SUPPLEMENTARY INFORMATION

Supplementary Figure Legends

Supplementary Figure 1 | LysM-KI mice show normal hematological parameters in peripheral blood.

a-e, Standard hematological analyses of peripheral blood from young (7-16 weeks) and older (42-46 weeks) mice (n=7 for control and n=6 for LysM-KI). Results for individual mice are shown (horizontal line = mean). WBC, white blood cells; RBC, red blood cells; Hct, hematocrit; Hb, hemoglobin; MCHC, mean corpuscular hemoglobin concentration. **f**, Summary and statistical analyses of parameters displayed in (**a-e**). Results are the mean ± SEM. No statistically significant differences were observed. p-values were determined using the unpaired t test with Welch's correction.

Supplementary Figure 2 | KI ES cells expressing R132H-Idh1 show normal *in vitro* differentiation into various hematopoietic lineages.

a-c, A control ES cell clone, a KI ES cell clone expressing R132H-Idh1, and a control ESR1 ES cell line were co-cultured in the ES cell-OP9/OP9-DL1 co-culture system to induce *in vitro* differentiation into different lineages of hematopoietic cells[9]. FACS staining to detect established lineage markers was performed as indicated for: **a**, myeloid differentiation after 12 days in culture; **b-c**, T cell differentiation after 16 days in culture. Results are representative of 2 independent experiments employing 2 different control and KI ES cell clones each. No statistically significant differences were observed.

Supplementary Figure 3 | LysM-KI HSCs show normal short-term and long-term competitive repopulation capacity.

a,b, Analysis of peripheral blood from lethally-irradiated (10.5 Gy) recipient mice after transplantation with $1x10^5$ donor control or LysM-KI nucleated BM cells ($CD45.2^+$) plus $1x10^5$ competitor nucleated BM cells ($CD45.1^+$) into $CD45.1^+$ hosts. **a**, Percentage of cells in the peripheral blood derived from donor BM after the indicated number of days as determined by FACS (control (n=2) or LysM-KI (n=3), data are representative for three independent experiments). **b**, Reconstitution in recipients of the indicated lineages of peripheral blood cells at 40 days after transplantation of control or LysM-KI BM cells (n=10/group). For a and b, results are the mean±SEM. No statistically significant differences were observed.

Supplementary Figure 4 | LysM-KI LSK cells show normal mRNA levels of Hif1α target genes.

mRNA was extracted from LSK cells sorted from the BM of young (9-16 weeks) control or LysM-KI mice (n=7/group), transcribed into cDNA, and analyzed by real-time RT-PCR to detect transcripts of the indicated HIF1α target genes. Hprt was used as the housekeeping gene control. Analyses were performed in triplicate for each sample and each mRNA. Shown are the relative mRNA expression levels relative to the average of the expression levels in control LSK cells (mean ± SEM, p-values were calculated using the unpaired t test with Welch's correction).

Supplementary Figure 5 | Proposed model for the effects of the Idh1-R132H mutation

This schematic representation summarizes our findings in mice in which the Idh1-R132H mutation was

inserted into the murine *Idh1* gene and conditionally expressed in myeloid cells under the control of LysMCre. We observed elevated 2HG levels, increased histone methylation and enhanced DNA methylation, as well as alterations to numbers of lineage-restricted precursors (LRPs). We speculate that either a partial block in cellular differentiation and/or an increase in symmetric cell division within the LRP pool may be responsible for the LRP accumulation in LysM-KI mice. As the size of the LRP pool expands, the number of cells at risk for acquiring a second tumorigenic hit increases, thereby raising the risk that these hematopoietic progenitors may give rise to AML cells.

Supplementary Table Legends

Supplementary Table 1 | LysM-KI mice are born at the expected Mendelian ratio.

For breeding mice heterozygous for either genotype ($LysMCre^{+/WT}$ and $Idh1^{LSL/WT}$) were crossed or LysM-KI mice were crossed with WT C57/B6 mice. Results are shown for a total of 41 litters (mean litter size = 6.78) (WT-: $Idh1^{WT/WT}LysMCre^{WT/WT}$, WT+: $Idh1^{WT/WT}LysMCre^{+/WT}$, HET-: $Idh1^{LSL/WT}LysMCre^{WT/WT}$, HET+ (LysM-KI): $Idh1^{LSL/WT}LysMCre^{+/WT}$).

