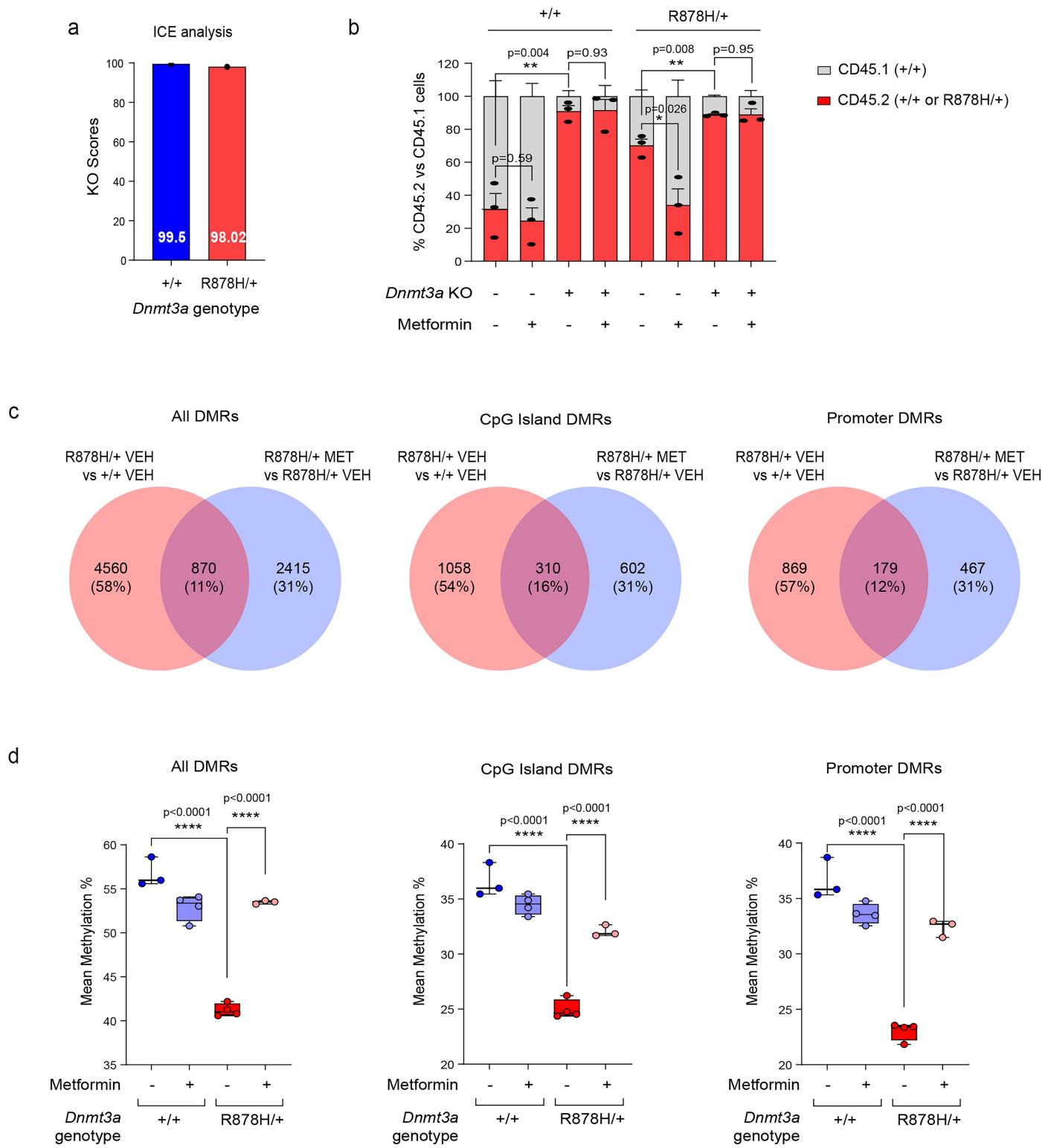
Extended Data Fig. 5 | See next page for caption.

