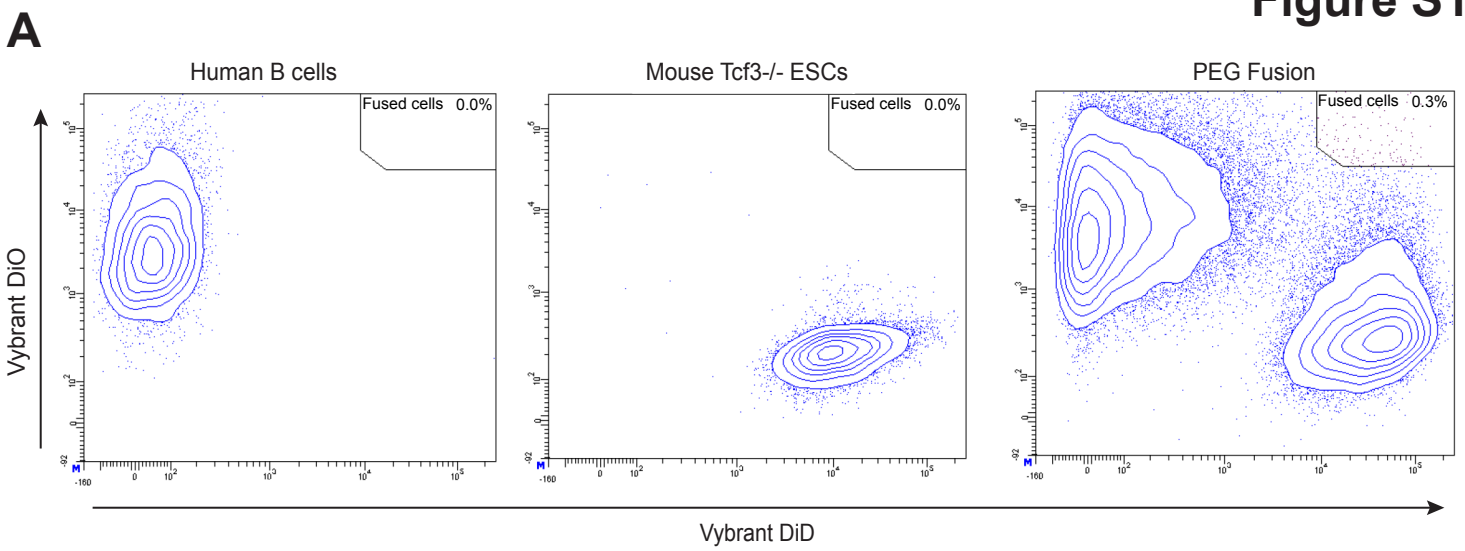
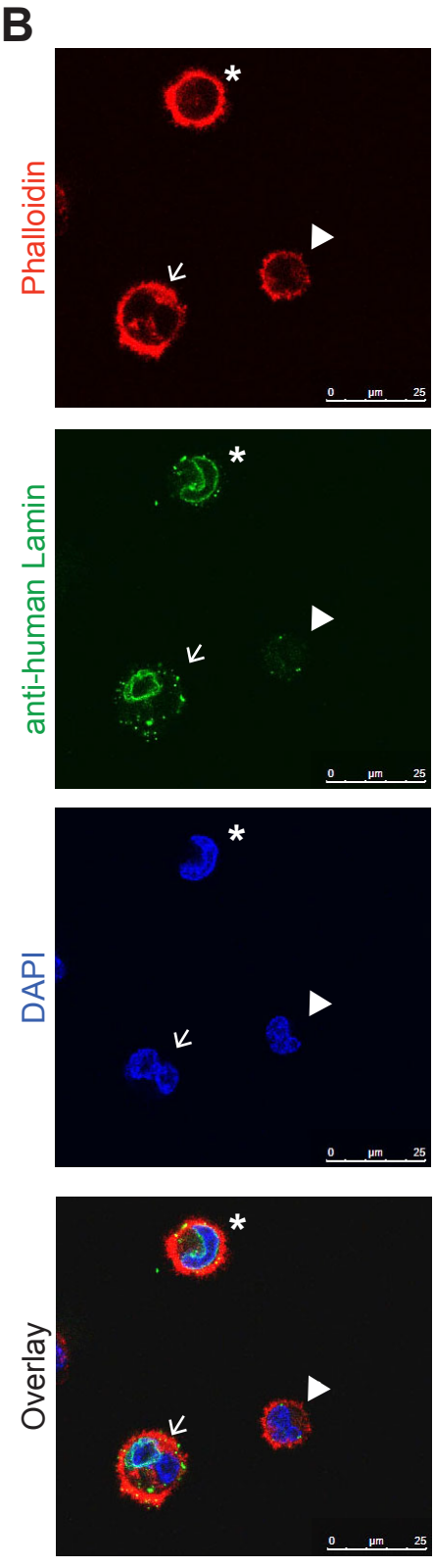
Figure S1


