## Supplemental methods

### Mouse complete blood cell count (CBC) analysis

Mouse peripheral blood was collected via retro-orbital bleeding into EDTA-coated tubes, and complete blood counts and immature platelet fraction (IPF) were obtained using a Sysmex XN-1000V multispecies hematology analyzer.

### Histopathology

Mice were euthanized using a $CO_2$ chamber, and organs were processed immediately thereafter. Spleen and tibias were dissected and fixed in 10% neutral buffered formalin. Tibias were decalcified in 10% EDTA, pH 7.4 for one week with regular changes of the solution. All organs were dehydrated in graded ethanol and embedded in paraffin using standard procedures. 4 µm sections were stained with hematoxylin and eosin (H&E) or with Wilder's method for reticulum staining. Bone marrow and spleen sections were evaluated for megakaryocyte numbers, megakaryocyte morphology, and myelofibrosis emperipolesis. Megakaryocytic emperipolesis was quantified as the percentage of megakaryocytes containing at least one intact intracytoplasmic hematopoietic cell, after evaluation of 200 megakaryocytes in well-preserved marrow areas. Emperipolesis index (%) = (megakaryocytes with emperipolesis / total megakaryocytes counted) × 100. For all histopathological analyses, the hematopathologists were blinded to the mouse genotypes. Images were acquired on a Olympus BX40 microscope using cellSens imaging software.

### Immunohistochemistry

Tissue slides were deparaffinized in Xylene, rehydrated in graded ethanol, and transferred to distilled water. Antigen retrieval was performed by simmering in citrate

buffer, pH 6, for 5 minutes. Slides were washed and blocked in 5% goat serum in TBS-T, followed by incubation with primary antibody (Phospho-Histone H3 (Ser10), D7N8E, Cell Signaling # 53348) overnight at 4 degrees. Slides were washed in TBS-T and incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies for 30 min. Nuclei were counterstained with hematoxylin. Imaging was performed on a Lionheart FX automated microscope.

Bone Marrow Flow Cytometry

For immunophenotyping of hematopoietic stem and progenitor cells (HSPC), bone marrow cells were isolated by crushing bones (femur, tibia and spine) using mortar and pestle. For the analysis of native megakaryocytes, bone marrow was spun out of femur and tibia by placing cut-end bones in nested tubes and centrifugation at 4000 xg for 30 seconds. Erythrocytes were lysed for 5 min on ice (BD Pharm Lyse, #555899), and cells were serially passed through 100, 70 and 40 µm cell strainers. Cells were stained in PBS with 3% fetal bovine serum (FBS) and 1% penicillin/streptomycin (P/S) for 30 min using the following antibodies: BD Biosciences B220-APC/Cy7 (#552094, clone RA3-6B2) and CD3-PE (#553240, clone 500A2), BioLegend Ly6C-APC (#128015, clone HK1.4), CD11b-Pacific Blue (#101224, clone M1/70), Ly6G-PerCP/Cy5.5 (#127615, clone 1A8) and ZombieRed (#423110). Spectral flow cytometry was performed using a Cytek Aurora spectral cytometer (Cytek Biosciences). Post-acquisition analysis of data was performed using FlowJo software V10. Gating strategies to discriminate cell populations are displayed in Supplemental Figure 1B.

Analysis of megakaryocyte polyploidization

Tibias and femurs were cut and placed in a punctured nested 0.5 mL tube in a 1.5 mL tube. The bones were spun at 2500 xg for 1 minute to collect bone marrow cells. Total bone marrow cells were passed through a 100 µm strainer and red blood cells were lysed for 5 minutes on ice. The cells were washed and stained with Biolegend CD41-APC antibody (#133914, clone MWReg30) for 30 min. After washing, the cells were fixed in ice-cold 70% ethanol/PBS and fix for 30 min at 4°C. After another washing step, the cells were resuspended in 3% FBS, 1% P/S in PBS containing 0.1 µg/mL DAPI (BD Pharmingen), and the samples were analyzed using a Cytek Aurora spectral cytometer (Cytek Biosciences).

Megakaryocyte Differentiation

Bone marrow cells were isolated by crushing femur, tibia and spine and processed as described above. Total bone marrow cells were then cultured in DMEM (Gibco #11-965-092) supplemented with 10% FBS, 1% P/S, and 50 ng/mL TPO (Peprotech, #315-14). After 5 days of differentiation, cells were harvested and centrifuged for 5 min at 200 x*g*. The pellet was resuspended in 1 mL PBS and subjected to BSA gradients as previously described [1] to remove smaller, incompletely differentiated cells and enrich mature megakaryocytes for RNAseq.

