RUNX1 and CBFβ-SMMHC transactivate target genes together in abnormal myeloid progenitors for leukemogenesis

By Tao Zhen et al.

Supplemental Methods

Flow cytometry

Peripheral blood, spleen and bone marrow cells from mice were isolated and stained as described previously.[1] Lineage cells were determined by staining with lineage-specific antibodies including CD3-PE (BD Biosciences, AB_395699), CD4-PE (BD Biosciences, AB_394585), CD8-PE (BD Biosciences, AB_394571), B220-PE (BD Biosciences, AB_394620), Gr1-PE (BD Biosciences, AB_394644), Mac1-PE (BD Biosciences, AB_394775), Ter119-PE (BD Biosciences, AB_394986) and CD127-PE (eBioscience, AB_465845). Other fluorescence-conjugated anti-mouse antibodies used in this study include CD19-PE (eBioscience, AB_396682), CD3-PerCP-Cyanine5.5 (eBioscience, AB_1107000), CD4-PerCP-Cyanine5.5 (eBioscience, AB_1107001), CD8-PerCP-Cyanine5.5 (eBioscience, AB_1107004), Sca1-Pacific blue (BioLegend, AB_493273), Mac1-APC-Cy7 (eBioscience, AB_1603193), c-Kit-APC (eBioscience, AB_469431), CD34-FITC (eBioscience, AB_465022), Fc□RII/III-PE-Cyanine7 (eBioscience, AB_469598), CD45.1-FITC (eBioscience, AB_465058), Ter119-APC (eBioscience, AB_469473) and CD71-PerCP-Cyanine (eBioscience, AB_925765). Data were acquired using a LSRII Flow Cytometer (BD Biosciences) and analyzed using FlowJo software (TreeStar).

Western blot analysis

Western blot was performed with standard protocols (Invitrogen, NuPAGE electrophoresis system). Internal control α-tublin and β-actin were detected with anti-α-tublin (Abcam, AB_2241126) and anti-β-actin (Millipore Sigma, AB_476743) antibodies respectively, followed by anti-mouse IgG-HRP antibody (Vector Laboratories, AB_2336177). RUNX1 and CBFβ/CBFβ-SMMHC were detected with anti-RUNX1 (Abcam, AB_2049267) and anti-CBFβ (Abcam, AB_725993) antibodies respectively, followed by anti-Rabbit IgG-HRP antibody (Vector

Laboratories, AB_2336198). Lamin B was detected with anti-Lamin B1 (HRP) antibody (Abcam, ab194109).

Immunofluorescence

293T cells grown on cover glasses were transfected with indicated plasmids. Twenty-four hours later, the cells were fixed with 4% paraformaldehyde for 15 minutes. Immunofluorescence staining was performed with standard protocols. Images were captured with laser scanning confocal microscope (Zeiss) and analyzed with Zen software (Zeiss).

Mouse bone marrow were attached to glass microscopic slides with cytospin (SHANDON). The cells on the slides were then fixed with 4% paraphomyldehyde for 15 mins, followed with 0.25% Triton X-100 for 10 mins. Immunofluorescent staining was performed with standard protocols. CBFβ/CBFβ-SMMHC was stained with anti-CBFβ (Abcam, AB_725993) antibody, followed with anti-rabbit Alexa Fluor 568 fluorescent antibody (Thermo Fisher Scientific, AB_143011). RUNX1 protein was first stained with anti-RUNX1 antibody (Abcam, AB_2238869), followed with anti-mouse Alexa Fluor 488 fluorescent antibody (Thermo Fisher Scientific, AB_2536161). Images were captured with a laser scanning confocal microscope (Zeiss) and analyzed with Zen software (Zeiss).

RNA-sequencing, ChIC-sequencing and single cell RNA-sequencing

Two to three weeks after the last pIpC treatment, bone marrow cells were harvested and AMP cells were sorted out with a BD FACSAria IIIu cell sorter (BD Biosciences).

For RNA-seq: Total RNA from AMP cells was extracted with AllPrep DNA/RNA/Protein Mini Kit (QIAGEN). Poly-A selected stranded mRNA libraries were constructed using the Illumina TruSeq Stranded mRNA Sample Prep Kits according to manufacturer's instructions. Unique barcode adapters were applied to each library. All libraries were combined in equimolar proportion into one pool for sequencing with a NovaSeq6000 sequencer. Raw reads were mapped to mouse genome mm10 with STAR (v2.5.1b)[2]. The gene expression levels were quantified and normalized by cuffnorm from cufflinks (v2.2.1)[3] package into an FPKM matrix. Genes were filtered off if their expression levels < 1 FPKM in all samples. The expression matrix was further log2-transformed. The differentially expressed genes (DEGs) were called with cuffdiff (v2.2.1),[4] requiring *q-value* <0.01, absolute fold change between the two conditions >=2, the sum of mean

expression level between the two conditions >=1 and genes in the filtered expression matrix. The enriched GO terms were identified by findGO.pl from HOMER package[5] for the DEGs. Gene set enrichment analysis of DEGs was performed using GenePattern 2.0[6] with GSEA preranked modules. Gene set enrichment analysis of entire RNA-seq data was carried out with gene expression matrix of each samples using standard method.[7] Ingenuity Pathway Analysis (Ingenuity Systems) was performed to identify upstream regulators of the DEGs. The RNA-seq dataset has been deposited to GEO (accession number GSE152573).

For ChIC-seq: AMP cells were fixed with 1% PFA and subjected to ChIC-seq as described.[8] Specifically, antibody+PA-MNase complex with 50,000 cells was used for each sample. RUNX1 (Abcam, AB_2184205), CBFβ (Diagenode, C15310002), MYH11 (Diagenode, C15310254), H3K27Ac (Abcam, AB_2118291) and H3K27me3 (Millipore Sigma, AB_310624) antibodies were used to identify the binding sites of the indicated proteins. All libraries were combined in equimolar proportion into one pool for sequencing with a HiSeq 3000 sequencer. Paired-end raw reads were mapped to mouse genome mm10 by Bowtie2 (v2.3.5)[9] with key parameters of --no-mixed --no-discordant --local --very-sensitive. For the mapped paired end reads, only MAPQ>=10, non-redundant and fragment size >=140bp were kept for subsequent analysis. The ChIC IgG data for *Mx1-CreCbfb*$^{+/56M}$ and *Runx1*$^{f/f}$*Mx1-CreCbfb*$^{+/56M}$ samples were pooled as control for calling peaks and normalization, and the corresponding samples from two batches were pooled together to call peaks peaks and aggregation profile comparison. ChIC-seq peaks were call by cLoops2 (will be available at: https://github.com/YaqiangCao/cLoops2) with key parameters of -minPts 5, -sen, and -eps 100,200 was used for all other data except H3K27me3. For H3K27me3, -eps 300,500 was used to call border peaks. Briefly, the new peak-calling algorithm for ChIC-seq was based on a similar idea of cLoops[10]: reads were clustered to find candidate peaks, and the candidate peaks were compared to local permutated background and control data to estimate significance. Generally, Poisson test was used and p-value <= 1e-5 was used to obtain significant peaks. The target genes of the peaks were annotated with annotatePeaks.pl from HOMER, and motif analysis for peaks changed by *Runx1* knockout were carried out with findmotifsgenome.pl from HOMER. The aggregated normalized reads per million (RPM) signals for genes TSS were obtained from cLoops2 agg command. The ChIC-seq dataset has been deposited to GEO (accession number GSE152573).

