Supplemental Experimental Procedures

Meis1a/Hoxa9 leukemic transformation assay

The day before transduction, 1 million c-Kit$^+$ cells were pre-stimulated with IMDM 10% FBS containing 40 ng/mL SCF, 20 ng/mL IL-3, and 20 ng/mL IL-6. Cells were transduced with retroviruses encoding Meis1a/Hoxa9 using retronectin-coated plates. 6 hours after the first transduction, cells were transduced for a second time in a new retronectin-coated plate containing retroviruses. Next day, medium containing antibiotics geneticin 418 (G418) (1 mg/ml) and puromicyn (1.5 µg/ml) was added to select transduced Meis1a/Hoxa9 cells. 72 hours later, 5,000 live cells were plated in CFC1 in M3231 (STEMCELL Technologies) and replated every 6 days for up to 3 rounds of CFC. Pre-LSCs were harvested and enriched for c-kit from CFC3 plates and transplanted into lethally-irradiated primary recipients. LSCs from moribund mice showing symptoms of AML were sorted and transplanted into lethally-irradiated secondary recipients.

Transplantation assays

For cell-autonomous transplantation, 500,000 CD45.2$^+$ unfractionated BM cells from *Gata2* mice, together with 500,000 unfractionated CD45.1$^+$ BM support cells, were injected in the tail vein of lethally-irradiated CD45.1$^+$ mice. For secondary transplantation, 500,000 CD45.2$^+$ BM cells were transplanted alongside 500,000 unfractionated CD45.1$^+$ BM support cells into lethally-irradiated CD45.1$^+$ mice. For niche studies, 500,000 unfractionated BM cells from CD45.1$^+$ mice were transplanted into the tail vein of lethally-irradiated *Gata2$^{fl/fl}$;Mx1-Cre* mice. For leukemia assays, 250,000 pre-LSCs were injected into lethally-irradiated CD45.1$^+$ mice alongside 200,000 unfractionated CD45.1$^+$ BM support cells. For secondary transplants of LSCs, 10,000 c-kit$^+$ CD45.2$^+$ BM cells from leukemic mice were injected into lethally-irradiated CD45.1$^+$ mice together with 200,000 unfractionated CD45.1$^+$ BM support cells.

Flow cytometry

Extracellular staining was carried out in PBS 2% FBS with an incubation time of 25’ in the dark at 4°C. To exclude dead cells from the analysis, 1:10,000 µL

of DAPI (20 µg/mL) was added prior to acquisition. For the annexin V assay, cells were incubated in annexin binding buffer (BioLegend) containing annexin V (BioLegend) for 25' in the dark at RT. Annexin binding buffer was added to stop the reaction, and 2 µL of diamidino-2-phenylindole (DAPI, 20 µg/ml)(Molecular probes) was added prior to acquisition. For intracellular flow cytometry, cells were fixed with 1% methanol-free PFA (ThermoFisher) for 15' on ice and then permeabilized in PBS 0.1% Saponin (Sigma) 2% BSA (ThermoFisher) for 15' on ice and stained for 1 hour on ice with a fluorescence-labelled antibody diluted in the permeabilisation buffer.

Generation of lentiviruses

Lentiviral vectors were diluted in water, and then mixed with calcium chloride (Sigma). This mixture was added dropwise to 2x HEPES buffered saline (HBS; Sigma) and after 15' incubation, added to the media of HEK293T cells at 70% confluency in a 10-cm dish. Supernatant containing lentiviruses was collected at 24 hours after transfection of HEK293T cells and filtered in a 0.45 µm filter (Sigma) prior to use or storage at -80C.

Cord blood $CD34^+$ HSPCs isolation and culture

Highly enriched human $CD34^+$ cells (>90%) were derived from CB mononuclear cells using MiniMACS (Miltenyi Biotech, UK) using manufacturer's instructions. CB derived $CD34^+$ HSPC purity was confirmed by two-color flow cytometric analysis using anti-CD45 PerCP (Biolegend, UK) and CD34-PE (BD Biosciences, UK). $CD34^+$ cells were pre-stimulated overnight in StemSpan (STEMCELL Technologies) supplemented with 100 ng/mL of SCF, 100 ng/mL of FLT3L, 10 ng/mL IL6, 10 ng/mL IL3, and 25ng/mL G-CSF, 1% Pen/Strep, 10 uM HEPES. Lentiviruses containing shRNA against human *GATA2* and scrambled vector (Genecopoeia) were bound to retronectin-coated plates by centrifugation at 2,100 xG, at RT, for 2:30h. $CD34^+$ cells were then added to viruses-retronectin-coated plates and incubated for 4 days. GFP^+ cells were sorted and plated into CFC assays in H4434 for 14 days. For LTC-IC assays, GFP^+ cells were plated onto irradiated MS5-stromal cells for 5 weeks, with half-media changes every 7 days. Media was then replaced with enriched H4435 and colonies from LTC-IC were

enumerated at day 14 and LTC-IC frequency was calculated using ELDA software.