RNAseq in Mature Megakaryocytes

RNA from ex vivo differentiated mature megakaryocytes (n=3-4 mice per group) was isolated using Qiagen RNeasy™ kit according to the manufacturer's protocol. The RNA quality was assessed using an Agilent 2100 Bioanalyzer. Libraries were prepared using Roche Kapa mRNA HyperPrep strand specific sample preparation kits from 200 ng of

purified total RNA according to the manufacturer's protocol on a Beckman Coulter Biomek i7. The finished dsDNA libraries were quantified by Qubit fluorometer and Agilent TapeStation 4200. Uniquely dual indexed libraries were pooled in an equimolar ratio and shallowly sequenced on an Illumina MiSeq to further evaluate library quality and pool balance. The final pool was sequenced on an Illumina NovaSeq 6000 targeting 40 million 150bp read pairs per library at the Dana-Farber Cancer Institute Molecular Biology Core Facilities. Sequenced reads were aligned to the UCSC mm10 reference genome assembly and gene counts were quantified using STAR (v2.7.3a) [2]. Differential gene expression testing was performed by DESeq2 (v1.22.1) [3]. RNAseq analysis was performed using the VIPER snakemake pipeline [4].

Global Chromatin Profiling in Mature Megakaryocytes using nanoLC-MS/MS

Mouse megakaryocytes (n=4-5 mice per group) were differentiated and purified as described above. Cells were then spun for 5 min at 200 xg, and pellets were frozen in liquid nitrogen until further processing. Histones were extracted following an established protocol [5]. Samples (each containing 10 µg of histones) were propionylated, desalted and digested overnight with trypsin. A second round of propionylation occurred at the peptide level and peptides were desalted using C18 Sep-Pak cartridges (Waters). A reference mixture of isotopically labeled synthetic peptides (heavies) for histones H3 and H4 was added to each sample prior to MS analysis. Peptides were separated on a C18 column (EASY-nLC 1000, Thermo Scientific) and analyzed on a Q ExactiveTM Plus (Thermo Scientific) using a PRM method targeting both endogenous and heavy peptides. Peak areas were extracted and integrated in Skyline [6]. The ratios of endogenous to heavy peak

areas were log2 transformed and normalized to unmodified regions (“Norm” peptide) of histone H3 (41-49) or H4 (68-78) respectively and samples were raw median normalized at the probe level. Histone posttranslational modification stoichiometry at the single amino acid level was calculated as previously described [7].

CUT&RUN

Bone marrow cells of mice 8 weeks after pIpC injection were isolated by crushing femur, tibia and spine and processed as described above. Cells were then cryopreserved in liquid nitrogen and thawed at the day of the assay. $cKit^+$ cells were incubated with Miltenyi anti-CD117 microbeads (#130-091-224) and purified using MACS™ Miltenyi columns as described above, and $3x10^5$ $cKit^+$ cells were used for each reaction. CUT&RUN was performed according to CUTANA™ CUT&RUN version 2 kit instructions, and antibodies used were Rabbit IgG Negative Control Antibody (CUTANA™ #13-0042), H3K4me3 (CUTANA™ #13-0041) H3K27me3 (Cell Signaling #9733, clone C36B11). ChIPseq libraries were prepared using IDT DNAseq library prep reagents on a Beckman Coulter Biomek i7 liquid handling platform from approximately 1ng of DNA according to manufacturer’s protocol and 14 cycles of PCR amplification. Finished sequencing libraries were quantified by Qubit fluorometer and Agilent TapeStation 2200. Library pooling and indexing was evaluated with shallow sequencing on an Illumina MiSeq. Subsequently, libraries were sequenced on an Illumina NovaSeq6000 targeting 20 million 150bp read pairs by the Molecular Biology Core Facilities at Dana-Farber Cancer Institute. Sequenced read were aligned to the UCSC mm10 reference genome assembly using

BWA (v0.7.17) [8] and peaking calling performed using MACS2 (v2.2.6) [9] as part of the CHIPS software package. IgG sample data were used as input controls.

CUT&RUN analysis

All samples were processed through the computational pipeline developed at the Dana-Farber Cancer Institute Center for Functional Cancer Epigenetics (CFCE) using primarily open-source programs [8,10,11]. Sequence tags were aligned with Burrows-Wheeler Aligner (BWA) [8] to build hg19 and uniquely mapped, non-redundant reads were retained. These reads were used to generate binding sites with Model-Based Analysis of ChIP-Seq 2 (MACS v2.1.1.20160309), with a q-value (FDR) threshold of 0.01 [9]. We evaluated multiple quality control criteria based on alignment information and peak quality: (i) sequence quality score; (ii) uniquely mappable reads (reads that can only map to one location in the genome); (iii) uniquely mappable locations (locations that can only be mapped by at least one read); (iv) peak overlap with Velcro regions, a comprehensive set of locations – also called consensus signal artifact regions – in the genome that have anomalous, unstructured high signal or read counts in next-generation sequencing experiments independent of cell line and of type of experiment; (v) number of total peaks (the minimum required was 10,000); (vi) high-confidence peaks (the number of peaks that are tenfold enriched over background); (vii) percentage overlap with known DHS sites derived from the ENCODE Project (the minimum required to meet the threshold was 80%); and (viii) peak conservation (a measure of sequence similarity across species based on the hypothesis that conserved sequences are more likely to be functional). For differential binding analyses, peaks from all samples were merged to create a union set of sites for

each transcription factor and histone mark using bedops [12]. Sample-sample correlation and differential peaks analysis were performed by the CoBRA pipeline [11]. Read densities were calculated for each peak for each sample and used for the comparison of cistromes across samples. Sample similarity was determined by hierarchical clustering using the spearman correlation between samples. Differential peaks were identified by DEseq2 with adjusted p-value ≤ 0.05. A total number of reads in each sample was applied to the size factor in DEseq2, which can normalize the sequencing depth between samples. Peaks from each group were used for motif analysis by the motif search findMotifsGenome.pl in HOMER (v3.0.0) [13], with cutoff q-value ≤ 1e-10.