A
Human B cells
Mouse Tcf3-/- ESCs
PEG Fusion
Fused cells 0.0%
Fused cells 0.0%
Fused cells 0.3%
Vybrant DiO
Vybrant DiD


B
Phalloidin
anti-human Lamin
DAPI
Overlay

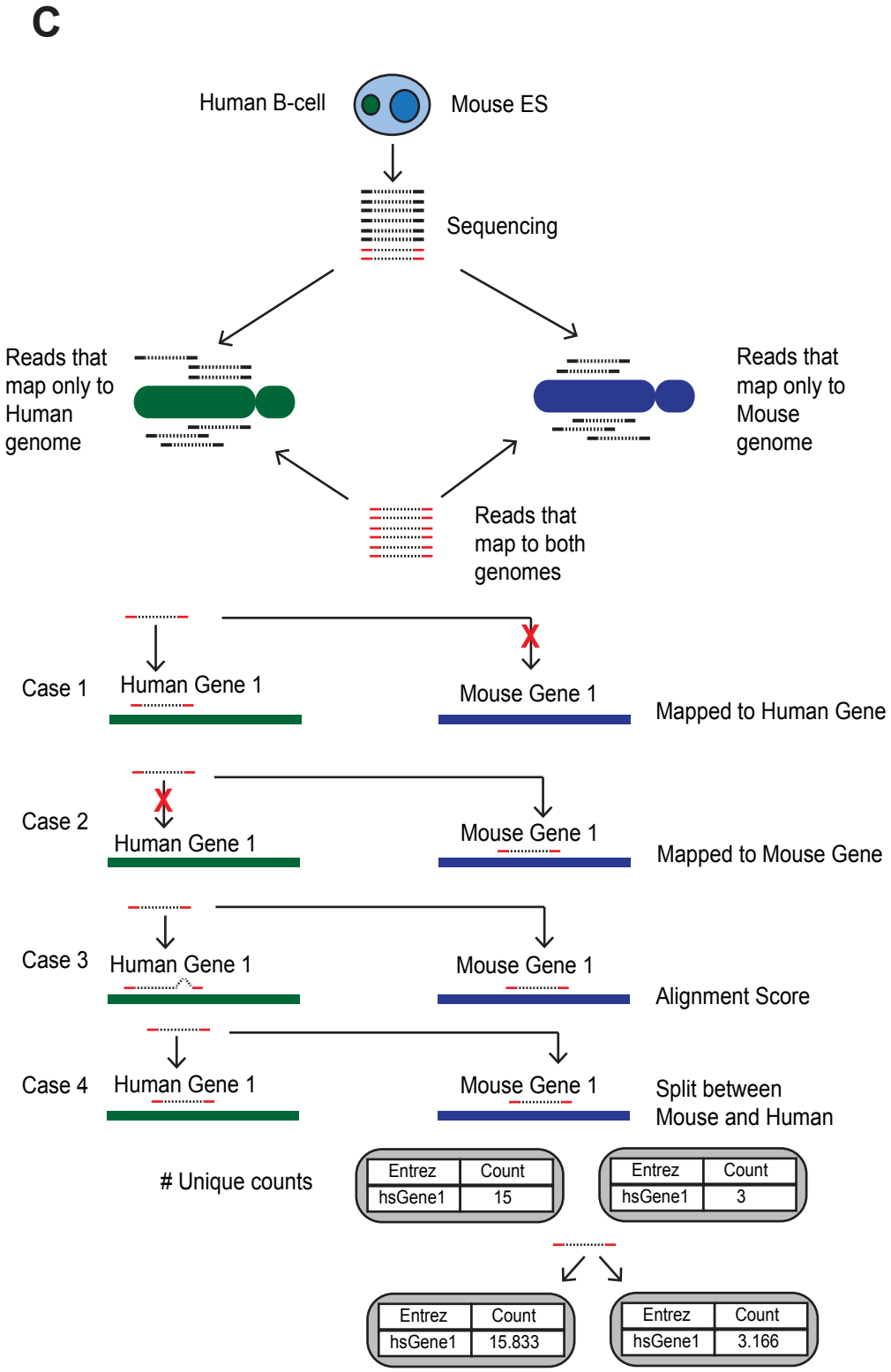
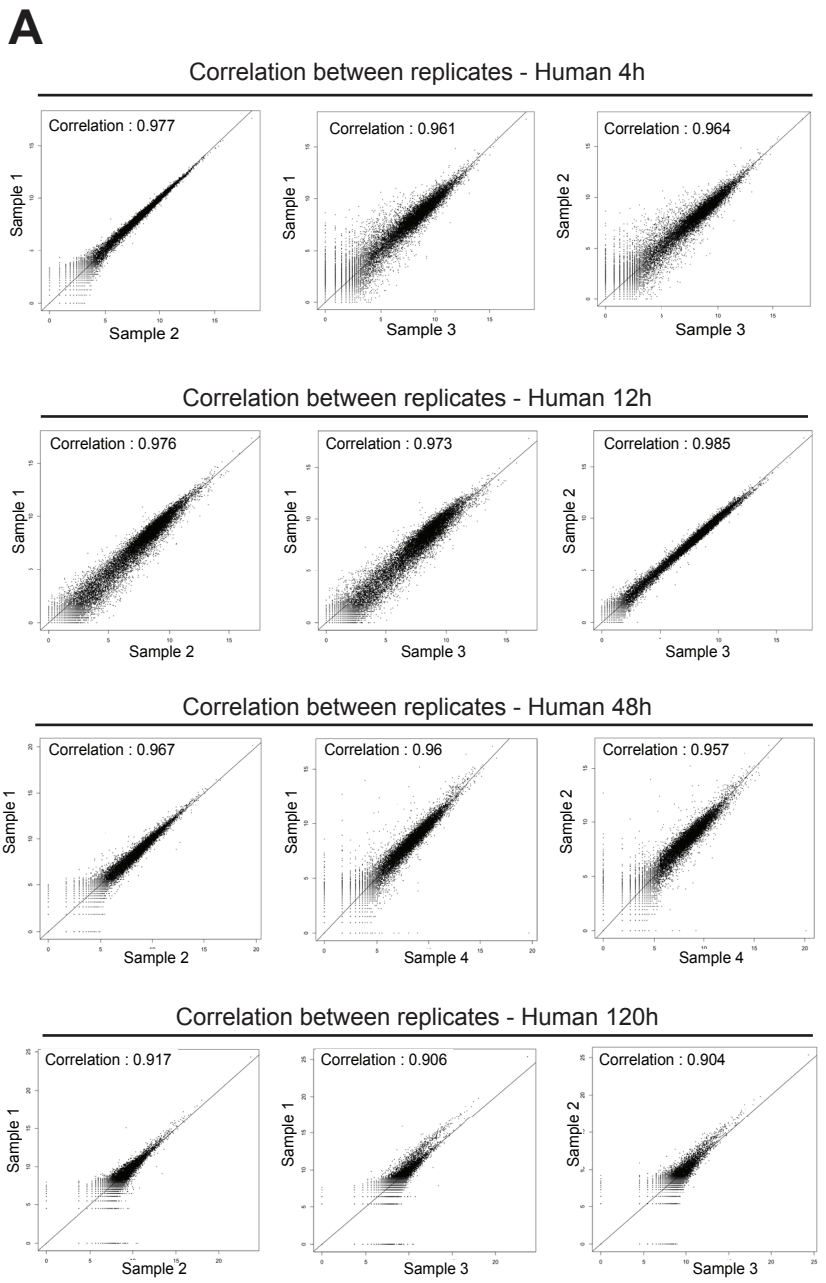
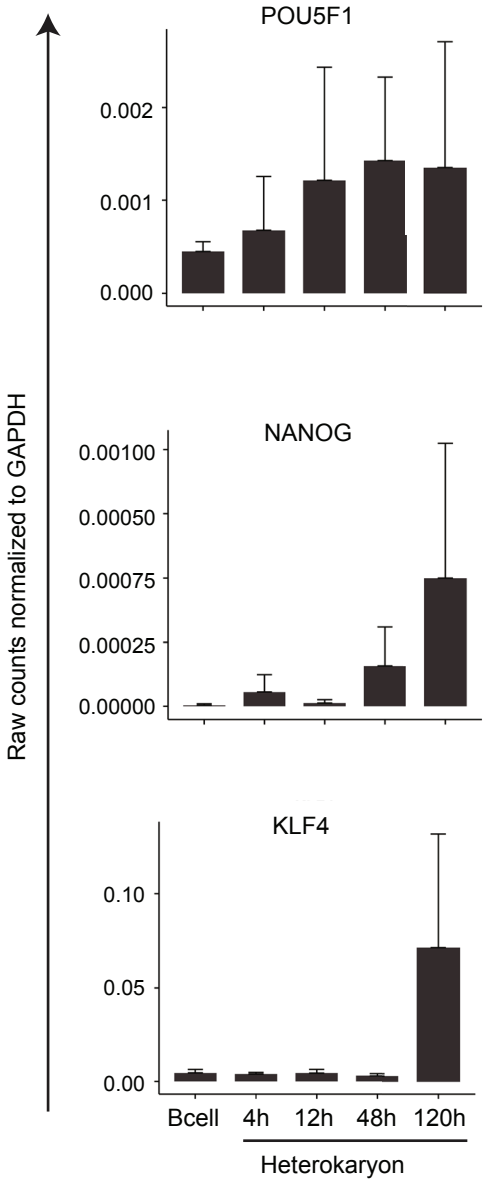