Article

Extended Data Fig. 5 | Metformin suppresses the competitive advantage of *Dnmt3a*^{R878H/+} HSPCs by enhancing their methylation potential.

a, Proportion of CD45.2⁺ and CD45.1⁺ cells in a competition assay between CD45.2⁺ *Dnmt3a*^{R878H/+} LK cells and CD45.1⁺ *Dnmt3a*^{+/+} LK cells in the presence or absence of the indicated compounds. n = 2 biologically independent samples for each condition. **b**, Expression of *Dnmt3b* in *Dnmt3a*^{+/+} LK cells transduced with a shRNA vector expressing a non-targeting shRNA (shNT) or sh*Dnmt3b*. n = 3 technical replicates for 1 biological sample in each condition. **c**, Proportion of CD45.2⁺ and CD45.1⁺ cells in a competition assay between CD45.2⁺ *Dnmt3a*^{R878H/+} LK cells and CD45.1⁺ *Dnmt3a*^{+/+} LK cells in the presence or absence of exogenous SAM. Both populations were transduced with the indicated shRNA vectors. n = 1 biological sample per condition. The mean value of 5 technical replicates is shown for each condition. **d**, Proportion of CD45.2⁺ and CD45.1⁺ cells in a competition assay between CD45.2⁺ *Dnmt3a*^{R878H/+} or *Dnmt3a*^{+/+} LK cells and CD45.1⁺ *Dnmt3a*^{+/+} LK cells in the presence or absence of folic acid supplementation at 100 μM. n = 2 biologically independent samples per condition. **e**, Proportion of CD45.2⁺ and CD45.1⁺ cells in a competition assay between CD45.2⁺ *Dnmt3a*^{R878H/+} LK cells and CD45.1⁺ *Dnmt3a*^{+/+} LK cells in the presence or absence of the indicated compounds. n = 3 biologically independent

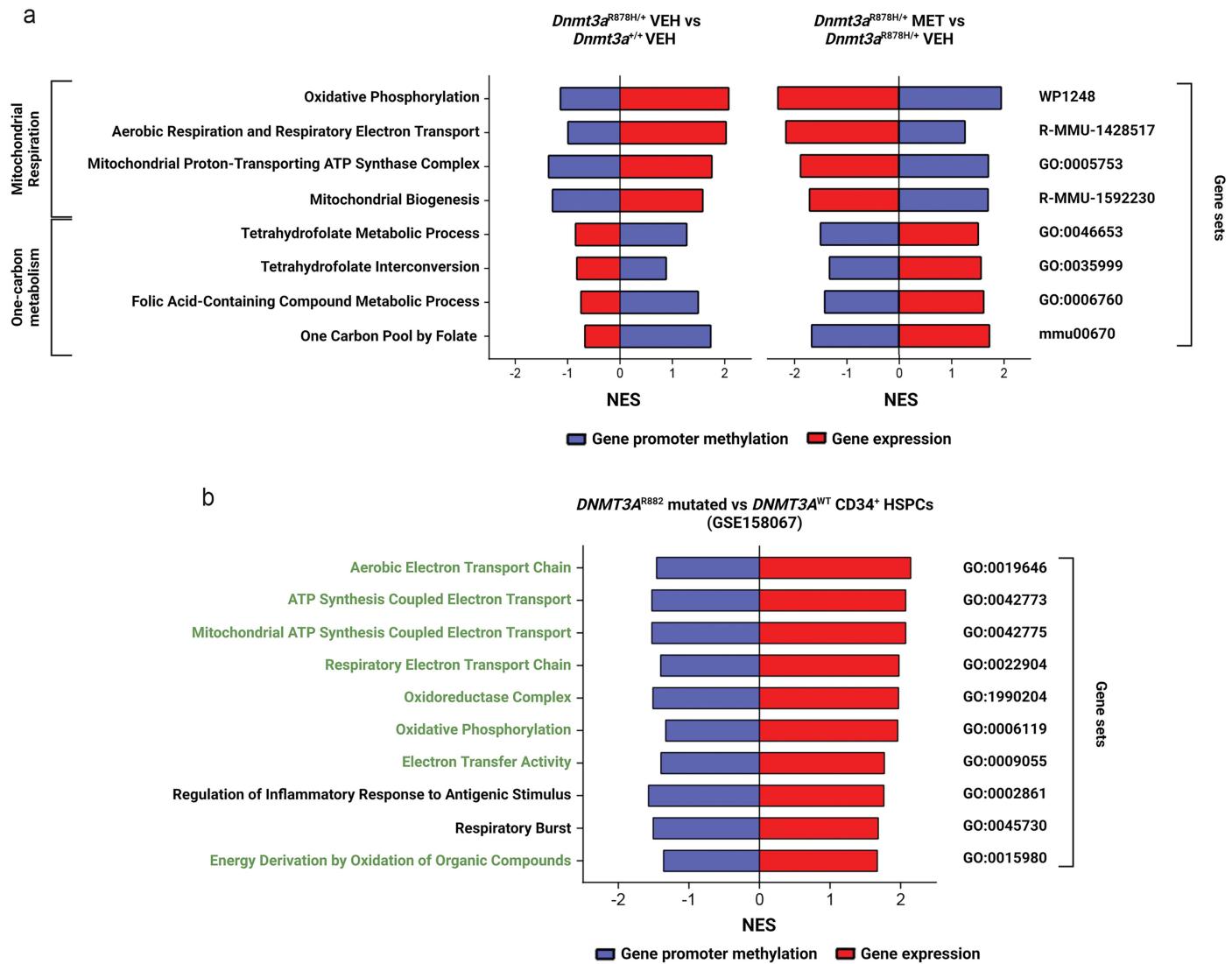