scRNAseq with feature barcoding of mouse HSPCs

Bone marrow cells of thirty-week-old primary mice (n=3-4 mice per genotype) were isolated by crushing femur, tibia and spine and processed as described above. Lineage-committed cells were depleted by incubation of the cells with biotin-conjugated antibodies (eBioscience CD3e #553060 clone 145-2C11, CD5 #553018 clone 53-7.3, CD11b #553309 clone M1/70, B220 #553086 clone RA3-6B2, Ly6C/G #553124 clone RB6-8C5, Ter-119 #553672 clone TRE-119) and negative selection though Dynabeads™ Biotin Binder (Invitrogen #11-047) according to the manufacturer's protocol. After lineage-depletion, the cells were resuspended in 45 µL Cell Staining Buffer (Biolegend). Cells were blocked using 5 µL of TruStain FcX™ anti-mouse Cd16/32 reagent for 10 min at 4°C. After washing, the cells were incubated with Biolegend cKit-APC (#105812, clone 2B8), Lineage-Pacific Blue (eBioscience CD5 #48-0051-82 clone 53-7.3, Biolegend CD3 #100214 clone 17A2, CD11b #101224 clone M1/70, B220 #103227 clone RA3-6B2,

Ly6C/G #108430 clone RB6-8C5 and Ter-119 #116232 clone TER-119) and, for CITEseq analysis, TotalSeq™-B antibodies from Biolegend (CD16/32 #B0109 clone 93, CD34 #B0857 clone HM34, CD41 #B0443 clone MWReg30, CD48 #B0429 clone HM48-1, CD71 #B0441 clone RI7217, CD105 #B0812 clone MJ7/18, CD115 #B0105 clone AFS98, CD135 #B0098 clone A2F10, CD150 #B0203 clone TC15-12F12.2, Epcr #B0439 clone RCR-16, Sca1 #B0130 clone D7) for 30 minutes at 4°C. After washing, lineage$^{-}$, cKit$^{high}$ cells (defined as the upper 50% of the cKit$^{+}$ population) were sorted using a BD FACS-Aria cell sorter (BD Biosciences). The cell concentration was adjusted to 1000 cells/µL before library preparation. The cells in mastermix were then added to the 10x chip G along with 10x gel beads and partitioning oil. The chip was run on the 10x Chromium Controller to create an emulsion of single cells, RT mastermix and a barcoded gel bead. After emulsions were created, samples were immediately put on reverse transcription thermalcycler protocol set at 53°C for 45 minutes, 85°C for 5 minutes before holding at 4°C. The emulsion was then broken, and debris was removed via a Silane bead cleanup. The purified cDNA was amplified with the following PCR conditions: Initial denaturation at 98°C for 3 minutes, followed by 12 cycles of denaturation at 98°C for 15 seconds, annealing at 63°C for 20 seconds, and extension at 72°C for 1 minute. A 0.6X SPRI (Beckman Coulter #B23317) cleanup was performed before quantification on a Qubit. The cDNA library construction was run on an entirely automated workflow. The cDNA was fragmented and repaired with the addition of an A-tail at 32°C for 5 minutes then 65°C for 30 minutes. A 0.8X double sided SPRI cleanup removed undesired fragments. Adapters were ligated onto the DNA at 20°C for 15 minutes followed by a 0.8X SPRI. 10x dual indexed barcodes were added before a PCR reaction with the following conditions:

an initial denaturation at 98°C for 45 seconds, followed by 12 cycles of denaturation at 98°C for 20 seconds, annealing at 54°C for 30 seconds, and extension at 72°C for 20 seconds. These libraries underwent a final 0.8X double sided SPRI cleanup for size selection. Following PCR cleanup, sample quality was assessed using Quant-iT PicoGreen and an Agilent TapeStation to confirm library size. Samples were pooled to accommodate the desired reads per cell. For Illumina Sequencing, pooled libraries were normalized to 2nM and denatured using 0.1 N NaOH prior to sequencing. Flowcell cluster amplification and sequencing were performed according to the manufacturer's protocols using either the NextSeq or NovaSeq. Libraries were sequenced with a read structure of 28x10x10x90 using Broad Clinical Labs Walkup Sequencing product.

scRNAseq and genotyping of transcriptomes (GOT) of Human Bone Marrow Cells

Cryopreserved human bone marrow samples were thawed and resuspended in PBS with 3% FBS and 1% P/S. The cells were then stained with CD34-APC (BD Biosciences # 555824 clone 581) for 30 min at 4°C. After washing, cells were resuspended in PBS with 3% FBS and 1% P/S containing 0.1 µg/mL DAPI. Viable, $CD34^+$ cells were sorted using BD FACS-Aria cell sorter (BD Biosciences), and the cell concentration was adjusted to 1000 cells/µL before library preparation. Library preparation and Illumina sequencing were done as described above. The cDNA was further used for genotyping of transcriptomes, which was performed as previously described [14].