Supplementary Table 2 | Altered hematopoietic subsets in LysM-KI mice.

FACS analyses of the indicated subsets among total nucleated BM cells from young (7-16 weeks) and older (42-46 weeks) control and LysM-KI mice. Numbers are the percentage of live nucleated BM cells. $p<0.05$, $**p<0.01$, as determined using the unpaired t test with Welch's correction.

Supplementary Table 3 | Primer sequences.

Sequences of oligonucleotide primers used for real-time RT-PCR profiling of HIF1α target gene mRNA expression. All PCR fragments span at least one exon-exon border.

Supplementary Methods

Histology

For histological analyses, mouse tissues were fixed in 10% buffered formalin. After fixation, femurs were decalcified in formic acid. Fixed tissues were embedded in paraffin, sectioned, and stained with hematoxylin and eosin (H&E) according to standard laboratory protocols.

FACS analyses

FACS analyses were performed according to standard protocols. In brief, mice were sacrificed, single-cell suspensions from bone marrow (BM) were generated by flushing out the bone marrow from tibia and femur in Iscove's modified Dulbecco's medium (IMDM2) supplemented with 2% heat-inactivated fetal bovine serum (HI-FBS). Single-cell suspensions of spleen were generated by mashing the spleen in IMDM2. Single-cell suspensions were passed through 70 μm cell strainers, centrifuged at 1350 rpm for 5 min, cell pellets were resuspended in ammonium-based red cell lysis buffer and incubated for 8 min at room temperature (Sigma). Nucleated cells were then washed in IMDM2, centrifuged and resuspended in FACS buffer [PBS without Mg^{2+} or Ca^{2+} ($PBS^{-/-}$) containing 2% HI-FBS and 5 mM EDTA, pH8.0]. Viable cell numbers were determined by cell counting after trypan blue staining.

For FACS analyses, cells were incubated in Fc block (BD Biosciences) and then immunostained with fluorophore-linked antibodies (Abs) from BD Biosciences Pharmingen and/or eBioscience. For immunostaining of Lineage-positive cells, a cocktail containing biotin-labeled primary Abs against CD5, B220, CD11b, 7-4, Gr-1, and Ter-119 (Miltenyi Biotec) was used, followed by staining with streptavidin-linked fluorophore-labeled secondary Abs. To detect total reactive oxygen species (ROS), immunostained cells were incubated with the live cell dye CM-H_2DCFDA (100 μM) at 37°C for 15 minutes followed by two washes in FACS buffer. FACS data were acquired on a BD FACSCantoII with

up to seven colors and analysed using FlowJo flow cytometry analysis software. For FACS analyses of differentiated cells, 10^4-10^5 events were acquired. For HSC and HPC analyses, up to $2.5x10^6$ events were acquired. Statistical analyses of FACS data were performed using the GraphPad Prism software.

Generation of CAGGS-KI ES cell clones

KI ES cells which had been generated while making the KI mouse as reported elsewhere[7] were grown up. They were then transiently transfected with pCAGGS-nlsCre in OPTI-MEM (Invitrogen) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. After transfection, cells were then trypsinized and plated at clonal density on gelatin-coated cell culture plates. Single colonies were picked and plated in duplicates into 96-well plates to test for puromycin sensitivity as the cre-mediated recombination of the LSL cassette results in loss of the puromycin-resistance cassette. Puromycin-sensitive clones expressing wild type Idh1 protein (clone F8) or Idh1-R132H protein (clone A12, H11, 1B3A, 1B3C, 1H5G and 1H5H; CAGGS-KI ES cell clones) and puromycin-resistant clones ($Idh1^{LSL/WT}$) were grown up, tested by PCR-genotyping, and then used for *in vitro* differentiation experiments.

OP9/OP9-DL1 ES cell *in vitro* differentiation system

Experiments including co-culture and differentiation induction of ES cells in the presence of OP9 or OP9-DL1 cells were carried out as described elsewhere[9].

LSK cell sorting

Single-cell suspensions from BM were generated as described above and cells were resuspended in MACS buffer ($PBS^{-/-}$ containing 0.5% bovine serum albumin and 5 mM EDTA pH8.0) and counted.