samples for each condition. **f**, Proportion of CD45.2⁺ and CD45.1⁺ cells in a competition assay between CD45.2⁺ *Dnmt3a*^{R878H/+} LK cells and CD45.1⁺ *Dnmt3a*^{+/+} LK cells in the presence or absence of the indicated compounds. n = 3 biologically independent samples for each condition. **g**, Proportion of CD45.2⁺ and CD45.1⁺ cells in a competition assay between CD45.2⁺ *Dnmt3a*^{R878H/+} LK cells and CD45.1⁺ *Dnmt3a*^{+/+} LK cells in the presence or absence of the indicated compounds. n = 3 biologically independent samples per condition. **h**, Expression of *Shmt2* in *Dnmt3a*^{+/+} LK cells transduced with an empty shRNA vector control (shEV) or a shRNA vector expressing sh*Shmt2*. n = 5 technical replicates for 1 biological sample in each condition. **i**, Proportion of CD45.2⁺ and CD45.1⁺ cells in a competition assay between CD45.2⁺ *Dnmt3a*^{R878H/+} LK cells and CD45.1⁺ *Dnmt3a*^{+/+} LK cells in the presence or absence of metformin. Both populations were transduced with the indicated shRNA vectors. n = 3 biologically independent samples per condition. In **a**, **c**, **d**, the bar blot represents the mean value. In **b**, **h**, the box represents the interquartile range with the median indicated by the line inside the box. Whiskers extend to the minimum and maximum values. In **e**, **f**, **g**, **i**, data are presented as mean values ± SEM. Statistical significance was calculated using two-sided Student's t-test for all comparisons. *P < 0.05, **P < 0.01, and ****P < 0.0001.