Lentiviral-Cre mediated deletion

Meis1a/Hoxa9 LSK and CMP derived pre-LSCs from *Gata2*$^{fl/fl}$ *or Gata2*$^{+/+}$ mice were transduced with lentiviruses encoding Cre and *iVenus* reporter marker using the retronectin-coated system as explained above. Transduced *iVenus*+ cells were plated in colony assays for up to three rounds, and were immunophenotypically assessed using Mac1, Gr1, and c-kit.

RNA-sequencing and Bioinformatics

RNA from CD45.2^{+}c-kit^{+} LSCs was extracted with the RNAeasy micro kit (Qiagen). Total RNA quality and quantity was assessed using Agilent 2100 Bioanalyser and a RNA Nano 6000 kit (Agilent Technologies). 30-260ng of Total RNA with a RIN value >8 was depleted of ribosomal RNA using the NEBNext® rRNA Depletion Kit (Human/Mouse/Rat), (New England BioLabs, NEB) and the sequencing libraries were prepared using the NEB® Ultra™ II Directional RNA Library Prep Kit for Illumina® (NEB). The steps included RNA fragmentation and priming, 1st strand cDNA synthesis, 2nd strand cDNA synthesis, adenylation of 3' ends, adapter ligation, PCR amplification (11-cycles) and validation. The manufacturer's instructions were followed. The libraries were validated using the Agilent 2100 Bioanalyser and a high-sensitivity kit (Agilent Technologies) to ascertain the insert size, and the Qubit® (Life Technologies) was used to perform the fluorometric quantitation. Following validation, the libraries were normalized to 10nM, pooled together and clustered on the cBot™2 following the manufacturer's recommendations. The pool was then sequenced using a 75-base paired-end (2x75bp PE) dual index read format on the Illumina® HiSeq4000 according to the manufacturer's instructions.

To remove adapter sequencer and poor quality ends of reads, trimming was performed using Trim Galore, a wrapper tool which uses cutadapt and FastQC to trim and perform QC on the trimmed. Trimming was performed in paired-end mode. Trimmed reads were mapped against the GRCm38.p6

mouse genome using STAR. STAR was run with the MultimapNMax=1 flag, meaning reads mapping to more than 1 location were considered unmapped. Expression counts for both exons and transcripts were calculated, using Subread feature Counts Version 1.6.2 (Liao et al., 2014), a program for assigning sequence reads to genomic features. To define the exon and transcript locations (raw reads calculation), the GENCODE M18 gene model was used. Read summarization (counting) was generated for paired end read fragments - it was summarized at exon level and then grouped at transcript level. To provide stringent and robust data, reads overlapping more than one feature were excluded from the count summary (according to authors recommendations (Liao et al., 2014).

Differentially expressed genes were identified using a DEseq2 Bioconductor package (Love et al., 2014) analysis within the R environment for statistical computing software on normalised count data. For multiple testing and false discovery issues, the generated p-values were corrected using the FDR method (Benjamini and Hochberg, 1995).

The Gene ontology over-representation analyses (GO ORA) for biological processes (BP) were undertaken using the AmiGO website service. Gene Set Enrichment Analysis (Subramanian et al., 2005) was applied against categories and mouse gene mappings obtained from the Gene Ontology Consortium using data ranked by a scoring metric generated by multiplying the sign of fold change by its inverse p-value". Heat maps of differentially significant genes were generated using the Morpheus online software tool (Broad Institute).

For Ingenuity Pathway Analysis, RNA-seq datasets with FDR <0.05 were investigated for the common canonical pathways between HSCs and LSCs using the Ingenuity Pathway Analysis software (Qiagen-Bioinformatics). Enriched pathways were selected by Fischer's Exact Test.