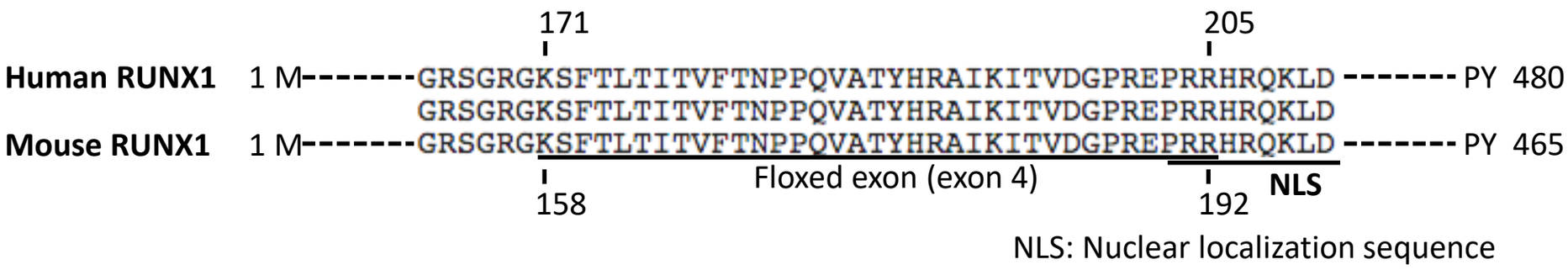
For single cells RNA-sequencing (scRNA-seq): 10X genomics chromium platform[11] was used to capture isolated AMP cells and all steps were performed according to manufacturer's instructions. The Chromium Single Cell 3' Library & Gel Bead Kit v2 was used. Libraries were sequenced on an Illumina NextSeq 550 sequencer. Reads were processed with 10x Cell Ranger (v3.1.0) command count, to the Cell Ranger mouse reference genome mm10-v3.0.0. Only cells detected at least 1,000 expressed genes and <=5% reads mapped to mitochondria genes were kept. Only genes expressed more than 10 cells were kept. The raw counts were further normalized by scaling the median counts through all cells, and two batches of data were separated for clustering analysis. The clustering of single cells were carried out with HDBSCAN (https://hdbscan.readthedocs.io/en/latest) (min_samples=20, min_cluster_size=100) of 2 reduced components from the expression matrix processed by UMAP[12] with key parameters of n_neighbors=100, n_components=2, min_dist=0, init="random",metric="euclidean", n_epoch=1000. The differentially expressed genes between conditions (or one cluster against all other clusters) were identified by two-sided t-test (scipy.stats.ttest_ind), requiring *p-value* <1e-5 and absolute fold change >=2. The scRNA-seq dataset has been deposited to GEO (accession number GSE152573).

References

1. Hyde RK, Zhao L, Alemu L, Liu PP. Runx1 is required for hematopoietic defects and leukemogenesis in Cbfb-MYH11 knock-in mice. *Leukemia*. 2015;29(8):1771-1778.
2. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
3. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*. 2010;28(5):511-U174.
4. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*. 2013;31(1):46-+.
5. 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;38(4):576-589.
6. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nat Genet*. 2006;38(5):500-501.

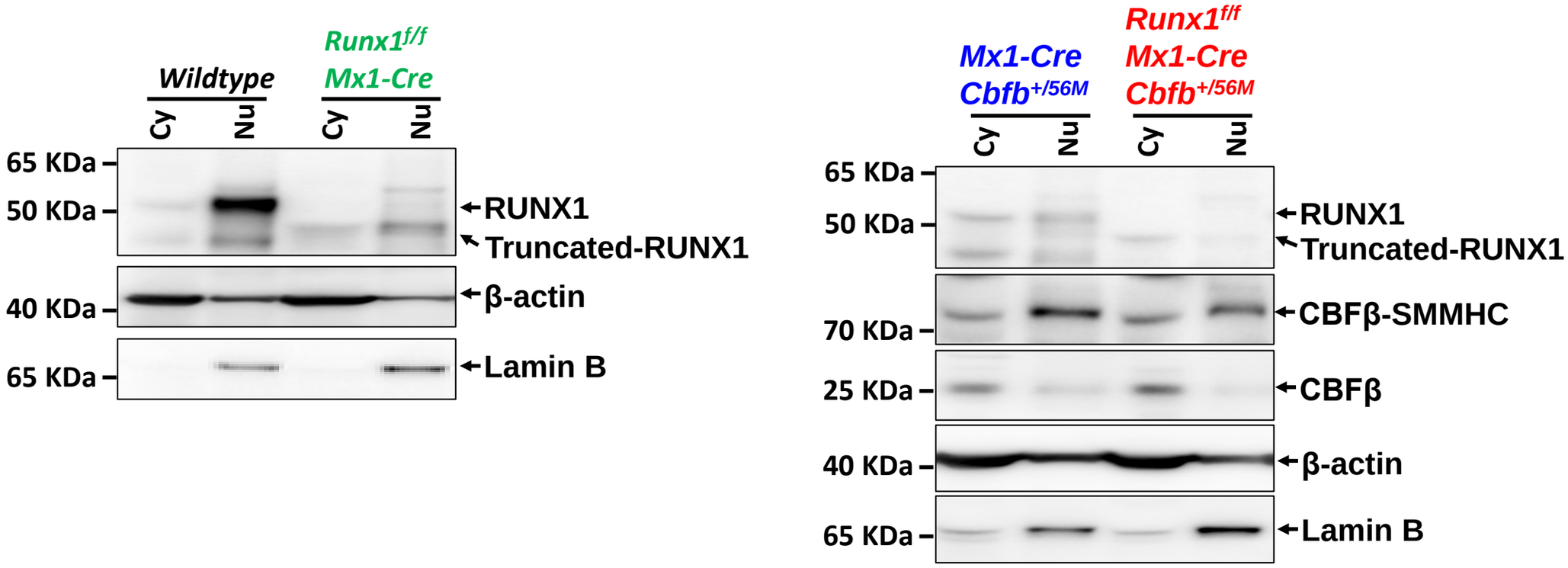
7. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550.
8. Ku WL, Nakamura K, Gao W, et al. Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat Methods*. 2019;16(4):323-325.
9. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359.
10. Cao YQ, Chen ZX, Chen XW, et al. Accurate loop calling for 3D genomic data with cLoops. *Bioinformatics*. 2020;36(3):666-675.
11. Zheng GX, Terry JM, Belgrader P, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*. 2017;8:14049.
12. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:180203426*. 2018.

Supplemental Figure 1



Supplemental Figure 1: Schematic representation of the human and mouse RUNX1 protein.
The amino acid sequence encoded by the floxed exon in the *Runx1$^{f/f}$* mice and the RUNX1 nuclear localization sequence are shown.