Extended Data Fig. 6 | See next page for caption.

Article

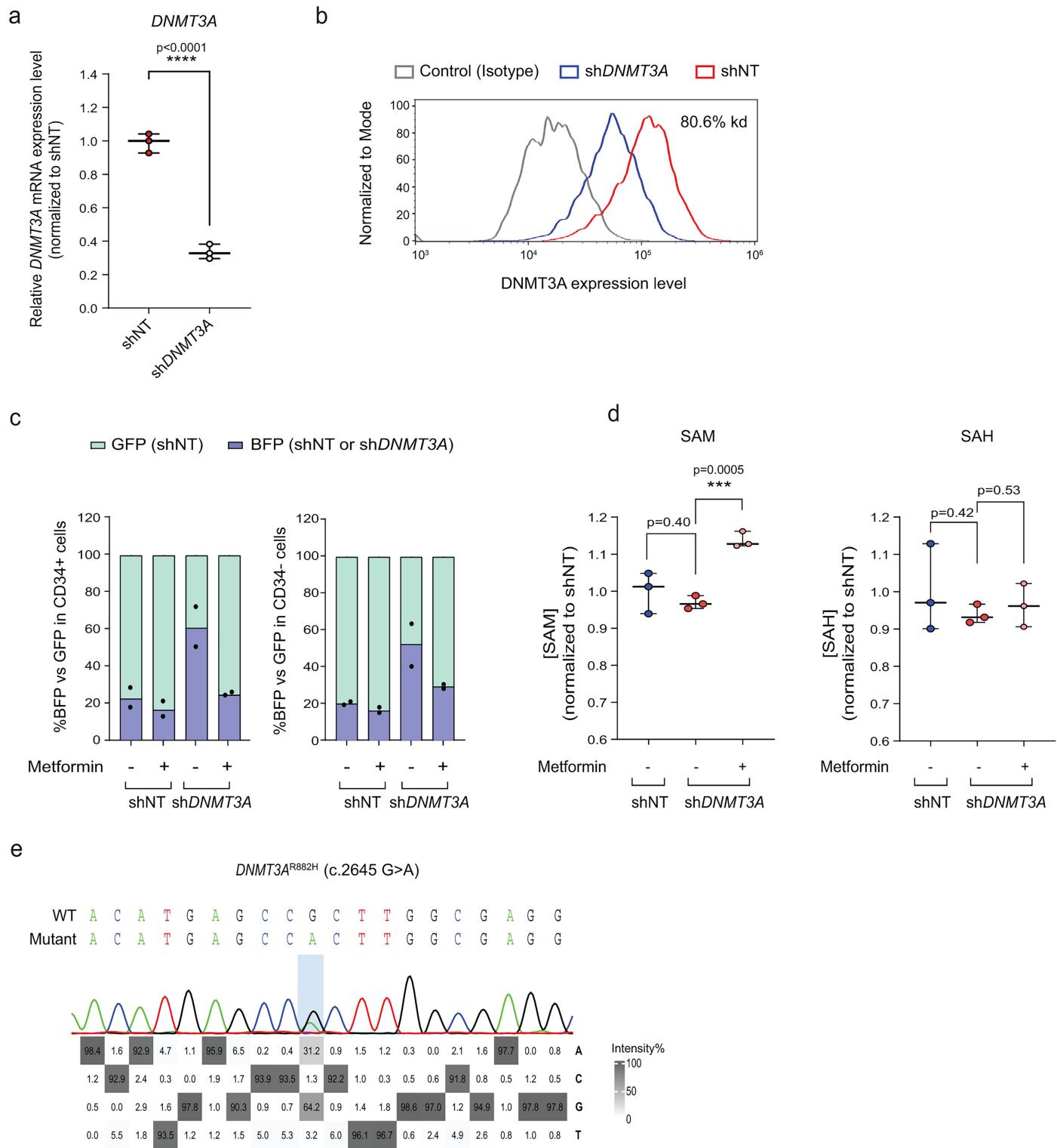
Extended Data Fig. 6 | Metformin reverses the aberrant DNA CpG methylation and H3K27me3 profiles in *Dnmt3a*^{R878H/+} HSPCs. **a**, Knockout score for *Dnmt3a* in *Dnmt3a*^{R878H/+} or *Dnmt3a*^{+/+} LK cells as determined by Inference of CRISPR Edits (ICE) analysis 4 days after nucleofection with CRISPR/Cas9 ribonucleoprotein (RNP) complexes targeting the *Dnmt3a* gene. **b**, Proportion of CD45.2⁺ and CD45.1⁺ cells in a competition assay between CD45.2⁺ LK cells of the indicated genotype and CD45.1⁺ *Dnmt3a*^{+/+} LK cells in the presence or absence of metformin. The CD45.2⁺ LK cells were nucleofected with Cas9 alone (-) or with RNPs targeting *Dnmt3a* (+) prior to mixing and plating. The competing CD45.1⁺ LK cells were nucleofected with Cas9 alone prior to plating. n = 3 biologically independent samples for each condition. **c**, Venn diagram showing the overlap of DMRs between untreated *Dnmt3a*^{R878H/+} samples versus untreated *Dnmt3a*^{+/+} samples and between metformin-treated *Dnmt3a*^{R878H/+} samples versus untreated *Dnmt3a*^{R878H/+} samples. **d**, Average methylation levels of DMRs in LK cells of the indicated genotype and treated with or without metformin for 1 month. DMRs were restricted to the ones that were hypomethylated in vehicle-treated *Dnmt3a*^{R878H/+} samples relative to *Dnmt3a*^{R878H/+} samples. Each dot represents the average value of the DMRs in one sample. n = 3 biologically independent samples for untreated *Dnmt3a*^{+/+} samples and metformin-treated *Dnmt3a*^{R878H/+} samples. n = 4 biologically independent samples for metformin-treated *Dnmt3a*^{+/+} samples and untreated *Dnmt3a*^{R878H/+} samples. In **d**, the box represents the interquartile range with the median indicated by the line inside the box. Whiskers extend to the minimum and maximum values. In **b**, data are presented as mean values ± SEM. Statistical significance was calculated using two-sided Student's t-test for all comparisons. *P < 0.05, **P < 0.01, and ***P < 0.0001.



Extended Data Fig. 7 | *Dnmt3a^{R878H}* mutation decreases promoter methylation and increases expression of genes involved in mitochondrial respiration. **a**, Normalized enrichment scores (NES) for the indicated gene sets from GSEA using genes ranked by differential promoter methylation or gene expression in the comparison between vehicle-treated *Dnmt3a^{R878H/+}* cells versus vehicle-treated *Dnmt3a^{+/+}* cells and between metformin treated *Dnmt3a^{R878H/+}* cells versus vehicle-treated *Dnmt3a^{R878H/+}* cells. **b**, Normalized

enrichment scores (NES) for the indicated gene sets from GSEA using genes ranked by differential promoter methylation or gene expression in the comparison between *DNMT3A^{R882}*-mutated versus *DNMT3A^{WT}* HSPCs as pseudobulk populations in the GSE158067 dataset. GSEA was performed using all Gene Ontology (GO) gene sets. The top 10 enriched GO gene sets ranked based on NES from gene expression are shown. The ones shown in green color are related to mitochondrial respiration.