C
Human B-cell
Mouse ES
Sequencing
Reads that map only to Human genome
Reads that map only to Mouse genome
Reads that map to both genomes
Case 1
Human Gene 1
Mouse Gene 1
Mapped to Human Gene
Case 2
Human Gene 1
Mouse Gene 1
Mapped to Mouse Gene
Case 3
Human Gene 1
Mouse Gene 1
Alignment Score
Case 4
Human Gene 1
Mouse Gene 1
Split between Mouse and Human
Unique counts
Entrez
Count
hsGene1
15
Entrez
Count
hsGene1
3
Entrez
Count
hsGene1
15.833
Entrez
Count
hsGene1
3.166

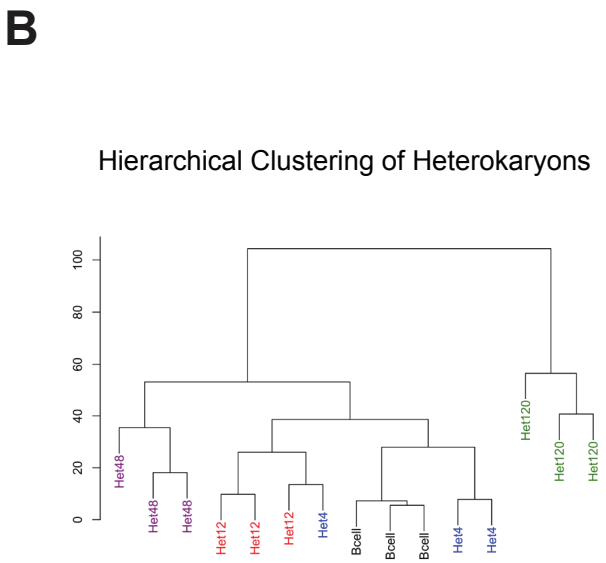
Figure S1. mRNA sequence read mapping of bi-species heterokaryons sorted by flow cytometry. Related to Figure 1.

(A) Human EBV-B cells labeled with Vybrant DiO and Mouse *Tcf7l1*-/- ESCs labeled with Vybrant DiD. Representative FACS plot of cells sorted 4 hours after PEG fusion. **(B)** Immunostaining of human EBV-B cells fused with mouse *Tcf7l1*-/- ESCs and sorted 4 hrs after fusion. One representative picture area was chosen to show unfused and hybrid cells. Phalloidin (in Red) stains the actin filaments. anti-human Lamin (in green) stains the EBV-B cell nuclear membrane. DAPI (in blue) stains the nucleus. The asterisk (*) points to a human EBV-B cell, the triangle (∇) points to a mouse ESC and the arrow (←) points to a heterokaryon. Scale bar 0-25 um. **(C)** Schematic for the mapping of the mouse and human reads. First the reads were mapped to the mouse and human genomes with high stringency. Multi-mapping reads were assigned to either the mouse or human genome using known gene information. If a read mapped to a gene in one of the genomes, but non-gene region in the other, the reads were assigned to the appropriate gene. Next mapping quality was considered to assign the read to the genome with the higher quality mapping. Finally, the reads that mapped perfectly to genes in both genomes were split between each of the genomes by considering how many unique reads had been already mapped to each gene.