Mouse HSPCs scRNAseq alignment, data processing, cell-type classification and clustering

For single cell sequenced mouse samples, the pooled scRNAseq and CITEseq were processed with Cell Ranger (v7.1.0) using the cellranger-count pipeline. The reads were aligned to the mm10 mouse genome with default parameters. The Seurat package (v5.0.2) was used to perform unbiased clustering of the sorted cells from each sample. In brief, for individual datasets, cells with UMI > or < 3 standard deviations from the mean UMI and mitochondrial gene percentage >10% were filtered and variation due to cell cycle was regressed out. These individual datasets were integrated and underwent batch-correction within Seurat, which implements principal component analysis (PCA) and the principles of mutual nearest neighbor. Recommended settings were used for the integration (30 principal components for the anchor determining procedure in IntegrateLayers function). Principal component analysis was performed using variable genes using recommended settings (i.e., top 2000 variable genes using variance stabilizing transformation). The first 30 statistically significant principal components were used as inputs to the UMAP algorithm for dimensionality reduction and visualization. Clusters were manually assigned based on differentially expressed genes using the FindAllMarkers function using default settings (using top 2000 variable genes, in a minimum of 10% of cells in either of the two comparison sets as input, and log-transformed fold change of 0.25 as the threshold, using Wilcoxon rank sum test). The clusters were annotated according to canonical lineage markers identified previously in scRNAseq data of murine HSPC into 11 progenitor subsets based on expression of levels of these canonical markers [15]. For Supplemental Figure 8C, the genes within the Hallmark gene sets ‘Inflammatory response’ and ‘Interferon alpha response’ were used as features in Seurat’s AddModuleScore function and the relative expression of the genes within the

HSC cluster were determined, split by genotype. The CITE-seq data was normalized with centered log-ratio (CLR) transformation and used to further distinguish and confirm cell types. Pseudotime analysis was performed by constructing cell lineage trajectory with Monocle [16-18]. The as.cell_data_set function from SeuratWrappers package was used to convert Seurat object to a CellDataSet object used by Monocle 3.

CRISPR/Cas9-mediated genetic knock out of Mef2c in mouse HSPCs

Mice were 8-weeks of age when MxCre was activated by polyinosinic polycytidylic acid pIpC injection. 12 weeks after pIpC injection, bone marrow cells were isolated by crushing the femur, tibia, and spine. cKit$^+$ cells were incubated with Miltenyi anti-CD117 microbeads (#130-091-224) and purified using MACS™ Miltenyi columns as described above.

Synthetic chemically modified single-strand guide RNAs (sgRNA) were purchased from Integrated DNA technology. sgRNAs were designed using CHOPCHOP [19]. The target sequence for the sgRNAs were as follows: *Mef2c* exon 4-sgRNA 5'-GAGTTTGTCCGGCTCTCGTG-3', *Mef2c* exon 3-sgRNA 5'-GATTCAGATTACGAGGATAA-3', and *Rosa26*-sgRNA 5'-ACTCCAGTCTTTCTAGAAGA-3'. HiFi SpCas9 protein was purchased from IDT. 3 μM Cas9 and 6 μM sgRNA were incubated for 15 min in P3 buffer (Lonza, #V4XP-3032) to complex ribonucleoprotein (RNP). c-Kit$^+$ HSPCs cells were resuspended in P3 buffer with complexed RNPs and electroporated using the Lonza 4D Nucleofector and 4D-Nucleofector X Unit (program: CA-137). Immediately following electroporation, the HSPCs were plated in megakaryocyte differentiation media (as described above) for 5

days. Either 3 days or 5 days after starting the megakaryocyte differentiation, the cells were collected and stained with CD41-APC antibody (Biolegend #133914 clone MWReg30) and CD61-PE antibody (Biolegend #104307 clone 2C9.G2) for 30 min. After washing, propidium iodide (Thermo Fisher Scientific) was used for dead cell staining. Flow cytometry was performed using FACSymphony A5 (Becton Dickinson). Post-acquisition analysis of data was performed using FlowJo software V11.

Since our scRNAseq mouse HSPC data revealed that *Mef2c* expression differed significantly but only modestly among genotypes (Figure 6B), we aimed to achieve a half dose reduction in *Mef2c* expression by generating predominantly heterozygous *Mef2c* knockout cells. Assuming independent editing of the two alleles, the expected frequency of heterozygous depletion cells follows a binomial distribution and is maximal when per-allele editing efficiency is approximately 50%. Therefore, moderate editing efficiencies (approximately 40–60%) are expected to enrich for heterozygous cells while limiting biallelic knockout. Consistent with this theoretical prediction, single-cell analyses have shown that increasing CRISPR editing activity is associated with progressively higher frequencies of biallelic editing[20].