Article



Extended Data Fig. 8 | Metformin decreases the competitive advantage of human *DNMT3A*^{R882H} HSPCs. **a**, Expression of *DNMT3A* mRNA in human CD34⁺ enriched HSPCs transduced with a lentiviral vector expressing a non-targeting shRNA (shNT) or *DNMT3A* shRNA. n = 3 technical replicates for 1 biological sample in each condition. **b**, Intracellular flow cytometry staining for DNMT3A protein in human CD34⁺ enriched HSPCs transduced with a lentiviral vector expressing a non-targeting shRNA (shNT) or *DNMT3A* shRNA. The histogram depicting isotype control antibody staining represents cells expressing the non-targeting shRNA (shNT). **c**, Proportion of BFP⁺ and GFP⁺ cells in a competition assay between BFP⁺ HSPCs expressing shNT or sh*DNMT3A* and GFP⁺ HSPCs expressing shNT in the absence or presence of metformin at 50 μ M.

Analysis was gated on the CD34⁺ cells in the left panel and CD34⁻ cells in the right panel. n = 2 biologically independent samples for each condition. **d**, Concentration SAM and SAH in human HSPCs expressing the indicated shRNA and treated with or without metformin. n = 3 biologically independent samples for each condition. **e**, Representative Sanger sequencing chromatogram of the sequences surrounding the *DNMT3A* p.R882H (c.2645 G > A) missense mutation in a prime-edited HSPC cell pool. In **a**, **d**, the box represents the interquartile range with the median indicated by the line inside the box. Whiskers extend to the minimum and maximum values. In **c**, bar plots represent mean values. Statistical significance was calculated using two-sided Student's t-test for all comparisons. ***P < 0.001.

Corresponding author(s): Steven M. Chan

Last updated by author(s): Feb 18, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Flow cytometry analyses were conducted with CytoFLEX software (v2.4). Seahorse analyses were performed using Wave (v2.6.3), and quantitative PCR results were evaluated with BioRad CFX Manager Software (v3.1).
Data analysis	CITE seq downstream and statistical analyses were conducted utilizing Cell Ranger (v6.1.2), Seurat package v4, scDblFinder 1.16.0, and R package scuttle (v1.0.4). For both bulk RNA-seq and scRNA data, R package edgeR 3.36.0 was used. RRBS downstream and statistical analyses were assessed using Trimmomatic-0.36, FastQC (v0.11.5), R package methylKit (v1.26.0), methylKit, plyranges R package (v1.20.0), and ChIPseeker R package (v1.36.0), R package ggplot2 (v3.5.1). ChIP-Seq analyses were performed using bcl2fastq2 v2.20, BWA v0.7.12 algorithm, MACS algorithm (v2.1.0), R package ggplot2 (v3.5.1), DeepTools 3.5.1. R v4.3.2. EditR software v1.0.0. GraphPad Prism v10. FlowJoTM v10. SnapGene v4.3.11.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The Gene Expression Omnibus (GEO) serves as the repository for all omics data associated with this study under the accession number GSE255101. Additionally, datasets GSE27187, GSE68833, and GSE158067 were utilized. Pathway and functional enrichment analyses were conducted using WikiPathway One Carbon Metabolism, KEGG Oxidative Phosphorylation, GO Oxidative Phosphorylation, and GO Mitochondrial Electron Transport NAD_NADH to UBIQUINONE databases, as well as hematopoietic stem cell signature datasets, including EPPERT-ce-HSC-LSC and Ivanova-Hematopoiesis Stem Cell Signatures.

All codes are available at the following link: <https://github.com/veroniquevoisin/DNMT3A>

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Cord blood samples were collected from female donors. Information on the age of the mother and sex of the infant was not provided.

Reporting on race, ethnicity, or other socially relevant groupings

Information on race, ethnicity, or other socially relevant groups of the donors was not provided.

Population characteristics

Consenting female donors of cord blood.

Recruitment

Donors in this study were approached for cord blood donation. Cord blood was collected from consenting donors. No selection was made based on demographic, clinical, or biological factors.

Ethics oversight

Cord blood (CB) samples were obtained with informed consent from Trillium Health, Credit Valley and William Osler Hospitals according to procedures approved by the University Health Network (UHN) Research Ethics Board.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The sample size was determined based on prior experience, not based on a power analysis.

Data exclusions

We excluded 2 biological replicates from RRBS analysis due to low number of read counts.

Replication

We assessed biologically independent and technical replicates. Please see the figure legends for details on the number of replicates used in each experiment.