Lineage-negative cells were isolated using the Miltenyi lineage cell depletion kit according to the manufacturer's protocol. Cells were counted and immunostained as above with anti-Lineage-biotin/streptavidin-APC-Cy7, anti-cKit-PE-Cy7 and anti-Sca-1-APC Abs. Immunostained cells were sorted on a BD FACSAria cell sorter, collected in IMDM2 and then centrifuged at 1350 rpm for 5 min at 4°C. Cell pellets were shock frozen on dry ice and stored at -80°C until further processing (DNA extraction, mRNA extraction).

Colony forming cell (CFC) assays

Single-cell suspensions from BM or spleen from control and LysM-KI mice were generated as described above. Cells were then plated at a density of $1x10^4$/ml for BM and $1x10^5$/ml for spleen in triplicates in complete methylcellulose media containing stem cell factor, Il-3, Il-6, and erythropoietin (Stem Cell Technologies, M3434). Colonies were counted after 7 days at 37°C and 5% CO2. Numbers of colonies per $1x10^5$/ml for BM and $1x10^6$/ml for spleen were calculated.

Bone marrow competitive repopulation experiments

For competitive BM transfer, donor BM cells from control or LysM-KI mice ($CD45.2^+$), as well as competitor BM cells from WT mice ($CD45.1^+$), were isolated as described above and then washed in $PBS^{-/-}$ twice. Cell numbers were adjusted to $1x10^6$/ml. Lethally-irradiated (10.5Gy) recipient mice ($CD45.1^+$) were injected with $1x10^5$ donor BM cells plus $1x10^5$ competitor BM cells ($2x10^5$ total). For FACS analyses of peripheral blood, mice were bled from the tail vein and immunostaining was carried out as described above. Immunostained blood samples were fixed in FoxP3 fixation buffer to lyse red blood cells as per the manufacturer's protocol (eBioscience), washed in FACS buffer, and analyzed by FACS.

Generation of bone marrow-derived macrophages (BMDMs)

BMDMs were generated from single-cell BM suspensions (prepared as described above) of control or LysM-KI mice using a standard *in vitro* M-CSF-based differentiation protocol. In brief, $5x10^6$ cells were plated on a 10 cm tissue culture plate in 5ml of RPMI1640 media supplemented with 10% HI-FBS and 50 ng/ml M-CSF. On days 2, 3, and 4 an additional 1ml of the complete media was added to every tissue culture plate. Cells were harvested on day 5 of the *in vitro* differentiation procedure and washed in PBS containing Mg^{2+} and Ca^{2+} ($PBS^{+/+}$) before being used in experiments.

Real-time RT-PCR analysis

RNA was purified from LSK cells isolated as described above using TRIZOL (Invitrogen) according to the manufacturer's protocol. RNA was resuspended in water and quantified on a NanoDrop spectrophotometer prior to transcription into cDNA using the iScript cDNA synthesis kit (BioRad) according to the manufacturer's protocol. Primers used for RT-PCR are listed in Table S3. Real-time RT-PCR analyses were performed using Power SybrGreen (Applied Biosystems) and a 7900HT Fast-Real Time PCR system (Applied Biosystems). Ct values were normalized to the housekeeping gene Hprt (ΔCt) and then to the average of the ΔCt values of all samples ($\Delta\Delta$Ct). Relative mRNA expression levels were calculated ($2^{-\Delta\Delta Ct}$) and then normalized to the average of the relative control mRNA expression levels.

Immunoblot analysis

For immunoblot analysis of histone-3 methylation, equal numbers of BMDMs generated as described above were washed 1x with $PBS^{+/+}$ and lysed in 1x cell lysis buffer (20 mM Tris-HCl pH7.5, 150 mM

NaCl, 1 mM EDTA pH8.0, 1 mM EGTA pH8.0, 1% Triton X-100) supplemented with EDTA-free protease inhibitor cocktail (Roche). SDS (1%) was added to lysates and incubation was continued for another 30 min. Lysates were centrifuged, loading buffer was added to supernatants, and samples were subjected to electrophoresis on 12% Bis-Tris gels using MES running buffer (Invitrogen). Separated proteins were transferred onto nitrocellulose membranes (Invitrogen) and the membranes blocked with 5% milk in TBST for 1 hour. Membranes were washed 3x with TBST and incubated with 1:1000 dilution of primary Ab in 5% BSA in TBST overnight. Primary Abs used in this study recognized: total histone-3 (Abcam ab10799), histone-3 K4me3 (Millipore 07-473), histone-3 K9me3 (Abcam ab8898), histone-3 K27me3 (Millipore 07-449), histone-3 K36me3 (Abcam ab9050), or histone-3 K79me2 (Cell Signaling Technology 9757). Membranes were washed 3x in TBST and incubated with secondary HRP-linked Ab diluted in 5% milk in TBST. Membranes were washed 3x in TBST and bands were visualized by chemiluminescence (Millipore WBLUR0100).