Genomic DNA extraction

2 days after the electroporation with RNP, the HSPCs were collected and washed with PBS. The cell pellets were suspended with QuickExtract DNA Extraction Solution (BIOSEARCH TECHNOLOGIES, #QE09050) to extract genomic DNA according to the manufacturer's protocol.

Cell lines

Ba/F3 cells were purchased from German Collection of Microorganisms and Cell Cultures (DSMZ).

Intracellular FACS

Murine Ba/F3 cells were stained with anti-mouse CD16/CD32 antibody (BD #553141) for Fc receptor blocking, followed by the fixation and permeabilization of the cells with BD Cytofix/Cytoperm™ Fixation/Permeabilization Kit (BD Biosciences, Cat. No. 554714) according to manufacturer's protocol. The cells were then stained with either anti-mouse MEF2C antibody (Invitrogen #MA5-17119 clone 6H2G2) or isotype control as a primary antibody (Invitrogen #MA110407 clone MOPC-21), followed by PE-labelled anti-mouse IgG1 antibody (eBiosciences #12-4015-82 clone M1-14D12) as a secondary antibody. Following a wash step, flow cytometric analysis was performed using a FACSymphony A5 with BD FACS Diva software. The results were analyzed with FlowJo software v11.

Quantification of knock out efficiency with Tracking of Indels by Decomposition (TIDE) analysis

The genomic regions targeted with CRISPR/Cas9 were amplified with Platinum™ SuperFi II PCR Master Mixes (ThermoFisher SCIENTIFIC, #12368010). Primer sequences for PCR of targeted genomic regions were designed by using CHOPCHOP [19] as follows:

*Mef2c* exon 4-forward 5'- GAGTTTGTCCGGCTCTCGTG-3',

*Mef2c* exon 4-reverse 5'- GAGTTTGTCCGGCTCTCGTG-3',

*Mef2c* exon 3-forward 5'-GATTCAGATTACGAGGATAA-3',

*Mef2c* exon 3-reverse 5'-GATTCAGATTACGAGGATAA-3',

*Rosa26*-forward 5'-ACTCCAGTCTTTCTAGAAGA-3',

and *Rosa26*-reverse 5'-ACTCCAGTCTTTCTAGAAGA-3'.

PCR amplicons were submitted to Genewiz for Sanger sequencing to generate ab1 files according to manufacturer's protocol. The results of Sanger sequencing with ab1 files were analyzed using Tracking of Indels by Decomposition using the default setting to quantify the frequency of insertions and deletions at the targeted genomic region [21].

GoT scRNAseq data processing, alignment, cell-type classification and clustering

For single-cell sequenced patient samples, the pooled scRNAseq data were processed with Cell Ranger (v7.1.0) using the cellranger-count pipeline. The reads were aligned to the human genome GRCh38 with default parameters. The Seurat package (v5.0.2) was used to perform unbiased clustering of the sorted cells from each sample. In brief, for individual datasets, cells with UMI > or < 3 standard deviations from the mean UMI and mitochondrial gene percentage >20% were filtered out. These individual datasets were integrated with five external *CALR* single mutant samples from Nam et al., 2019 and underwent batch-correction within Seurat, which implements principal component analysis (PCA) and the principles of mutual nearest neighbor. Recommended settings were used for the integration (30 principal components for the anchor determining procedure in IntegrateLayers function). Principal component analysis was performed using top 3000 variable genes using variance stabilizing transformation. The first 30 statistically significant principal components were used as inputs to the UMAP algorithm for dimensionality reduction and visualization.

Clusters were manually assigned based on differentially expressed genes using the FindAllMarkers function using top 3000 variable genes, in a minimum of 10% of cells in either of the two comparison sets as input, and log-transformed fold change of 0.25 as the threshold, using Wilcoxon rank sum test. The clusters were annotated according to canonical lineage markers identified previously in scRNAseq data into 11 progenitor subsets based on expression of levels of these canonical markers [14].

IronThrone-GoT for processing targeted amplicon sequences and mutation calling

Analysis of the GoT library was carried out as described previously [14] using IronThrone pipeline (v.2.1). Amplicon reads (Read 2) were screened for the presence of the primer sequence and the shared sequence (i.e., the expected sequence between the primer sequence and the mutation locus). Reads (Read 1) from GoT-IM experiments were also assessed for matching to the cell barcode list of the 10x dataset. A mismatch of 20% was allowed for all sequence matching steps. For *CALR* genotyping, only UMIs with at least 2 or more supporting reads were retained for final genotyping assignments, after the UMI collapse algorithm, as previously described [14]. For *ASXL1* mutations, given the known technical artifact of high background noise [22], greater stringency for mutation calling was applied. Specifically, only UMIs with at least 5 or 6 reads were retained and UMIs with a mutant read ratio at least greater than .75 were called mutant and at least less than .25 were called wildtype, using bulk DNA sequencing VAF as a reference. Filtered cells were then genotyped as follows: cells with at least one mutant UMI were categorized as mutant cells whereas cells with no mutant UMI and at least one wildtype UMI were identified as wildtype. For patients with co-occurring mutations, cells labeled as having insufficient data

indicate cases where either the relevant mutation was not detected or no mutation was captured. *CALR+ASXL1* Patient 1 had an *ASXL1* VAF too low to proceed with genotyping.