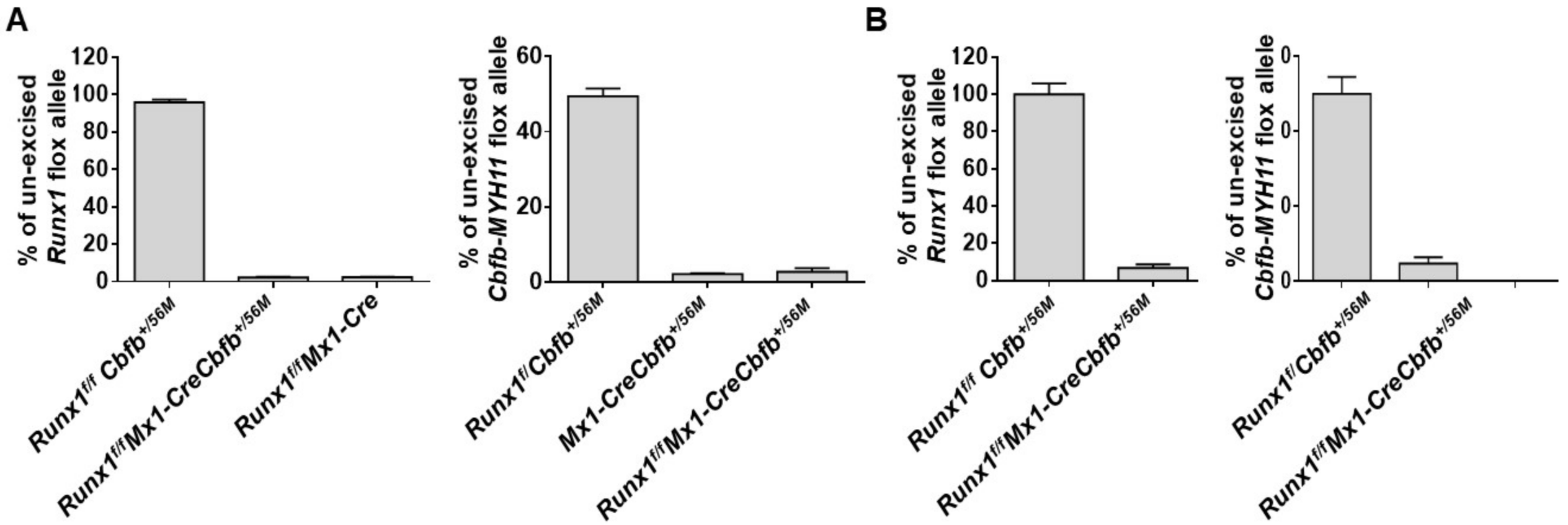
Supplemental Figure 2



Supplemental Figure 2: The mutant RUNX1 protein was expressed in the *Runx1$^{f/f}$ Mx1-Cre* bone marrow cells after Cre-mediated recombination.

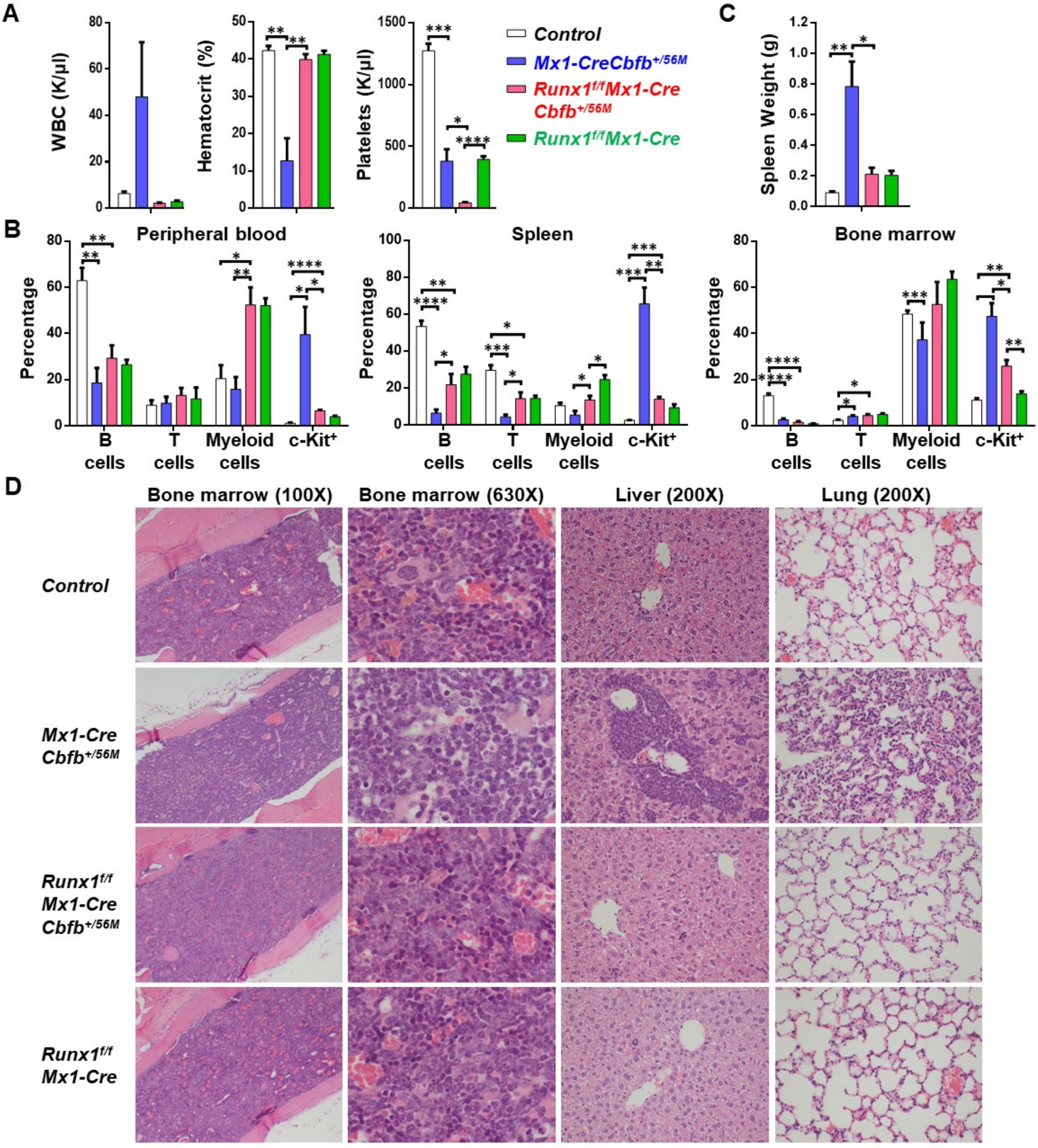
Indicated mice were treated with pIpC every other day for a total of 3 doses. 2 weeks after the last dose of injection, mice were sacrificed, and bone marrow cells were harvested. Cytoplasmic (Cy) and nuclear (Nu) protein fractions from these bone marrow cells were extracted. Western blot was performed to detect the expression of RUNX1, Truncated-RUNX1, CBFβ and CBFβ-SMMHC in cytoplasmic and nuclear fractions. β-actin and Lamin B were used as internal control for cytoplasmic and nuclear protein, respectively.