Randomization

In our mouse treatment experiments, we used cage-level randomization rather than individual animal assignment. Each cage—housing up to 5 mice—was randomly allocated to receive either water (control) or metformin (treatment) using a computer-generated randomization schedule.

Blinding

Blinding was not implemented due to experimental constraints inherent to our study design. Mice were randomized at the cage level—with each cage receiving a uniform treatment of either water or metformin. Because the treatments required distinct administration protocols and handling procedures, effective blinding of the investigators was not feasible.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

Antibody	/Conjugate/Clone/	Supplier/	Cat. #/	Dilutions
Anti-mouse CD45.2/	AF700/Clone: 104	/Biolegend/	109822/	1:100
Anti-mouse CD45.1	/APC-Fire 750/	Clone: A20/Biolegend/	110752	1:100
Anti-mouse Lineage markers	/FITC/	Multiple clones/	Biolegend/	133302
Anti-mouse Lineage markers	/Pacific Blue/	Multiple clones/	Biolegend/	133306
Anti-mouse c-Kit (CD117)/	PE/	Clone: 2B8/BD Biosciences/	553355/	1:100
Anti-mouse Sca-1/BV421/	Clone: D7 /Biolegend/		108128/	1:100
Anti-mouse CD150	/BV711/	Clone: TC15-12F12.2/	Biolegend/	115941
Anti-mouse CD48	/APC/Clone: HM48-1/	Biolegend/	103412	1:100
Anti-mouse TER-119	/FITC/	Clone: TER-119/	Biolegend/	116206
Anti-mouse CD11b	/PE-Cyanine7/	Clone: M1/70/	Biolegend/	101216
Anti-mouse Ly-6G/Ly-6C (Gr-1)/	BV711/	Clone: RB6-8C5/	Biolegend/	108443
Anti-mouse CD3ε	/APC/	Clone: 145-2C11	/Biolegend/	100312
Anti-mouse CD45R(B220)/	PE/	Clone: RA3-6B2/	Biolegend/	103208
Anti-Ki-67/FITC/	Clone: B56 /BD Biosciences/	556026		1:200
Mouse IgG1k isotype antibody/	FITC/	Clone: MOPC-21/	Biolegend/	400107/
Anti-mouse CD45.2/	TotalSeq™-B0157	/Clone: 104/	Biolegend/	109859
Anti-mouse CD45.1	/TotalSeq™-B0178	/Clone: A20	/Biolegend/	110755
Anti-DNMT3A/	Alexa Fluor 488/	Clone: D2H4B/	Cell Signaling	Technologies/83694S
Rabbit (DA1E) mAb IgG XP® Isotype Control/Alexa Fluor 488/	2975S	Clone: DA1E/	Cell Signaling Technologies/	1:50
Tri-Methyl-Histone H3 (Lys27)/	unconjugated	/Clone: C36B11	/Cell Signaling Technologies/	97335
Anti-rabbit IgG (H+L), F(ab')2 Fragment/Alexa Fluor 488	/Clone: Not applicable/		Cell Signaling Technologies/	
4412S	/1:2000			

Validation

The following antibodies were used, with their respective manufacturer validation sources:

Anti-mouse CD45.2: <https://www.biolegend.com/en-us/products/alex-fluor-700-anti-mouse-cd45-2-antibody-3393>
 Anti-mouse CD45.1: <https://www.biolegend.com/en-us/products/apc-fire-750-anti-mouse-cd45-1-antibody-13657>
 Anti-mouse Lineage markers: <https://www.biolegend.com/en-us/products/fitc-anti-mouse-lineage-cocktail-with-isotype-ctrl-5803>
 Anti-mouse Lineage markers: <https://www.biolegend.com/en-us/products/pacific-blue-anti-mouse-lineage-cocktail-with-isotype-ctrl-7556>
 Anti-mouse c-Kit (CD117): https://www.bdbiosciences.com/en-ca/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/pe-rat-anti-mouse-cd117.553355?tab=product_details
 Anti-mouse Sca-1: <https://www.biolegend.com/en-us/products/brilliant-violet-421-anti-mouse-ly-6a-e-sca-1-antibody-7198>
 Anti-mouse CD150: <https://www.biolegend.com/en-us/products/brilliant-violet-711-anti-mouse-cd150-slam-antibody-13494>
 Anti-mouse CD48: <https://www.biolegend.com/en-us/products/apc-anti-mouse-cd48-antibody-3622>
 Anti-mouse TER-119: <https://www.biolegend.com/en-us/products/fitc-anti-mouse-ter-119-erythroid-cells-antibody-1865>
 Anti-mouse CD11b: <https://www.biolegend.com/en-us/products/pe-cyanine7-anti-mouse-human-cd11b-antibody-1921>
 Anti-mouse Ly-6G/Ly-6C(Gr-1): <https://www.biolegend.com/en-us/products/brilliant-violet-711-anti-mouse-ly-6g-ly-6c-gr-1-antibody-8896>
 Anti-mouse CD3ε: <https://www.biolegend.com/en-us/products/apc-anti-mouse-cd3epsilon-antibody-21>
 Anti-mouse CD45R/B220: <https://www.biolegend.com/en-us/products/pe-anti-mouse-human-cd45r-b220-antibody-447>
 Annexin V : <https://www.biolegend.com/en-us/products/apc-annexin-v-8144>
 SYTOX blue: <https://www.thermofisher.com/order/catalog/product/S34857>
 SYTOX green: <https://www.thermofisher.com/order/catalog/product/S34860>
 Anti-mouse Ki-67: https://www.bdbiosciences.com/en-ca/products/reagents/flow-cytometry-reagents/research-reagents/panels-multicolor-cocktails-ruo/fitc-mouse-anti-ki-67-set.556026?tab=product_details
 Mouse IgG1k isotype antibody: https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/flow-cytometry-controls-and-lysates/fitc-mouse-igg1-isotype-control.551954?tab=product_details
 Anti-mouse CD45.2: <https://www.biolegend.com/en-us/products/totalseq-b0157-anti-mouse-cd45-2-antibody-18438>
 Anti-mouse CD45.1: <https://www.biolegend.com/en-us/products/totalseq-b0178-anti-mouse-cd45-1-antibody-18449>
 Anti-DNMT3A (D2H4B): https://www.cellsignal.com/products/antibody-conjugates/dnmt3a-d2h4b-rabbit-mab-alex-fluor-488-conjugate/836947srslid=AfmBOopxJFn1J4Bwkpc1lyVfxld5Cis0_5vWN6Vs15bd6klj1IKNj61
 Rabbit (DA1E) mAb IgG XP® Isotype Control (Alexa Fluor® 488 Conjugate): <https://www.cellsignal.com/products/antibody-2975S>

conjugates/rabbit-da1e-mab-igg-xp-isotype-control-alexa-fluor-488-conjugate/2975?srsltid=AfmBOorUp_TlwsQuYZsp8vSuOLo2USJ_h6jrTStdshfqvBFErvRO6Fp
 Tri-Methyl-Histone H3 (Lys27) (C36B11): <https://www.cellsignal.com/products/primary-antibodies/tri-methyl-histone-h3-lys27-c36b11-rabbit-mab/9733?srsltid=AfmBOorq9-qtgGZl6hIKDW-Ije9PVyVGUCD9oL6-oGudINhgfMMMeA08h>
 Anti-rabbit IgG (H+L), F(ab')2 Fragment: https://www.cellsignal.com/products/secondary-antibodies/anti-rabbit-igg-h-l-f-ab-2-fragment-alexa-fluor-488-conjugate/4412?srsltid=AfmBOoo8Vxo6vL417V_siL9ro1OWaTnmCND4j96l1FocgCT-RkOq5vEF
 Anti-human CD34 (581): <https://www.biologics.com/en-us/products/apc-anti-human-cd34-antibody-6090>

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	HEK293T cells were purchased from ATCC (CRL-3216).
Authentication	STR profiling was used to verify the identity of the cell line.
Mycoplasma contamination	The cells were tested and confirmed to be mycoplasma-free.
Commonly misidentified lines (See ICLAC register)	None

Animals and other research organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	1. C57BL/6J mice from The Jackson Laboratory 2. B6.SJL-Ptpcra Pepcb/BoyJ (Pep-Boy) mice from The Jackson Laboratory. 3. Dnmt3afl-R878H/+ mice (Jackson Laboratory, Strain# 032289) 4. Mx-Cre mice (Jackson Laboratory, Strain# 003556)
Wild animals	No wild animals were included in this study.
Reporting on sex	All mice were female.
Field-collected samples	There were no field-collected samples used in this study.
Ethics oversight	All in vivo experiments were performed in accordance with institutional guidelines approved by the University Health Network Animal care committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A

ChIP-seq

Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE255101
<i>May remain private before publication.</i>	

Files in database submission	FASTQ and bigwig and narrow peak (BED) files
Genome browser session (e.g. UCSC)	Genome browser session will be provided prior to publication.