RT–quantitative PCR

Gene expression analyses were performed using Taqman probes for mouse

Gata2 [Mm00492301_m1] or human *GATA2* [Hs00231119_m1]. Differences in input cDNA were normalized with mouse *Hprt* [Mm03024075_m1] or human *GAPDH* (Hs02758991_g1) mRNA expression levels were determined by the $2^{-\Delta\Delta CT}$ method of relative quantification (Schmittgen and Livak, 2008).

Western Blot

BM cells were washed with cold PBS and re-suspended in lysis buffer containing 10x RIPA buffer (CST), protease inhibitor cocktail (Merck), phosphatase inhibitor cocktail 2 (Merck), phosphatase inhibitor cocktail 3 (Merck) and 1.5 mM Phenyl methane sulfonyl fluoride (PMSF; Merck) in ddH_2O. Samples were separated in 4-15% Mini-protean pre-cast gel (Bio-rad) and transferred to a 0.2 µm PVDF membrane with trans-blot turbo (Bio-rad). Membrane was blocked in 5% milk and immunobloted with Rb anti-GATA2 ab (Abcam ab109241) and mouse anti-GAPDH (Merck CB1001) overnight at 4 degrees. Membranes were incubated with HRP-conjugated anti-rabbit (abcam ab97051) or anti-mouse IgG (Sigma A4416) and detected with enhanced chemiluminescent (ECL) reagents (WBLUF0100, Merck Millipore).

Statistical analyses

FACS data were analyzed using FlowJo 10.0.8 (Tree Star, Inc) software and all results graphed using GraphPad Prism 7 (GraphPad Software Inc, CA). Data are presented as mean ± standard error of mean (SEM). Significant differences were calculated using Mann–Whitney U test or one-way ANOVA (*$p<0.05$, **$p<0.01$, ***$p<0.001$, **** $p<0.0001$). Kaplan-Meier survival curve statistics were determined using the log-rank (Mantel-Cox) test.

Supplementary references

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. J.R.Stat.Soc.Ser.B Stat.Methodol. *57*, 289–300.

Liao, Q., Shen, J., Liu, J., Sun, X., Zhao, G., Chang, Y., Xu, L., Li, X., Zhao, Y., Zheng, H., et al. (2014). Genome-wide identification and functional annotation of Plasmodium falciparum long noncoding RNAs from RNA-seq data. Parasitol. Res. *113*, 1269–1281.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) methodSchmittgen, T.D., Livak, K.J., 2008. Analyzing real-time PCR data by the comparative C(T) method. Nat. Protoc. 3, 1101–1108.. Nat. Protoc. *3*, 1101–1108.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. *102*, 15545–15550.

Table S1. List of antibodies used in study.

Antigen	Clone	Reactive specie	Fluorochome	Concentration	Company
CD3	17A2	Mouse	Biotin, APC, FITC	1/1000	BioLegend
CD4	GK1.5	Mouse	Biotin, PE, PE-Cy7	1/1000	BioLegend
CD8a	53-6.7	Mouse	Biotin, PE, APC-Cy7	1/1000	BioLegend
CD11b	M1/70	Mouse/Human	Biotin, PE, APC	1/1000	BioLegend
CD16/32	93	Mouse	PE-Cy7	1/25	BioLegend
CD34	RAM34	Mouse	FITC, APC	1/25	eBiosciences
CD45.1	A20	Mouse	BV510, APC	1/500	BioLegend
CD45.2	104	Mouse	Pacific blue, BV510, PE	1/500	BioLegend
CD45R/B220	RA3-6B2	Mouse/Human	Biotin, FITC, APC	1/1000	BioLegend
CD48	HM48-1	Mouse	FITC	1/50	BioLegend
CD117	2B8	Mouse	PE, APC	1/100	BioLegend
CD150	TC15-12F 12.2	Mouse	PE-Cy7	1/100	BioLegend
Gr1	RB6-8C5	Mouse	Biotin, FITC, PE-Cy7	1/1000	BioLegend
Ter119	Ter119	Mouse	Biotin, APC-Cy7	1/1000	BioLegend
Sca-1	D7	Mouse	PE, APC-Cy7	1/25	BioLegend
Streptavidin		Mouse/Human	Pacific blue, PerCP	1/100	BioLegend, eBiosciences
BCL2	REA356	Mouse	PE	1/100	Miltenyi Biotec
REA Control-PE	REA293		PE	1/100	Miltenyi Biotec
Fc block	93	Mouse		1/100	BioLegend
BCL-XL	54H6	Mouse	PE	1/100	Cell signaling
P53	1C12	Mouse	APC	1/100	Cell Signaling
pSer473-Akt	REA359	Mouse	PE	1/100	Miltenyi Biotec
Mouse IgG1 Isotype Ctrl	MOPC-21		APC	1/100	Cell signaling