Supplemental Figure 3



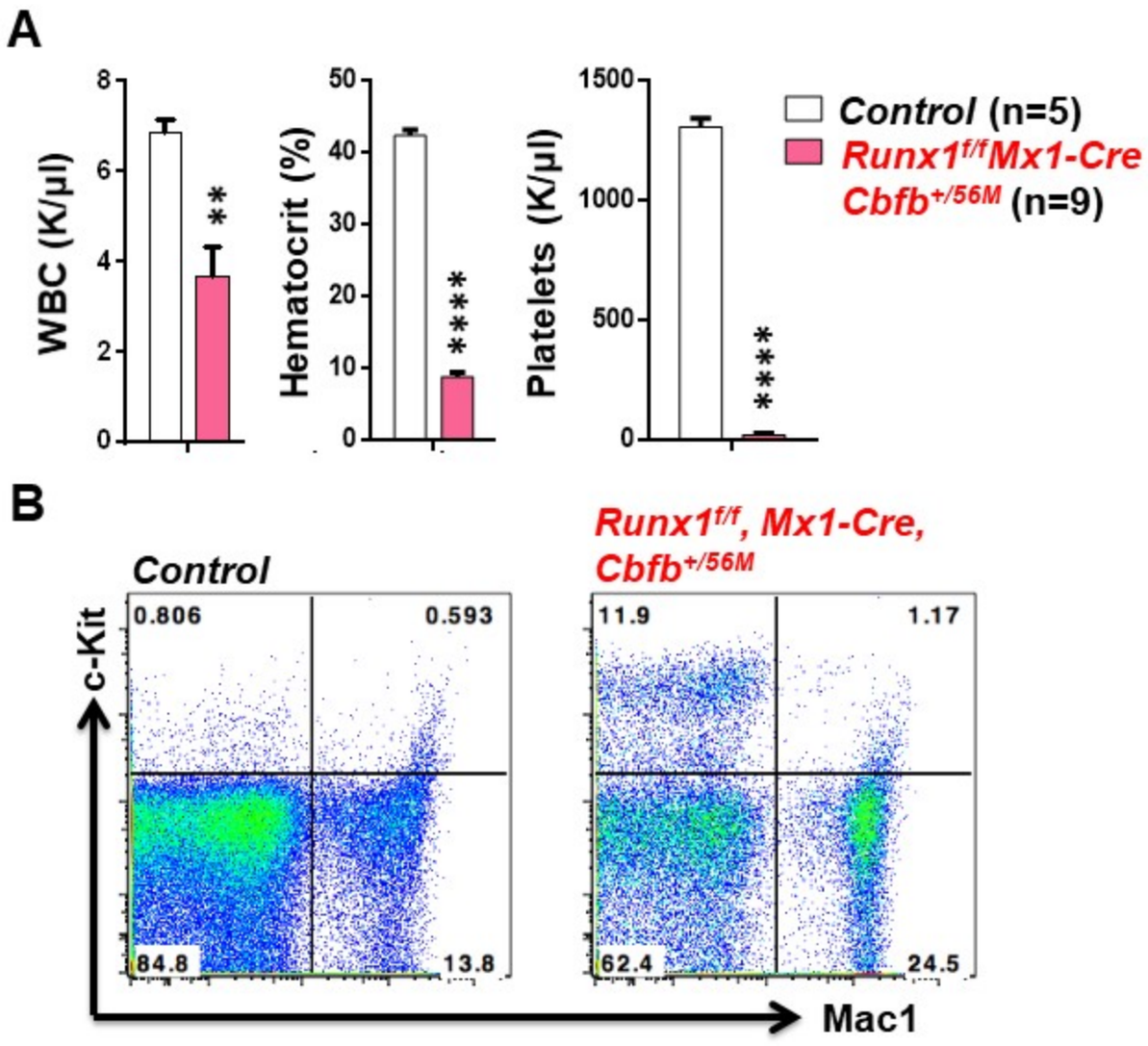
Supplemental Figure 3: Near complete excision of *Runx1* and *Cbfb-MYH11* flox alleles following pIpC treatment.
(A) Mice from each group (*Runx1$^{f/f}$Cbfb$^{+/56M}$, Mx1-CreCbfb$^{+/56M}$*,*Runx1$^{f/f}$Mx1-CreCbfb$^{+/56M}$ and Runx1$^{f/f}$Mx1-Cre,* 4 mice for each genotype) were treated with pIpC every other day for a total of 3 doses. Four months after the last injection, mice were sacrificed, and bone marrow cells were isolated. Genomic DNA from each bone marrow sample was extracted. The presence of un-excised *Runx1* and *Cbfb-MYH11* flox alleles in these DNA samples was detected by qPCR and normalized to un-excised *Runx1* (100%) and *Cbfb-MYH11* (50%) flox alleles in *Runx1$^{f/f}$Cbfb$^{+/56M}$*, respectively. (B) At 12 months after pIpC injection as mentioned above, treated mice from *Runx1$^{f/f}$Cbfb$^{+/56M}$* (n=3) *and Runx1$^{f/f}$Mx1-CreCbfb$^{+/56M}$* (n=4) were sacrificed, and bone marrow cells were isolated. Genomic DNA from each bone marrow sample was extracted. The presence of un-excised *Runx1* and *Cbfb-MYH11* flox alleles in these DNA samples was detected by qPCR and normalized to un-excised *Runx1* (100%) and *Cbfb-MYH11* (50%) flox alleles in *Runx1$^{f/f}$Cbfb$^{+/56M}$*, respectively.

Supplemental Figure 4



Supplemental Figure 4: Leukemia developed in *Mx1-CreCbfb$^{+/56M}$* mice but not *Runx1$^{f/f}$Mx1-CreCbfb$^{+/56M}$* mice. (A-D) Indicated mice (4 mice for each genotype) were treated with pIpC every other day for a total of 3 doses. Four months after the last injection, mice were sacrificed for analysis. (A) Bar graphs showing the white blood count (WBC), platelet count, and hematocrit in these mice (mean ± SEM). (B) Bar graphs showing the percentages of B cells (CD19+), T cells (CD3+, CD4+, CD8$^+$), myeloid cells (Gr1+, Mac1$^+$) and immature cells (c-Kit$^+$) in the peripheral blood, spleen and bone marrow of these mice (mean ± SEM). (C) Bar graph showing the spleen weight of these mice (mean ± SEM). **: $p < 0.05$; **: $p < 0.01$*; ****: $p < 0.001$; ****: $p < 0.0001$.* (D) Representative H&E stained bone marrow, liver and lung sections of these mice.