Figure S2


A
Correlation between replicates - Human 4h
Correlation : 0.977
Correlation : 0.961
Correlation : 0.964
Correlation between replicates - Human 12h
Correlation : 0.976
Correlation : 0.973
Correlation : 0.985
Correlation between replicates - Human 48h
Correlation : 0.967
Correlation : 0.96
Correlation : 0.957
Correlation between replicates - Human 120h
Correlation : 0.917
Correlation : 0.906
Correlation : 0.904
Sample 1
Sample 2
Sample 3
Sample 4


C
POU5F1
NANOG
KLF4
Raw counts normalized to GAPDH
Bcell
4h
12h
48h
120h
Heterokaryon


B
Hierarchical Clustering of Heterokaryons
Het48
Het12
Het4
Bcell
Het120

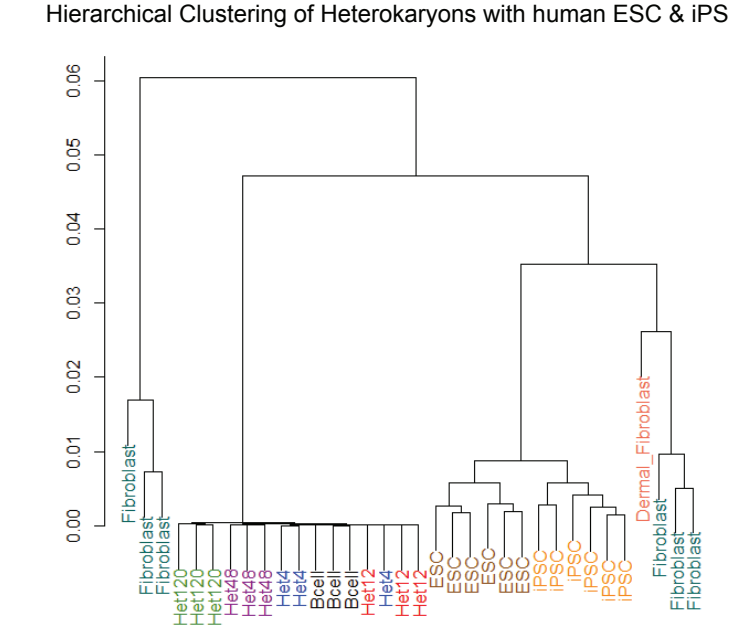
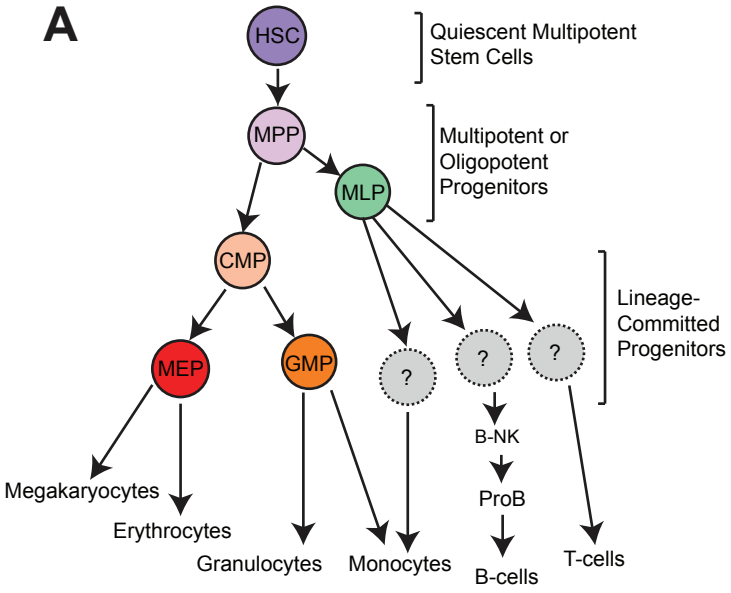

D
Hierarchical Clustering of Heterokaryons with human ESC & iPS
Fibroblast
Het120
Het48
Het4
Bcell
Het12
ESC
iPSC
Dermal_Fibroblast

Figure S2. Correlation of human gene expression in the heterokaryon sample replicates. Related to Figure 1.