Supplementary Figure 1 – linked to figure 1

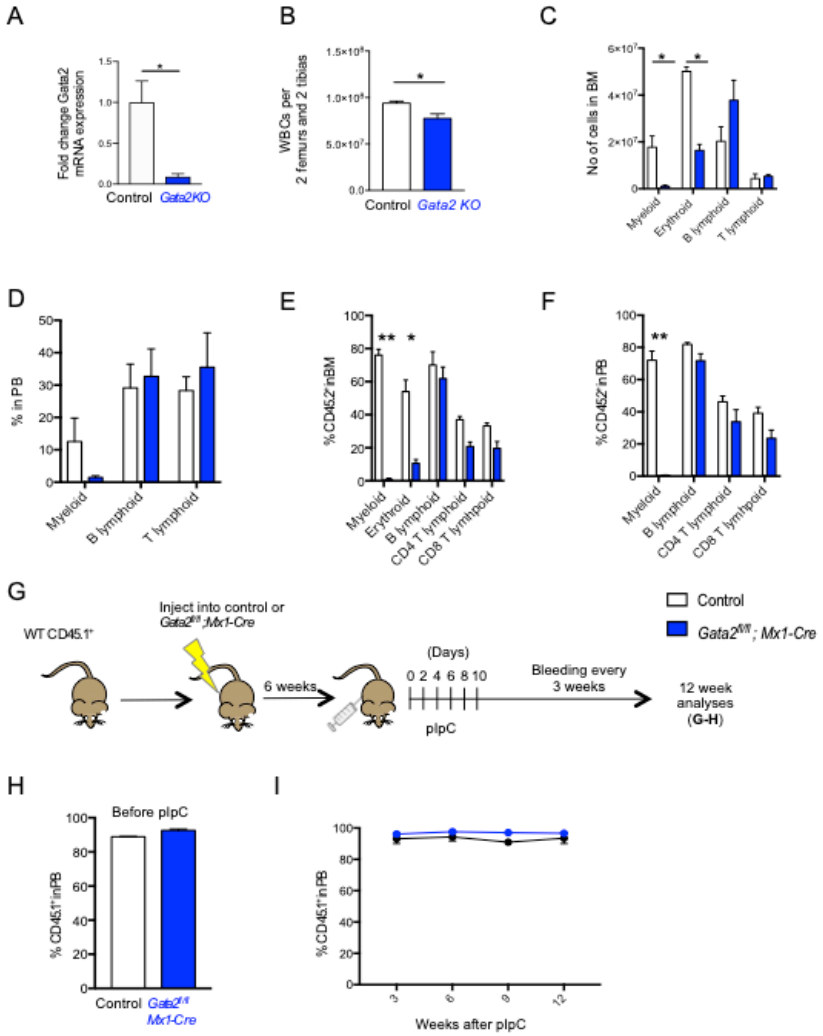

Figure S1 (linked to Figure 1). *Gata2* is essential for cell autonomous adult HSC maintenance. (A) *Gata2* mRNA levels in BM cells from control and *Gata2 KO* at day 24 after pIpC administration. Hprt was used as endogenous control for mRNA quantification (n = 4). (B) WBC count from 2 femurs and 2 tibias at day 24 after pIpC administration (n = 4). (C) Number of myeloid ($Gr1^+Mac1^+$), erythroid ($Ter119^+$), B lymphoid ($B220^+$), and T lymphoid ($CD3^+$) cells in BM at day 24 (n = 4). Statistical analyses: one-way ANOVA. (D) Frequency of myeloid, B lymphoid, and T lymphoid cells in PB at day 24 (n = 4). Statistical analyses: one-way ANOVA. (E) Frequency of $CD45.2^+$ cells of the indicated populations in BM at day 24 in cell-autonomous setting (n = 4). Statistical analyses: one-way ANOVA. (F) Frequency of $CD45.2^+$ cells of the indicated populations in PB at day 24 in cell-autonomous setting (n = 4). Statistical analyses: one-way ANOVA. (G) BM cells from $CD45.1^+$ WT mice were transplanted into lethally-irradiated untreated *Gata2*$^{fl/fl}$*;Mx1-Cre*$^+$ and control recipient mice. Six weeks later, the mice received six doses of pIpC. The mice were monitored for up to twelve weeks after last pIpC dose to analyze the impact of the BM microenvironment. (H) Frequency of $CD45.1^+$ cells in PB at week 6, before pIpC (n = 2-5). (I) Frequency of $CD45.1^+$ cells in PB at the indicated time points after pIpC administration (n = 2 control and n = 5 *Gata2 KO* recipients). Statistical analyses: one-way ANOVA. Data are mean ± SEM. Statistical analyses: Mann–Whitney U test unless otherwise indicated