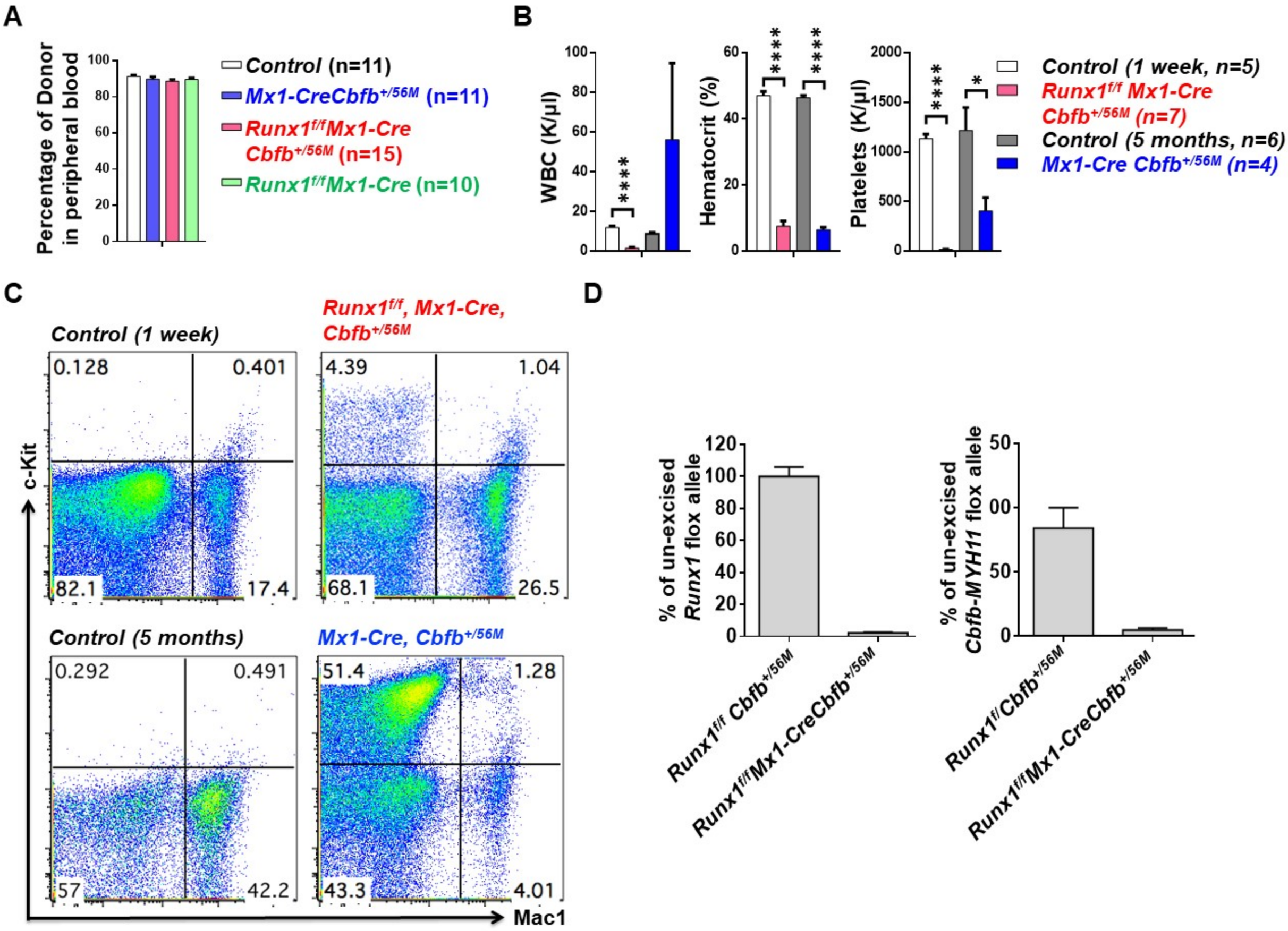
Supplemental Figure 5



Supplemental Figure 5: Early moribund *Runx1$^{f/f}$Mx1-CreCbfb$^{+/56M}$* mice most likely died with severe anemia but not leukemia.

(A-C) Analysis of peripheral blood cells from *Runx1$^{f/f}$Mx1-CreCbfb$^{+/56M}$* mice died within 2 weeks after the last dose of pIpC, together with control mice sampled at the same time. (A) Bar graphs showing the WBC, hematocrit and platelet counts (mean ± SEM). (B) Representative fluorescence-activated cell sorting (FACS) plots of c-Kit$^+$ and Mac1$^+$ cells in the peripheral blood of these mice. (C) Representative Wright-Giemsa stained peripheral blood smears from these mice. Scale bar, 10 μm. (D) Kaplan-Meier survival curves of C57BL/6 x 129/SvEv F1 mice (CD45.1), sub-lethally irradiated (650 rads) and transplanted with spleen cells from moribund *Runx1$^{f/f}$Mx1-CreCbfb$^{+/56M}$* mice (within two weeks after pIpC injection, n=4) and end stage *Mx1-CreCbfb$^{+/56M}$* leukemic mice (n=3). $p < 0.0001$, compared with *Mx1-CreCbfb$^{+/56M}$* group.

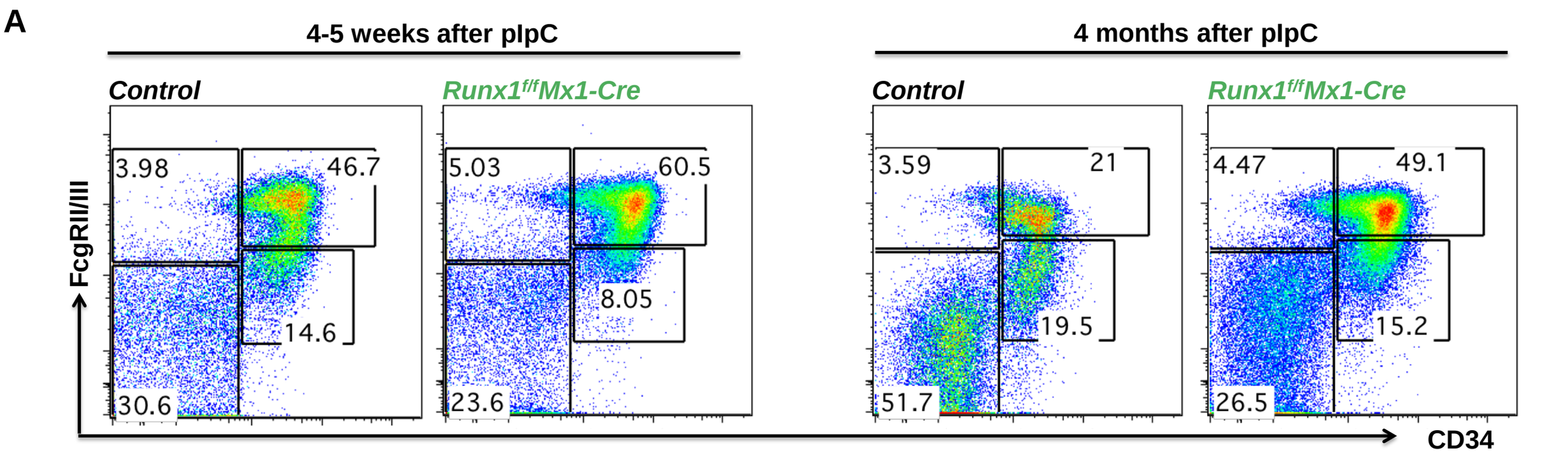
Supplemental Figure 6



Supplemental Figure 6: Leukemia developed in mice transplanted with bone marrow cells from *Mx1-CreCbfb*$^{+/56M}$ mice but not *Runx1*$^{f/f}$*Mx1-CreCbfb*$^{+/56M}$ mice.