(A) Comparison of replicates for human gene expression during cell-fusion mediated reprogramming over 5 days. Spearman correlation between each sample is shown. Sample numbers are shown on the x and y axis, and the plots show the correlation shown for samples 4 hours after fusion, 12 hours after fusion, 48 hours after fusion and 120 hours after fusion. **(B)** Hierarchical clustering of heterokaryons samples show strong separation according to time-points. Results show that almost all the sample replicates separated according to their respective time-points after cell fusion. Plotting was performed using the complete agglomeration method using the expression values of the top-2000 genes with the highest variance across all the samples. **(C)** Expression of a representative set of human pluripotency markers in the heterokaryon samples during reprogramming. Results show the raw count of *POU5F1*, *NANOG*, and *KLF4* normalized to *GAPDH*, prior to fusion (B-cell) and 4h, 12h, 48h, and 120h, after cell fusion. **(D)** Hierarchical clustering was performed between heterokaryon samples, fibroblasts and iPSC samples from a reference dataset (GSE41716). Results show a strong separation between heterokaryon samples and both the fibroblast and iPSC samples, suggesting a significant difference in the gene expression patterns. Plotting was performed using the complete agglomeration method using the expression values of the top-2000 genes with the highest variance across all samples.

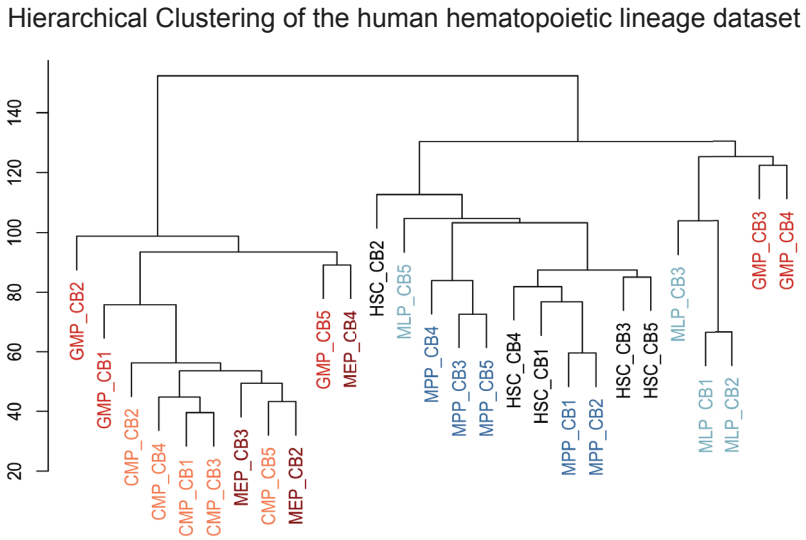
Figure S3



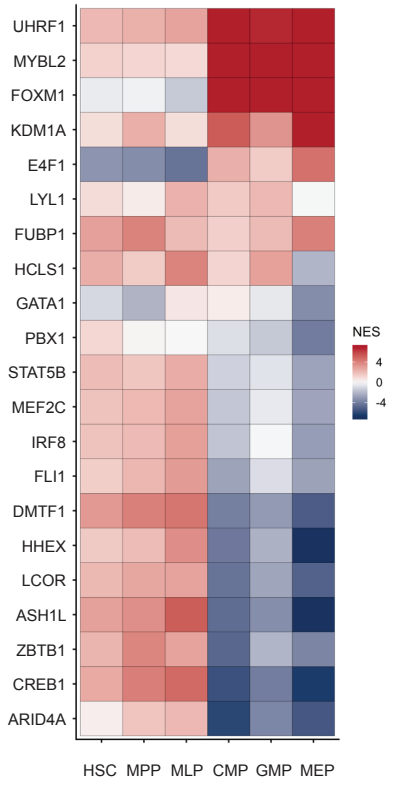
B

HSC	Lin-CD34+CD38-CD45RA-CD90+CD49f+	Hematopoietic Stem Cell
MPP	Lin-CD34+CD38-CD45RA-CD90-CD49f-	Multipotent Progenitor
MLP	Lin-CD34+CD38-CD45RA+CD10+	Multi-Lymphoid Progenitor
CMP	Lin-CD34+CD38+CD10-CD7-CD45RA-CD135+	Common Myeloid Progenitor
MEP	Lin-CD34+CD38+CD10-CD7-CD45RA+CD135+	Megakaryocyte Erythroid Progenitor
GMP	Lin-CD34+CD38+CD10-CD7-CD45RA-CD135-	Granulocyte Monocyte Progenitor