Figure S2 – linked to Figure 3

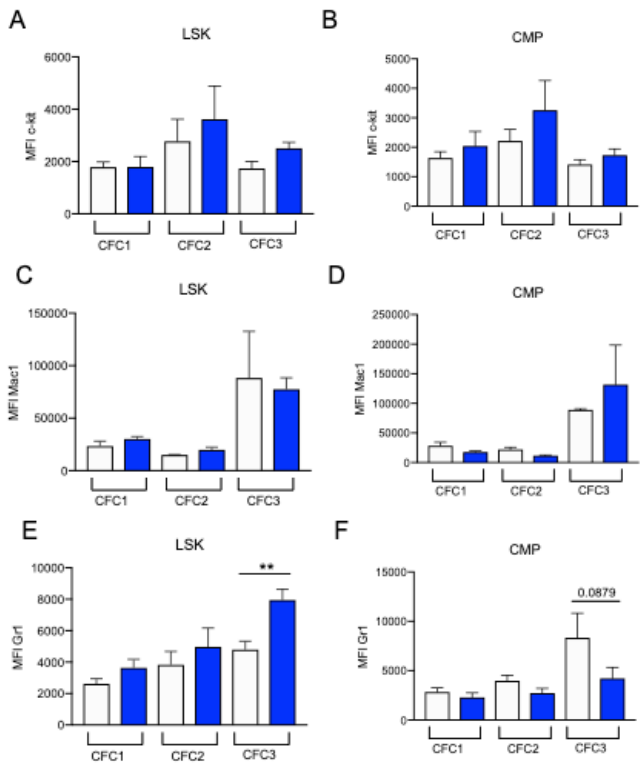

Figure S2 (linked to Figure 3). Cell of origin for GATA2 mediated transformation of *Meis1a/Hoxa9* AML (A-B) MFI of c-kit at CFC1-to-CFC3 from (A) LSK and (B) CMP derived *Meis1a/Hoxa9 Gata2*$^{fl/fl}$ and *Gata2*$^{+/+}$ transduced with lentiviral-Cre (n = 3). (C-D) MFI of Mac1 at CFC1-to-CFC3 from (C) LSK and (D) CMP derived *Meis1a/Hoxa9 Gata2*$^{fl/fl}$ and *Gata2*$^{+/+}$ transduced with lentiviral-Cre (n = 3). (E-F) MFI of Gr1 at CFC1-to-CFC3 from (E) LSK and (F) CMP derived *Meis1a/Hoxa9 Gata2*$^{fl/fl}$ and *Gata2*$^{+/+}$ transduced with lentiviral-Cre (n = 3). White bar in graph depicts *Gata2*$^{+/+}$ and Cre and blue bar depicts *Gata2*$^{fl/fl}$ and Cre. Statistical analyses: two-way ANOVA.