Mass spectrometry

$CD11b^+$ cells were isolated from single-cell suspensions of BM or spleen cells (prepared as described above) using a CD11b-magnetic bead purification kit (Miltenyi Biotec). BMDMs were generated by *in vitro* differentiation as described above. Metabolites were extracted from and equivalent number of $CD11b^+$ cells or BMDMs using dry ice temperature 80% methanol and centrifugation. Supernatants were dried down under nitrogen gas, resuspended in 175 µl 50% methanol, and centrifuged 13000g for 5 min at 4°C. The supernatant was transferred to HPLC sample vials and 10 µl of each extract was analyzed by reverse-phase liquid chromatography (LC) with tributylamine as an ion pairing reagent. The LC was coupled to a triple-quadrupole mass spectrometer running in negative mode (Thermo Quantum Ultra). Specific chromatography conditions and mass spectrometry parameters were as described[20]. Peak heights of specific metabolites were measured by metabolite-specific multiple

reaction monitoring (MRM) scans as previously reported[21], with the exception of hydroxyglutarate, which was monitored using an MRM scan consisting of a 147 to 129 m/z transition and a collision energy of 13 eV. The mzrock mass spectrometry tool kit (http://code.google.com/p/mzrock/) was utilized for data analysis and visualization. To minimize the impact of interday instrument variation, extracted chromatogram peak heights for a given metabolite were normalized by the maximum sample value for that specific metabolite measured in the same experiment. These relative values were then scaled by 10,000.

DNA methylation analysis

High molecular weight genomic DNA (gDNA) was isolated from LSK cells sorted from BM cells of two control and three LysM-KI mice using the PureGene kit (Qiagen) according to the manufacturer's protocol. The gDNA was then characterized using a modified Reduced Representation Bisulfite Sequencing (RRBS)[22] approach called Enhanced Representation Bisulfite Sequencing (ERBS).

Generation of libraries

i) 2.5-10 ng gDNA were digested with 200 U of *Msp*I (New England Biolabs) in a 100 µl reaction for up to 16 hours at 37ºC. Digested DNA was isolated using a standard phenol chloroform extraction followed by ethanol precipitation and resuspended into 30 µl of 10 mM Tris pH8.0.

ii) End repair of digested DNA was performed in a 100 µl reaction using 15 U of T4 DNA polymerase, 5 U of Klenow DNA polymerase, 50 U of T4 Polynucleotide Kinase, 4 µl of premixed dNTPs each at 10mM using T4 DNA ligase buffer (all enzymes from New England Biolabs). The reaction was incubated at 20ºC for 30 minutes and products were purified using QIAquick PCR purification columns (Qiagen) and eluted into 32 µl of elution buffer.

iii) Adenylation was performed in a 50µl reaction using 15 U Klenow fragment 3' to 5' exo minus

(New England Biolabs), 10µl of dATP at 1 mM concentration in Klenow buffer. The reaction was incubated at 37ºC for 30 minutes and products were purified using MinElute PCR purification columns (Qiagen) and eluted into 10 µl of elution buffer.

iv) Adenylated DNA fragments were ligated with pre-annealed 5-methylcytosine-containing Illumina adapters in a 20 µl reaction using 2000 U T4 DNA ligase (New England Biolabs) and 1.2 µM final concentration of methylated adapters at 16ºC for a minimum of 16 hours. Products were isolated using MinElute columns (Qiagen) and eluted into 10 µl of elution buffer.

v) Library fragments of 150-250 bp and 250-400 bp were gel isolated from a 1.5% agarose gel using the QIAquick Gel Extraction kit (Qiagen) and eluted into 40 µl of elution buffer.