Differential Expression, Gene Module Scoring, Statistical Analysis

The tested genes included the variable genes from the CCA integration, which were filtered for those expressed in at least 10% of either group. In aggregated differential gene expression analysis, the two groups were compared via the linear mixed model framework. For each gene, the variable in question was entered as the fixed effect and samples as random effects. P-values were obtained from likelihood ratio tests of the full model with the effect in question against the model without the effect in question. Individual genes with abs(avg_log2FC) > 0.1 and adjusted p-value < 0.05 were significant for further analysis.

For examining gene module expression in Figure 7D, the top 200 upregulated genes in the MkP cluster were used as features in Seurat's AddModuleScore function and the relative expression of the genes within the HSC cluster, split by genotype, were determined.

For statistical analysis, when the variable in question was cell type identity, cell type identity was entered as a fixed effect and sample as a random effect in a linear mixed-effects model. When the variable in question was genotype status, genotype status was entered as a fixed effect and sample as a random effect in a linear mixed-effects model. P-values for overall effects were obtained by likelihood ratio tests comparing full and reduced models. For pairwise genotype comparisons, estimated marginal means were calculated using the emmeans package and P-values were adjusted for multiple testing

using the Benjamini–Hochberg procedure. Linear mixed-effects analyses were performed using the lme4 package (v.1.1.3).

**Supplemental references**


1. Jutzi JS, Marneth AE, Jiménez-Santos MJ, et al. CALR-mutated cells are vulnerable to combined inhibition of the proteasome and the endoplasmic reticulum stress response. *Leukemia*. 2022/12/06 2022;37(2):359-369. doi:10.1038/s41375-022-01781-0
2. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2012/10/25 2012;29(1):15-21. doi:10.1093/bioinformatics/bts635
3. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014/12/05 2014;15(12)doi:10.1186/s13059-014-0550-8
4. Cornwell M, Vangala M, Taing L, et al. VIPER: Visualization Pipeline for RNA-seq, a Snakemake workflow for efficient and complete RNA-seq analysis. *BMC Bioinformatics*. 2018/04/12 2018;19(1)doi:10.1186/s12859-018-2139-9
5. Dele-Oni DO, Christianson KE, Egri SB, et al. Proteomic profiling dataset of chemical perturbations in multiple biological backgrounds. *Scientific Data*. 2021/08/25 2021;8(1)doi:10.1038/s41597-021-01008-4
6. MacLean B, Tomazela DM, Shulman N, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*. 2010/02/09 2010;26(7):966-968. doi:10.1093/bioinformatics/btq054
7. Rossmann MP, Hoi K, Chan V, et al. Cell-specific transcriptional control of mitochondrial metabolism by TIF1γ drives erythropoiesis. *Science*. 2021/05/14 2021;372(6543):716-721. doi:10.1126/science.aaz2740
8. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009/05/18 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324
9. Zhang Y, Liu T, Meyer CA, et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*. 2008/09/17 2008;9(9)doi:10.1186/gb-2008-9-9-r137
10. Qin Q, Mei S, Wu Q, et al. ChiLin: a comprehensive ChIP-seq and DNase-seq quality control and analysis pipeline. *BMC Bioinformatics*. 2016/10/03 2016;17(1)doi:10.1186/s12859-016-1274-4
11. Qiu X, Feit AS, Feiglin A, et al. CoBRA: Containerized Bioinformatics Workflow for Reproducible ChIP/ATAC-Seq Analysis. *Genomics, Proteomics & Bioinformatics*. 2021/07/18 2021;19(4):652-661. doi:10.1016/j.gpb.2020.11.007
12. Neph S, Kuehn MS, Reynolds AP, et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics*. 2012/05/09 2012;28(14):1919-1920. doi:10.1093/bioinformatics/bts277
13. Heinz S, Benner C, Spann N, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*. 2010/05 2010;38(4):576-589. doi:10.1016/j.molcel.2010.05.004
14. Nam AS, Kim K-T, Chaligne R, et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature*. 2019/07/03 2019;571(7765):355-360. doi:10.1038/s41586-019-1367-0

15. Izzo F, Lee SC, Poran A, et al. DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nature Genetics*. 2020/03/23 2020;52(4):378-387. doi:10.1038/s41588-020-0595-4
16. Satpathy AT, Granja JM, Yost KE, et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nature Biotechnology*. 2019/08 2019;37(8):925-936. doi:10.1038/s41587-019-0206-z
17. Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nature Methods*. 2017/08/21 2017;14(10):979-982. doi:10.1038/nmeth.4402
18. Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*. 2014/03/23 2014;32(4):381-386. doi:10.1038/nbt.2859
19. Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res*. Jul 2 2019;47(W1):W171-W174. doi:10.1093/nar/gkz365
20. Kalter N, Gulati S, Rosenberg M, et al. Precise measurement of CRISPR genome editing outcomes through single-cell DNA sequencing. *Mol Ther Methods Clin Dev*. Jun 12 2025;33(2):101449. doi:10.1016/j.omtm.2025.101449
21. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res*. Dec 16 2014;42(22):e168. doi:10.1093/nar/gku936
22. Abdel-Wahab O, Adli M, LaFave Lindsay M, et al. ASXL1 Mutations Promote Myeloid Transformation through Loss of PRC2-Mediated Gene Repression. *Cancer Cell*. 2012/08 2012;22(2):180-193. doi:10.1016/j.ccr.2012.06.032