Figure S3 – linked to Figure 4 and Figure 5

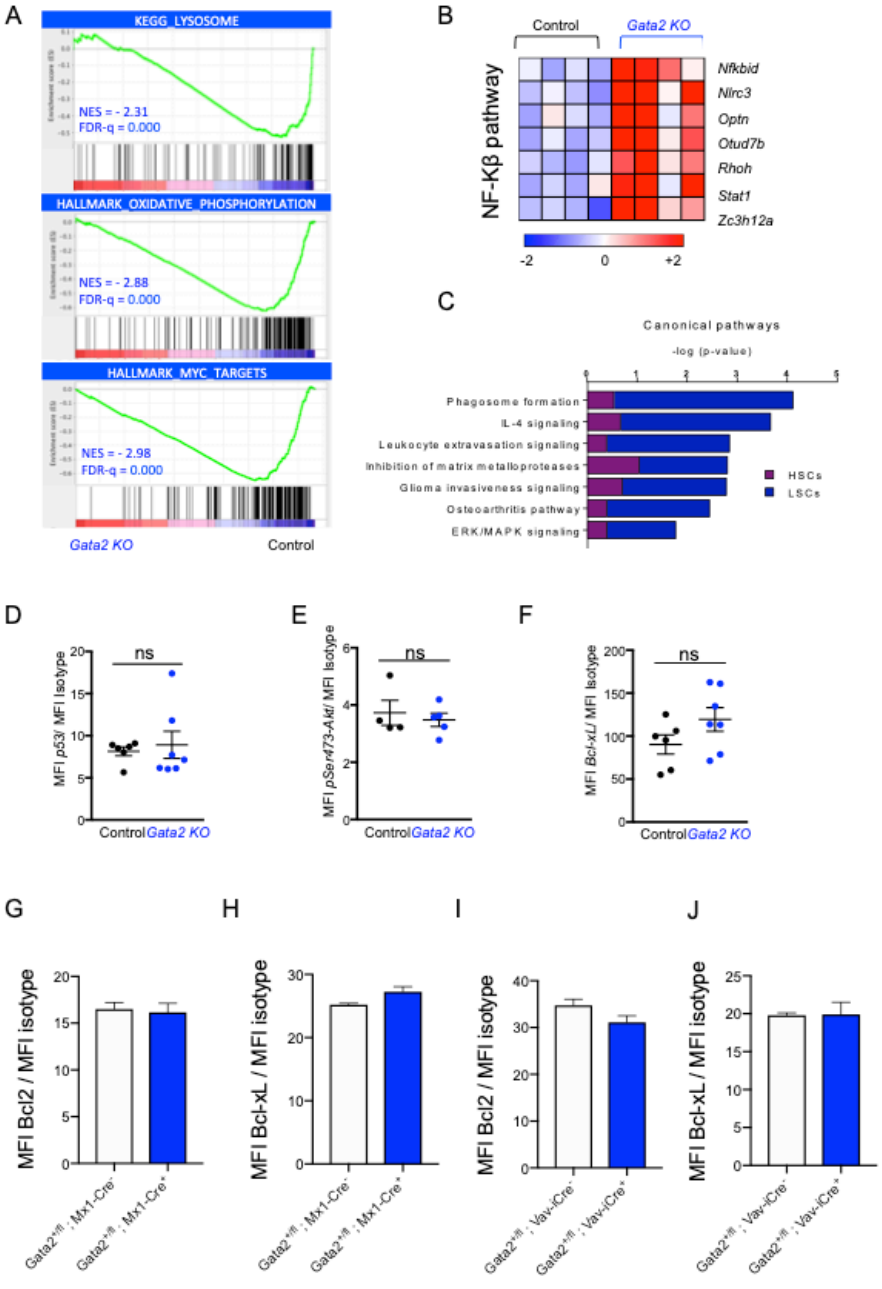

Figure S3 (linked to Figure 4 and Figure 5). Transcriptional signature of *Gata2 KO* LSCs shows deregulated apoptotic and myeloid differentiation pathways. (A) GSEA plots showing lysosome, oxidative phosphorylation signature and c-Myc targets gene sets in control LSCs compared to *Gata2 KO* LSCs. (B) Heat map reflecting expression levels of NF-Kβ pathway genes that are significantly differentially expressed between control and *Gata2 KO* LSCs. (C) Biological processes commonly shared between *Gata2 KO* HSC and LSC target genes. Analyses was performed using the Ingenuity Pathway Analysis software. Data is shown as –log10 (p-value). (D-F) MFI of (D) *p53,* (E) *pSer473-Akt,* and (F) *Bcl-xL* in LSCs in relation to the MFI of isotype control (n = 6-7 ; [E] n = 4). (G and H) MFI of (G) BCL2 and (H) BCL-XL in HSCs from control and *Gata2*$^{+/fl}$*; Mx1-Cre*$^{+}$ mice 24 days after the first pIpC injection (n = 3). (I and J) MFI of (I) BCL2 and (J) BCL-XL in HSCs from 8-week old control and *Gata2*$^{+/fl}$ *; Vav-iCre*$^{+}$ mice (n = 4). Data are mean ± SEM. Statistical analyses: Mann–Whitney U test.