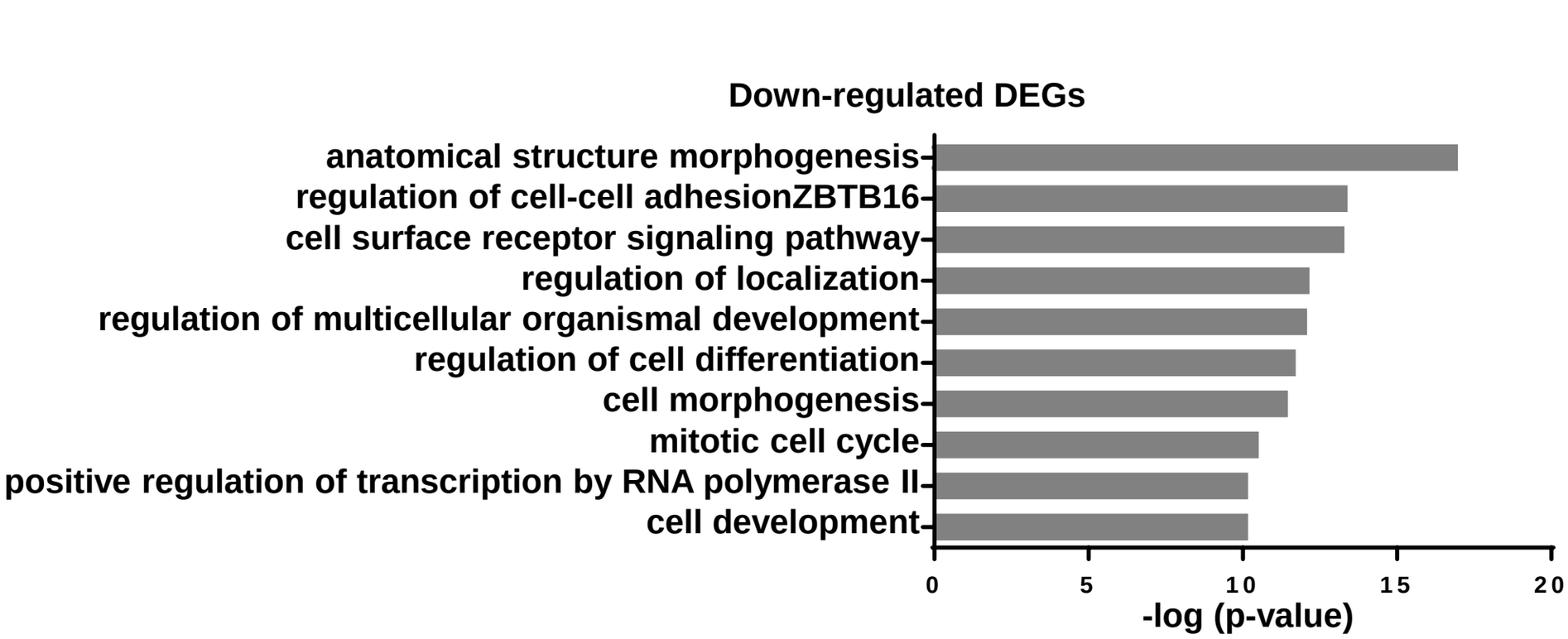
Non-competitive transplantation assay was performed by transplanting one million total bone marrow cells from donor mice of the indicated genotypes to recipients, which were treated with pIpC (every other day for a total of 3 doses) 9 weeks later to induce the expression of *Cbfb-MYH11* and deficiency of *Runx1*. These recipients were monitored for one year for hematopoietic malignancy. (A) Bar graphs showing the percentage of donor-derived cells in peripheral blood right before pIpC injection. (B) Bar graph showing the WBC and hematocrit of end stage *Mx1-CreCbfb*$^{+/56M}$ and *Runx1*$^{f/f}$*Mx1-CreCbfb*$^{+/56M}$ mice, together with related control (mean ± SEM). **: p < 0.05; ****: p < 0.0001*. (C) Representative FACS plots of c-Kit^{+} and Mac1^{+} cells in the peripheral blood of end stage *Mx1-CreCbfb*$^{+/56M}$ and *Runx1*$^{f/f}$*Mx1-CreCbfb*$^{+/56M}$ mice, together with related control. (D) At 12 months after pIpC injection as mentioned above, treated mice from *Runx1*$^{f/f}$*Cbfb*$^{+/56M}$ (n=4) *and Runx1*$^{f/f}$*Mx1-CreCbfb*$^{+/56M}$ (n=5) groups were sacrificed, and bone marrow cells were isolated. Genomic DNA from each bone marrow sample was extracted. The presence of un-excised *Runx1* and *Cbfb-MYH11* flox alleles in these DNA samples was detected by qPCR and normalized to un-excised *Runx1* (100%) and *Cbfb-MYH11* (50%) flox alleles in *Runx1*$^{f/f}$*Cbfb*$^{+/56M}$, respectively.

Supplemental Figure 7



Supplemental Figure 7: **(A-B)** The indicated groups of mice were treated with pIpC to induce the expression of Cbfb-MYH11 and/or Runx1 deficiency. At certain time points after pIpC treatment, the mice were sacrificed, and flow cytometry assays perform. (A) Representative FACS plots of bone marrow cells gated on LK cells from *control* and *Runx1$^{f/f}$Mx1-Cre* mice treated with pIpC for 4-5 weeks and 4 months. **(B)** Bar graph showing the percentages of indicated cell population in the bone marrow cell of mice 2-3 weeks, 4-5 weeks and 4 months after pIpC treatment (mean ± SEM). **: p < 0.05, **: p < 0.01*, ****: p < 0.001*, *****: p < 0.0001.*