vi) Bisulfite treatment was performed using the EZ DNA Methylation Kit (Zymo Research) as per the manufacturer's recommended protocol with the following modification: incubation after the addition of CT conversion reagent was conducted in a thermocycler (Eppendorf MasterCycler) with the following conditions: 30 seconds at 95ºC followed by 15 minutes at 50ºC for 55 cycles. After clean up products were eluted into 40 µl of nuclease free water.

vii) PCR amplification for each library was performed using the FastStart High Fidelity PCR System (Roche) in a 200 µl reaction containing 2 µl of FastStart High Fidelity DNA Polymerase (Roche), 0.5 µM each of Illumina PCR primers PE1.0 and 2.0, 0.25 mM of each dNTP using buffer 2 as per manufacturer's recommendation and divided into four 50µl reactions. The thermocycler conditions were: 5 minutes at 94ºC, 18 cycles of 20 seconds at 94ºC, 30 seconds at 62ºC, 1 minutes at 72ºC, followed by 3 minutes at 72ºC. PCR products were isolated using AMPure XP beads as per manufacturer's recommended protocol (Agencort) and eluted into 50 µl of elution buffer.

viii) All amplified libraries underwent quality control steps including analysis with a Qubit 1.0 fluorometer and a Quant-iT dsDNA HS Assay Kit for quantitation (Invitrogen) as well as bioanalyzer

visualization (Agilent 2100 Bioanalyzer).

Primary data analysis

Alignment of Bisulfite treated reads and methylation calls: Reads were filtered from the adapter sequences using FAR software. Adapter sequence contamination usually occurs towards 3'ends of some reads. The adapter matching part of the read was removed if it aligned with the adapter sequence at least 6 base-pairs and had at most 0.2 mismatch error rate. Reads were aligned to whole genome using the bismark alignment[23] with a maximum of 2 mismatches in a directional manner and only uniquely aligning reads were retained. In order to call methylation score for a base position, we required that read bases aligning to that position have at least 20 phred quality score and the base position should have at least 10X coverage. Only CpG dinucleotides that satisfied these coverage and quality criteria were retained for subsequent analysis. Percentage of bisulfite converted Cs (representing unmethylated Cs) and non-converted Cs (representing methylated Cs) were recorded for each C position in a CpG context.

Downstream data analysis

CpG islands, refseq genes and repeat sequences for the MM9 genome were downloaded from the UCSC genome browser[24]. CpG shores were defined as 1000 bp flanking regions on upstream and downstream of a given CpG island. If a 1000 bp shore overlapped with another island, then the shore was clipped so that its last base falls before the start of the overlapping CpG island. Similarly, if shores were overlapping they were merged into a single shore. In addition, the genome was partitioned into intergenic, intronic, exonic and promoter regions. Promoter regions were defined as the 2 kb window centered around the transcription start sites (TSS) of refseq genes. We classified CpG dinucleotides as promoter, intronic, exonic or intergenic based on their overlap with these predefined regions. In addition, we classified CpG dinucleotides as CpG island or shore overlapping.

Calculating differential methylation

Methylation values for genomic regions (intergenic, intronic, exonic and promoters, CpG islands and island shores) between different samples were compared by taking the mean methylation percentage of CpG dinucleotides overlapping those regions. In order to calculate the correlation between different samples and generate the appropriate scatter plots we required that in any given region at least 3 CpG dinucleotides were covered by reads in both control and KI samples. Testing for differential methylation was performed at both the single base and predefined region levels, where regions were defined as promoters, intronic, exonic, CpG islands, island shores and 100 bp sliding windows over the genome. For the base level comparison, coverage on base positions for each sample was required while for the region level comparison we required at least 3 covered bases on all samples. The number of methylated and unmethylated Cs aligning to each base/region were counted and compared across samples. To determine significant differential methylation between two groups of samples we applied logistic regression and the likelihood ratio test was used. Observed p-values were adjusted with the q-value method[25].

Supplementary References

20. Munger J, *et al*. Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. Nat. Biotechnol. **26**, 1179-1186 (2008).

21. Bajad SU *et al*. Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. J. Chromatogr. A. **1125**, 76-88 (2006).

22. Meissner A *et al*. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. **33**, 5868-5877 (2005).

23. Krueger F & Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics **27**, 1571-1572 (2011).

24. Fujita PA *et al*. The UCSC Genome Browser database: update 2011. Nucleic Acids Res. **39**, D876-882 (2010).

25. Storey JD & Tibshirani R. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U S A. **100**, 9440-9445 (2003).