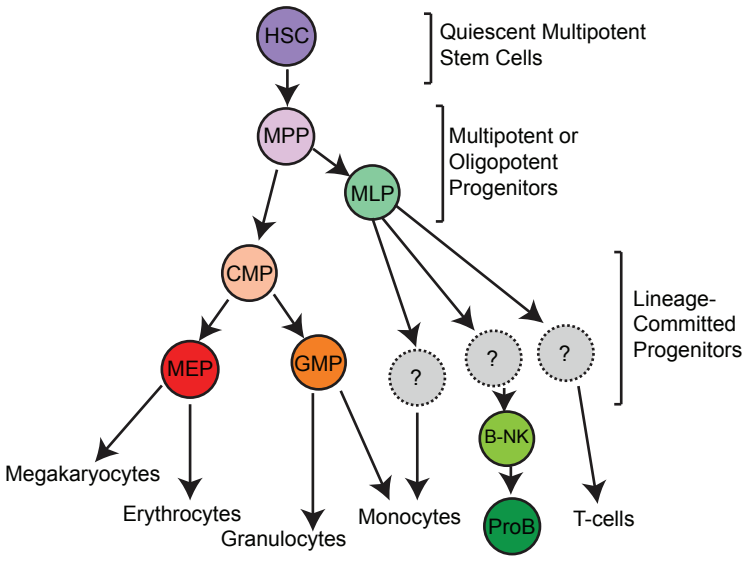
C



D



E



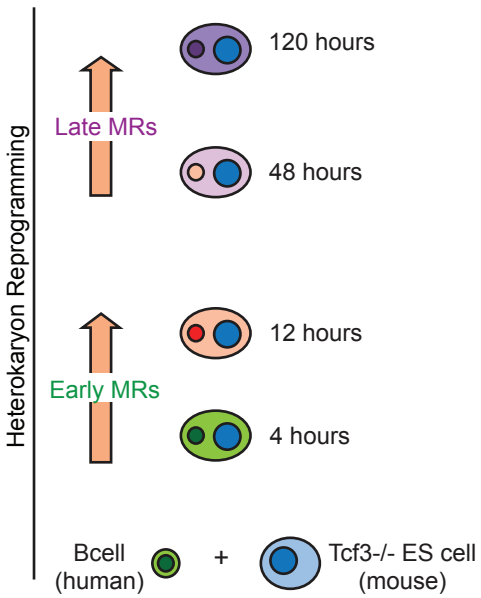
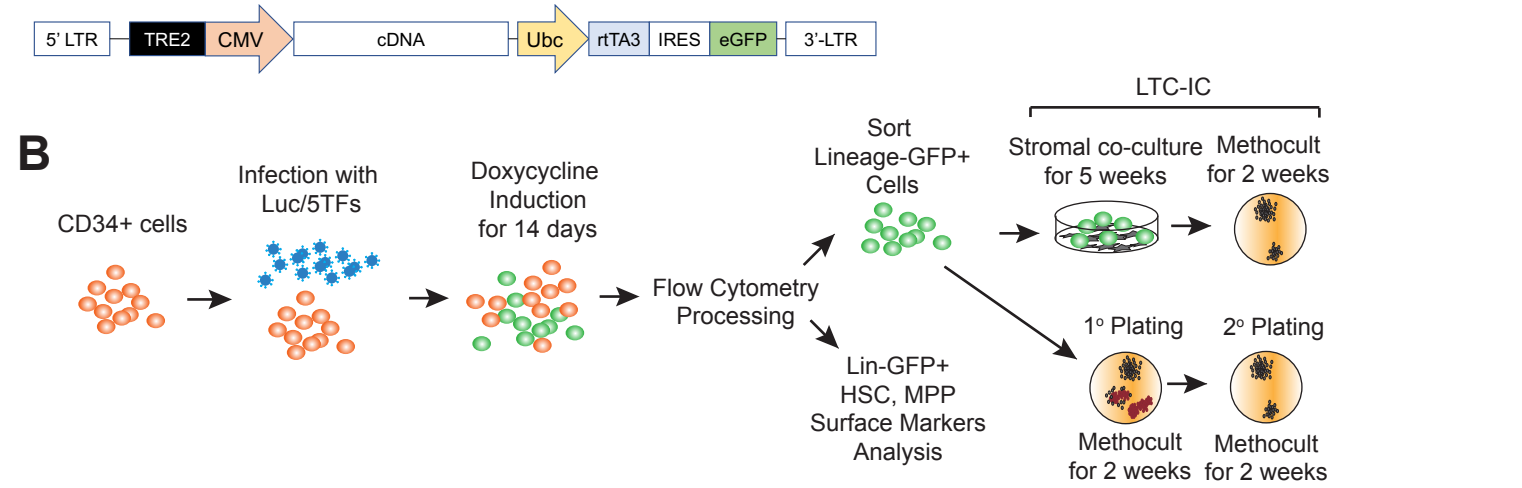
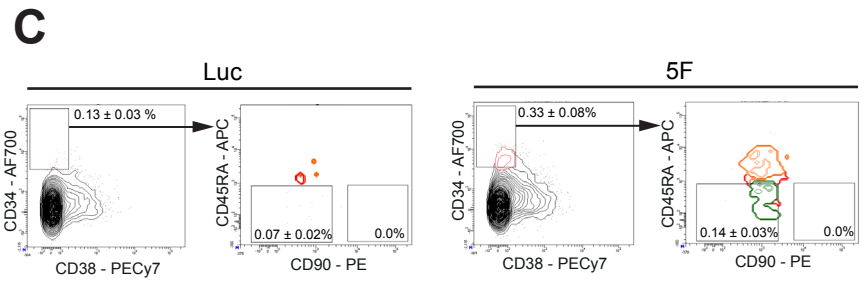