Supplemental Figure 8



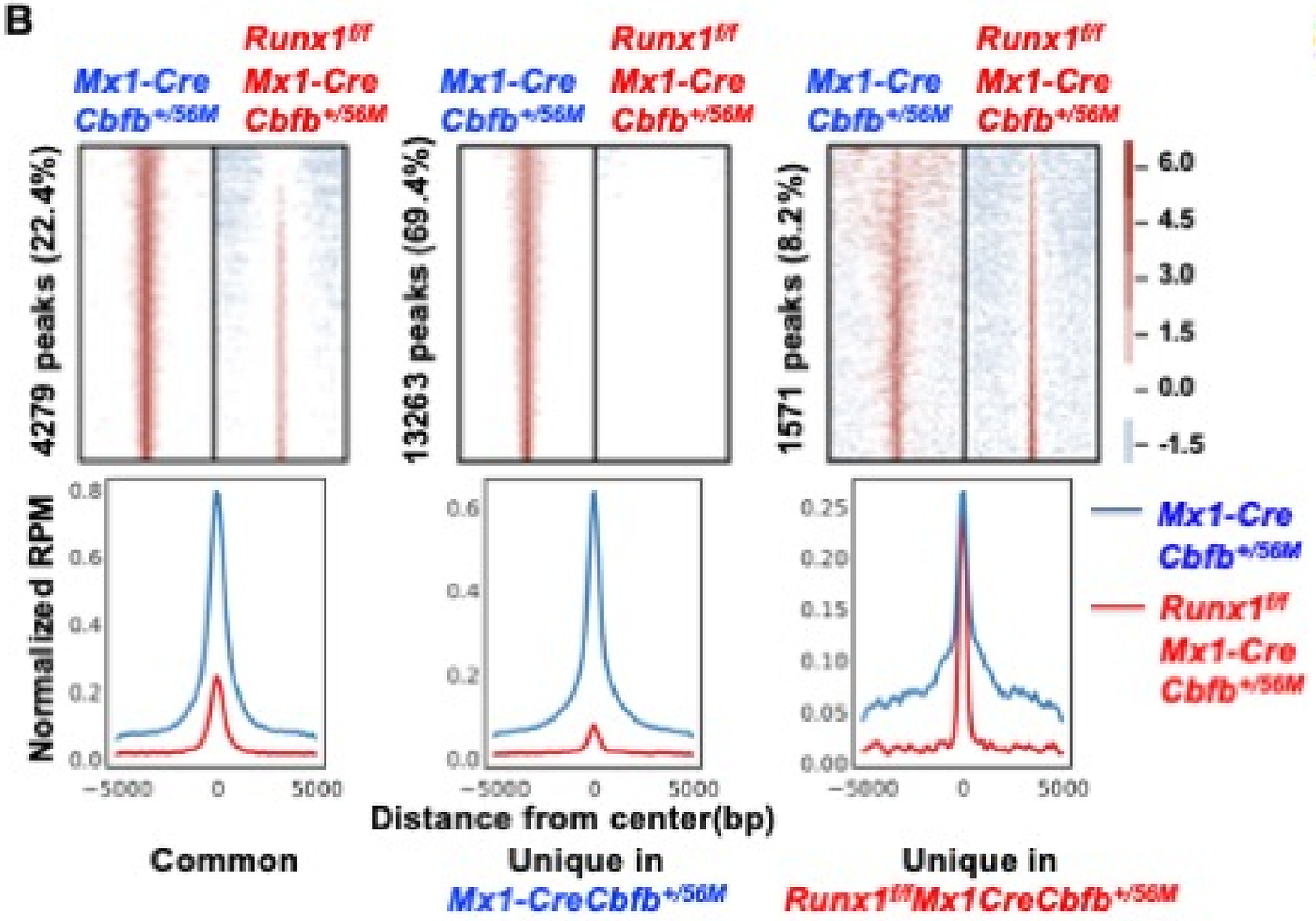
Supplemental Figure 8: Bioinformatic analysis of the differentially expressed genes (DEGs) from RNA-seq
(A-B) Gene ontology analysis of the down-regulated DEGs (A) and the up-regulated DEGs (B) shown in Figure 4B. (C) Upstream pathways associated with DEGs shown in Figure 4B, identified by Ingenuity pathway analysis. Gata2 was the highest scored upstream transcription factor.

Supplemental Figure 9

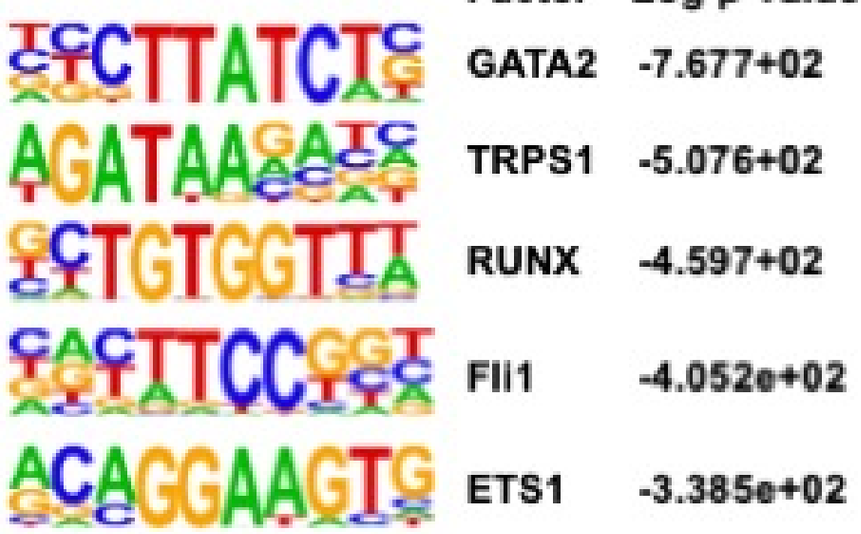
A

	Mx1-CreCbfb$^{+/56M}$		*Runx1*$^{f/f}$ *Mx1-CreCbfb*$^{+/56M}$	
Antibody	**peaks**	**Genes**	**peaks**	**Genes**
RUNX1	17542	9741	5844	4777
CBFβ	7824	5497	11421	7687
SMMHC	5546	4016	1397	1282
H3K27ac	23026	11962	21548	11501
H3K27me3	11678	8725	6601	5802

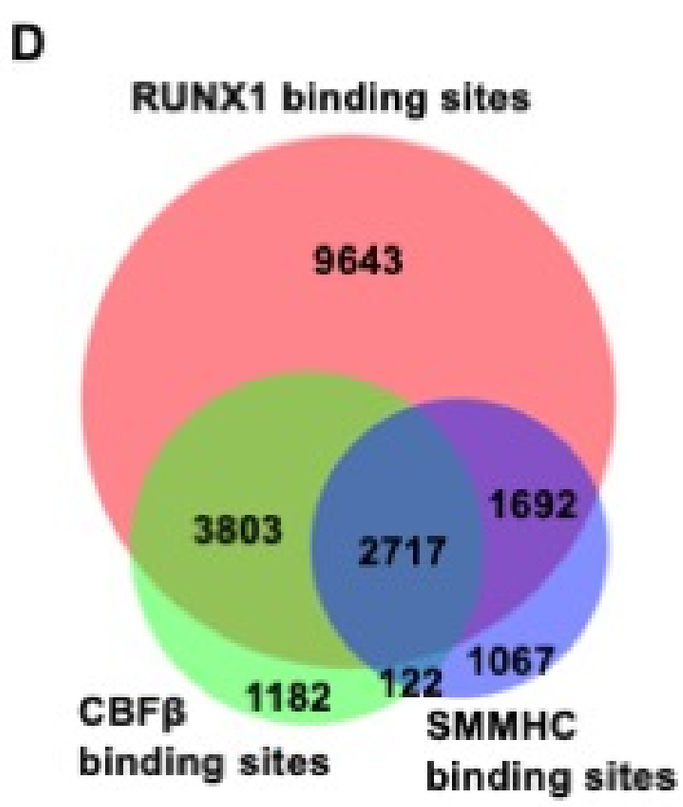
B



C



D



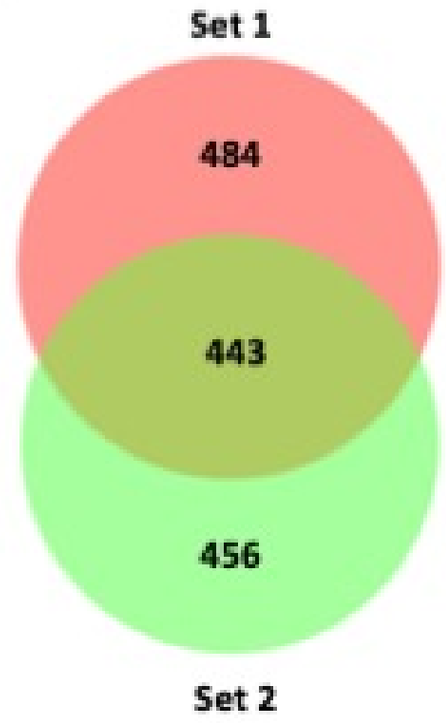
Supplemental Figure 9: Bioinformatic analysis of ChIC-seq data. (A) Summary of the number of binding sites (peaks) and genes assigned to the binding sites of each sample. (B) Heat map (upper panel) and average binding profile at the transcription start site (bottom panel) displaying peaks observed in cells of both genotypes (Common), peaks unique in *Mx1-CreCbfb*$^{+/56M}$ cells and peaks unique in *Runx1*$^{f/f}$*Mx1-CreCbfb*$^{+/56M}$ cells. (C) Top 5 motifs significantly enriched in peaks unique to *Mx1-CreCbfb*$^{+/56M}$ cells.(D) Venn diagrams representing the overlap of RUNX1, CBFβ and SMMHC binding sites in the AMP population from *Mx1-CreCbfb*$^{+/56M}$ mice with Intervene.