## Supplementary Figure Legends

**Supplemental Figure 1: *Calr-Asxl1* co-mutant mice display a more severe MPN phenotype. A** Mean platelet volume (MPV) of peripheral blood (PB) platelets and mean corpuscular volume (MCV) of PB erythrocytes. **B** Gating strategy used to phenotype myeloid and lymphoid BM cells. n=6-8 per group. **C** Proportion of long term (LT-) and short term (ST-) hematopoietic stem cell (HSCs) as well as common myeloid progenitors (CMPs), granulocyte-monocyte progenitors (GMPs), and megakaryocyte-erythroid progenitors (MEPs) within bone marrow. n=5-6 per group. *p≤0.05, **p≤0.01, ***p≤0.001, ****p≤0.0001 by one-way ANOVA. Absence of asterisks indicates non-significance. ns: not significant.

**Supplemental Figure 2: *Calr-Asxl* double-mutant mice demonstrate increased emperipolesis**. **A** Representative H&E-stained bone marrow sections from 65-week-old mice, showing dysmorphic, clustered megakaryocytes with hypolobulated to bulbous nuclei and increased emperipolesis in C+A double mutant (yellow arrow). A representative result from n = 3-4 mice per each genotype. **B** Quantification of emperipolesis index (%) showing a significant increase in *Calr–Asxl1* double-mutant (C + A) compared with wild-type (WT) and *Calr* (C) mice. Data are presented as mean ± SEM. Statistical significance was determined by one-way ANOVA followed by Tukey's multiple comparisons test (*p<0.05,**P < 0.01, ***P < 0.001, ****P < 0.0001).

**Supplemental Figure 3: Bone marrow transplantation of floxed *Calr-Asxl1* co-mutant cells, followed by Cre recombinase activation, induces thrombocytosis in wild-type recipient mice.** Experimental design for bone marrow transplantation (BMT) of c-Kit+ cells from MxCre (WT), MxCre-*Asxl1*$^{mut/mut}$ (A), MxCre-*Calr*$^{del52/+}$ (C) and MxCre-

*Calr*$^{del52/+}$-*Asxl1*$^{mut/mut}$ (C+A) mice into lethally irradiated CD45.1+ WT recipient mice. Cre recombination was activated by injecting polyinosinic polycytidylic acid (pIpC) 4 weeks after BMT (n=4-5 per group). **B** White blood cell (WBC), red blood cell (RBC) and platelet (PLT) counts over time in BMT mice

**Supplemental Figure 4: Dysregulated transcriptome of *Calr-Asxl1* co-mutant megakaryocytes. A** Experimental design to assess the transcriptome of ex vivo differentiated megakaryocytes of MxCre (WT), MxCre-*Asxl1*$^{mut/mut}$ (A), MxCre-*Calr*$^{del52/+}$ (C) and MxCre-*Calr*$^{del52/+}$-*Asxl1*$^{mut/mut}$ (C+A) mice. 8 weeks post induction, total bone marrow cells were differentiated in medium supplemented with 50 ng/mL thrombopoietin (TPO). After 5 days, mature megakaryocytes were enriched and subjected to RNA sequencing (RNAseq, n=3-4 per group). **B** Principal component analysis (PCA) of WT, A, C and C+A megakaryocyte transcriptomes. **C** Venn diagram showing the number of differentially expressed genes that are unique to and shared between the genotypes compared to WT control megakaryocytes. **D** Volcano plot showing differentially expressed genes and implicated pathways in C+A vs C megakaryocytes. **E** Gene set enrichment analysis (GSEA) of Hallmark gene sets in C+A and C megakaryocytes. Red (up) and blue (down) bars indicate significantly enriched gene sets. *p<0.05, **p≤0.01, ***p≤0.001, ****p≤0.0001 by one-way ANOVA. Absence of asterisks indicates non-significance.

**Supplemental Figure 5: Altered histone H3 protein tri-methylation dysregulates the transcriptome of *Calr-Asxl1* co-mutant megakaryocytes. A** Experimental design to assess the transcriptome of ex vivo differentiated megakaryocytes of MxCre (WT),

MxCre-*Asxl1*$^{mut/mut}$ (A), MxCre-*Calr*$^{del52/+}$ (C) and MxCre-*Calr*$^{del52/+}$-*Asxl1*$^{mut/mut}$ (C+A) mice. 8 weeks post induction, total bone marrow cells were differentiated in medium supplemented with 50 ng/mL thrombopoietin (TPO). After 5 days, mature megakaryocytes were enriched and subjected to global chromatin profiling (GCP, n=4-5 per group). **B** Heatmap showing the abundance of histone 3 peptides detected by GCP. **C** Volcano plots comparing detected histone marks between A vs WT and C+A vs WT megakaryocytes. Each dot represents a single modification at the amino acid level **D** Representative pictures of immunohistochemistry stain of H3S10ph of bone marrow (BM) of aged mice. Scale bar = 100 µm. **E** Boxplot showing differential expression of *Aurka* and *Aurkb* in differentiated megakaryocytes. *p<0.05, **p≤0.01, ***p≤0.001, ****p≤0.0001 by one-way ANOVA. Absence of asterisks indicates non-significance.