Methodology

Replicates	Mutant and wild-type mice untreated and treated with metformin had 2 biological replicates each for ChIP-Seq assay.
Sequencing depth	A minimum threshold of 30 million reads per sample was set.
Antibodies	15ug chromatin and 4ul of antibody against H3K27me3 (Active Motif cat# 39155) were used to immunoprecipitated genomic DNA regions.
Peak calling parameters	Peak locations were determined using the MACS algorithm (v2.1.0) with default parameters and a cutoff of p-value =1e-7. Peaks that were on the ENCODE blacklist of known false ChIP-Seq peaks were removed.
Data quality	18556 peaks that were identified in at least one sample with a cutoff p-value of 1e-7 were all merged in a common matrix and the total number of present peaks as well as averaged peak values were calculated and plotted for each sample and condition
Software	bcl2fastq2 (v2.20): processing of Illumina base-call data and demultiplexing. bwa (v0.7.12): alignment of reads to reference genome. Samtools (v0.1.19): processing of BAM files. BEDTools (v2.25.0): processing of BED files. MACS2 (v2.1.0): peak calling; narrow peaks. SICER (v1.1): peak calling: broad peaks. wigToBigWig (v4): generation of bigWIG files and DeepTools 3.5.1 for generating profile heatmap.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	The isolated mouse bone marrow cells were treated with RBC lysis buffers (Biolegend, Cat#420302) in order to remove red blood cells. Then the cells stained for 30min at 4°C with antibodies (listed in Supplementary Table 4) at the suggested dilutions in 100ul of FACS buffer (HBSS supplemented with 2% FBS and 0.1% sodium azide) and washed with PBS 1X once prior to flow cytometry analysis.
Instrument	Beckman Coulter CytoFLEX
Software	FlowJoTM V10
Cell population abundance	The bone marrow (BM) cells were enriched for Lin-Kit+ cells (HSPCs) using the EasySep mouse hematopoietic progenitor isolation kit (StemCell Technologies, Cat# 19856), followed by further enrichment using the c-KIT positive enrichment kit (StemCell Technologies, Cat# 18757). After performing the sorting, the proportion of lin- and kit+ cells was determined to be between 90 to 95%. This was achieved by utilizing particular antibodies against the lin cocktail (Biolegend, Cat#133302) and the CD117 antibody (Biolegend, Cat#133306), and analyzing the cells using a CytoFlex flow cytometer.
Gating strategy	The gating strategy for all immunophenotypic experiments involved first selecting healthy cells and excluding debris based on the Forward Scatter (FSC) and Side Scatter (SSC) axes. For the Peripheral blood (PB) chimerism analysis we performed peripheral blood sampling and red blood cells were then excluded using RBC lysis buffer and the remaining cells were stained with the following antibodies listed below: CD45.2, CD45.1, TER-119, CD11b, B220, CD3, and SYTOX Blue. According to our gating strategy, we initially gated on SYTOX-negative and TER-119-negative cells to remove non-viable and erythroid cells, respectively. Subsequently, we used two different approaches: the first was to determine the ratio of CD45.2 to CD45.1 stained cells. The second approach involved categorizing the CD45.2 to CD45.1 ratios within the myeloid population, B cells, and T cells. To achieve this, the myeloid population was recognized as CD11b-positive cells, and CD11b-negative cells stained with B220 were classed as B cells, while CD3-positive cells were identified as T cells. Following that, we gated the labeled cells with CD45.2 versus CD45.1 to calculate the CD45.2 vs CD45.1 ratio in the myeloid, B, and T cell populations of each mouse. In regard to the HSC, HPC1 and HPC2 gating strategy, mouse bone marrow cells were stained with the following antibodies:

Lin cocktail, CD117, Sca-1, CD48, CD150, and SYTOX green, subsequent to the removal of red blood cells. Viable HSC cells were identified as lineage negative cells that stained negatively for SYTOX but positively for Sca-1 and CD117; they belonged to the CD48-negative and CD150-positive populations. HPC1 and HPC2 Viable cells were identified as lineage negative cells that stained negatively for SYTOX we selected Sca-1 and CD117 positive cells; HPC1 belonged to the CD48-positive and CD150-negative populations and HPC2 belonged to the CD48-positive and CD150-populations. The supplementary table 4 contains all the specific information regarding the antibodies.

For the identification of CMP, GMP, and MEP populations, mouse bone marrow cells were stained with a panel of antibodies, including a lineage cocktail, CD117, Sca-1, CD16/32, CD34, and SYTOX Green, following the lysis of red blood cells. Viable cells were identified based on SYTOX negativity. CMPs were characterized as lineage-negative, CD117-positive, Sca-1-negative, CD16/32-, and CD34-positive. GMPs were defined as Lin-, CD117+, Sca-1-, CD16/32+, and CD34+. MEPs were identified as Lin-, CD117+, Sca-1-, CD16/32-, CD34-.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.