Supplemental Figure 10

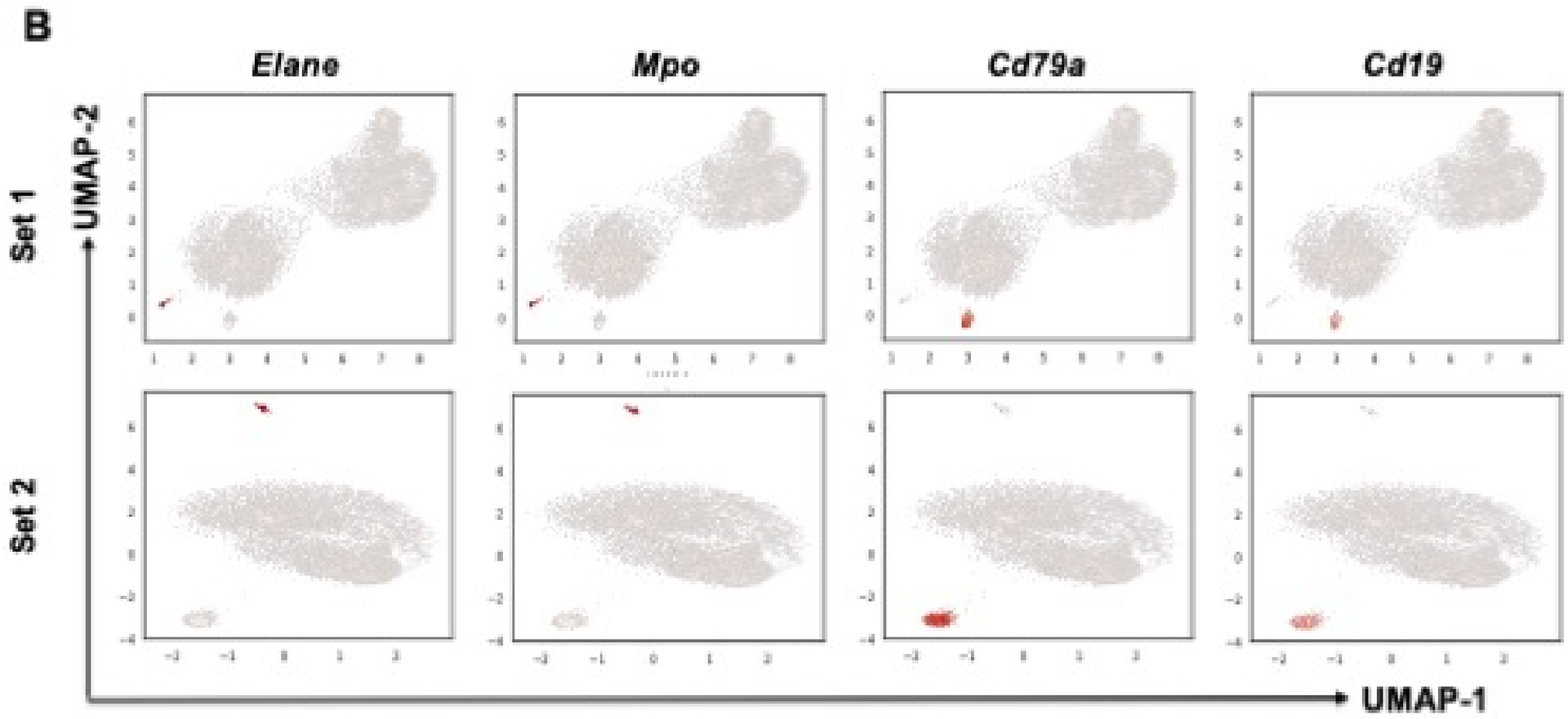
A

	Set 1		Set 2	
	Mx1-Cre Cbfb$^{+/56M}$	*Runx1*$^{f/f}$ *Mx1-Cre Cbfb*$^{+/56}$	*Mx1-Cre Cbfb*$^{+/56M}$	*Runx1*$^{f/f}$ *Mx1-Cre Cbfb*$^{+/56}$
Estimated Number of Cells	6456	7636	5963	7121
Mean Reads per Cell	14393	15249	19961	19609
Median Genes per Cell	1687	1997	1862	1906
Number of Reads	92927458	1.16E+08	1.19E+08	1.4E+08
Valid Barcodes	97.50%	97.50%	97.40%	97.20%
Sequencing Saturation	51.60%	45.10%	56.90%	55.60%
Q30 Bases in Barcode	97.30%	97.40%	97.40%	97.40%
Q30 Bases in RNA Read	81.90%	82.10%	82.00%	82.20%
Q30 Bases in Sample Index	97.10%	96.80%	97.20%	96.60%
Q30 Bases in UMI	97.20%	97.20%	97.20%	97.20%
Reads Mapped to Genome	90.00%	90.00%	89.80%	90.40%
Reads Mapped Confidently to Genome	86.50%	86.70%	86.10%	86.80%
Reads Mapped Confidently to Intergenic Regions	2.00%	2.00%	2.10%	2.10%
Reads Mapped Confidently to Intronic Regions	9.60%	9.20%	10.30%	9.80%
Reads Mapped Confidently to Exonic Regions	74.90%	75.60%	73.70%	74.80%
Reads Mapped Confidently to Transcriptome	72.20%	73.00%	70.90%	71.90%
Reads Mapped Antisense to Gene	0.90%	0.90%	1.10%	1.10%
Fraction Reads in Cells	93.50%	91.80%	92.70%	90.30%
Total Genes Detected	15255	15184	15717	15236
Median UMI Counts per Cell	4368	5192	5081	5106

C



B



Supplemental Figure 10: Bioinformatic analysis of single cell RNA-seq (scRNA-seq) data. (A) Statistics of the scRNA-seq data. (B) Feature plot depicting the expression of selected genes across different cell populations (Red is high, grey is low) in both sets of scRNA-seq experiments. (C) Venn diagrams representing the overlap of differentially expressed genes (p value < $1e^{-5}$, absolute fold change ≥ 2) observed between *Mx1-CreCbfb*$^{+/56M}$ *and Runx1*$^{f/f}$*Mx1-CreCbfb*$^{+/56M}$ cells in both sets of experiments.