**Supplemental Figure 6: Mutant *Asxl1* alters the hematopoietic stem and progenitor cell (HSPC) landscape in *Calr*-mutant mice.** Heatmap of population defining markers of hematopoietic stem and progenitor cells (HSPC) determined by single cell RNA sequencing (scRNAseq). CLP: common lymphoid progenitor, EP: erythrocyte progenitor, HSC: hematopoietic stem cell, CMP: common myeloid progenitor, Mast/Baso: Mast cell/basophil progenitor, EP: erythrocyte progenitor, MkP: megakaryocyte progenitor, MP: monocyte progenitor, NP: neutrophil progenitor).

**Supplemental Figure 7: Mutant *Asxl1* alters the hematopoietic stem and progenitor cell (HSPC) landscape in *Calr*-mutant mice.** Heatmap of population defining cell surface markers of hematopoietic stem and progenitor cells (HSPC) determined by single cell RNA sequencing (scRNAseq) with feature barcoding. CLP: common lymphoid progenitor, EP: erythrocyte progenitor, HSC: hematopoietic stem cell, CMP: common

myeloid progenitor, Mast/Baso: Mast cell/basophil progenitor, EP: erythrocyte progenitor, MkP: megakaryocyte progenitor, MP: monocyte progenitor, NP: neutrophil progenitor).

**Supplemental Figure 8: Mutant *Asxl1* alters the hematopoietic stem and progenitor cell (HSPC) landscape in *Calr*-mutant mice. A** Histone 3 lysine 4 trimethylation (H3K4me3) CUT&RUN tracks in the *Mef2c*, *Cd34* and *Flt3* promoter region. **B** Histone 3 lysine 27 trimethylation (H3K27me3) CUT&RUN tracks in the *Mef2c*, *Cd34* and *Flt3* promoter region.

**Supplemental Figure 9: *Mef2c* contributes to megakaryocyte differentiation *in vitro.***

**A** *Mef2c* expression of MxCre (WT), MxCre-*Calr*$^{del52/+}$ (C) and MxCre-*Calr*$^{del52/+}$-*Asxl1*$^{mut/mut}$ (C+A) across progenitor clusters (NP: neutrophil progenitor, MP: monocyte progenitor, EP: erythrocyte progenitor, Baso/Mast: Basophil/Mast cell progenitor, CLP: common lymphoid progenitor). Significance was determined using linear mixed-effects models with sample as a random effect and genotype as a fixed effect, followed by Benjamini–Hochberg-adjusted pairwise comparisons (*P < 0.05, **P < 0.01, ***P < 0.001). ns: not significant. **B** Representative morphology of WT murine c-Kit$^+$ bone marrow cells on day 0 (left), 3 (middle) and 5 (right) during 5-day in vitro megakaryocyte differentiation. A representative result from N = 4 experimental replicates. Scale bars, 100 μm **C** Gating strategy used to identify megakaryocytes differentiated *in vitro*. **D** Representative flow cytometry of WT murine c-Kit$^+$ bone marrow cells on day 0 (left), 3 (middle) and 5 (right) during 5-day *in vitro* megakaryocyte differentiation. A representative result from n = 4 experimental replicates. **E** Mean ± s.d. of the percentage of Cd41$^+$Cd61$^+$ cells 3 days after *in vitro* megakaryocyte differentiation. n = 4 experimental replicates per each condition. ****p≤0.0001 by two-way ANOVA. ns indicates not significant. *Mef2c* KO1 and *Mef2c* KO2

conditions were both compared to the negative control (*Rosa26* KO) for each genotype. ns: not significant.

**Supplemental Figure 10: Co-mutation of *CALR* and *ASXL1* in MPN patients promotes transcriptional skewing toward megakaryopoiesis and a deficit in lymphoid lineage potential. A** Boxplots showing number of UMIs (left panel), and genes (right panel) detected per cell in sorted $CD34^+$ hematopoietic progenitors from each patient after filtering based on quality control (QC) metrics (methods). **B** Uniform Manifold Approximation and Projection (UMAP) of sorted $CD34^+$ progenitors (n = 27,202 cells, samples from this study and Nam et al., 2019) highlighted by patient ID (n = 10 individuals) **C** Dot plot showing expression levels of cell type specific gene markers in each progenitor subset. **D** Barplot showing the proportion of cells carrying a specific genotype within MkP cluster, in relation to all cells of this genotype. Overlaid counts indicate the number of cells contributing to each genotype's proportion.