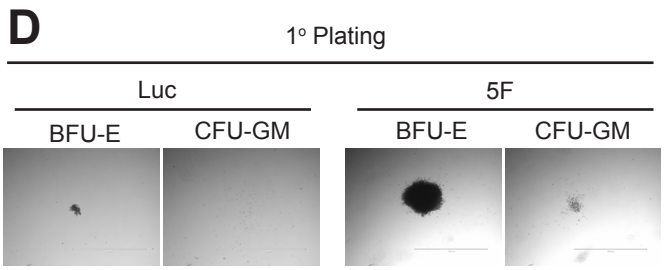
Figure S3. A human hematopoietic reference dataset used to track the reprogramming of the human EBV-B nuclei in the heterokaryons. Related to Figure 3.

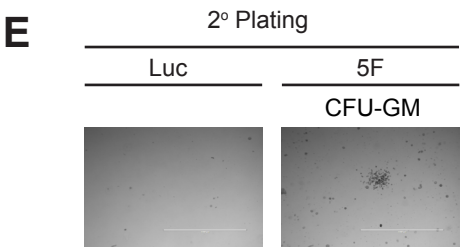
(A) Hierarchical model of human hematopoietic system, adapted from (Doulatov et al., 2010; Laurenti et al., 2013). The samples consisted of a population of quiescent HSC, which are multi-potent stem cells that can differentiate into all blood cell types of the hematopoietic lineage (Doulatov et al., 2012). The HSCs after exiting quiescence form proliferating multipotent progenitor (MPP) cells that can differentiate into committed progenitors of the myeloid and the lymphoid lineages (Notta et al., 2011). In the lymphoid branch, the multi-lymphoid progenitors (MLP) can differentiate through as yet unknown intermediate progenitors into T lymphocytes and precursors of B lymphocytes and natural killer cell (B-NK) (Doulatov et al., 2010). MLPs are not entirely committed to the lymphoid lineage and can also differentiate into monocytes from the myeloid lineage (Doulatov et al., 2010). MLP population is highly similar to HSCs and MPPs (Laurenti et al., 2013), and can differentiate into the myeloid and lymphoid lineage, with a bias toward the lymphoid lineage. In the myeloid branch, the common myeloid progenitors (CMP) are committed to the myeloid lineage and can differentiate into the myeloid progenitors of granulocytes and monocytes (GMP), megakaryocytes and erythrocytes (MEP). **(B)** The table includes the surface markers that were used to sort the different hematopoietic fractions in the human hematopoietic dataset. **(C)** Hierarchical clustering of the hematopoietic lineage data set shows significant variations between cells of the same type. Plotting was performed using the complete agglomeration method, with the expression values of the top-2000 genes with the highest variance across all the samples. **(D)** VIPER-predicted activity of a representative set of human genes in the reference human hematopoietic stem and progenitor populations. Positive NES values indicating active MRs are shown in red, and negative NES values indicating silenced MRs are shown in blue. **(E)** Model of heterokaryon reprogramming, shown alongside hierarchical model of hematopoietic development, adapted from (Doulatov et al., 2010; Laurenti et al., 2013).

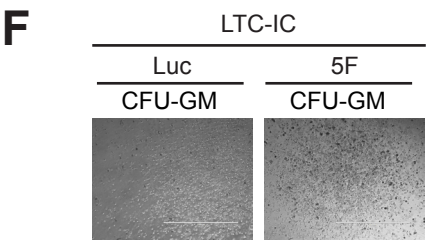
Figure S4

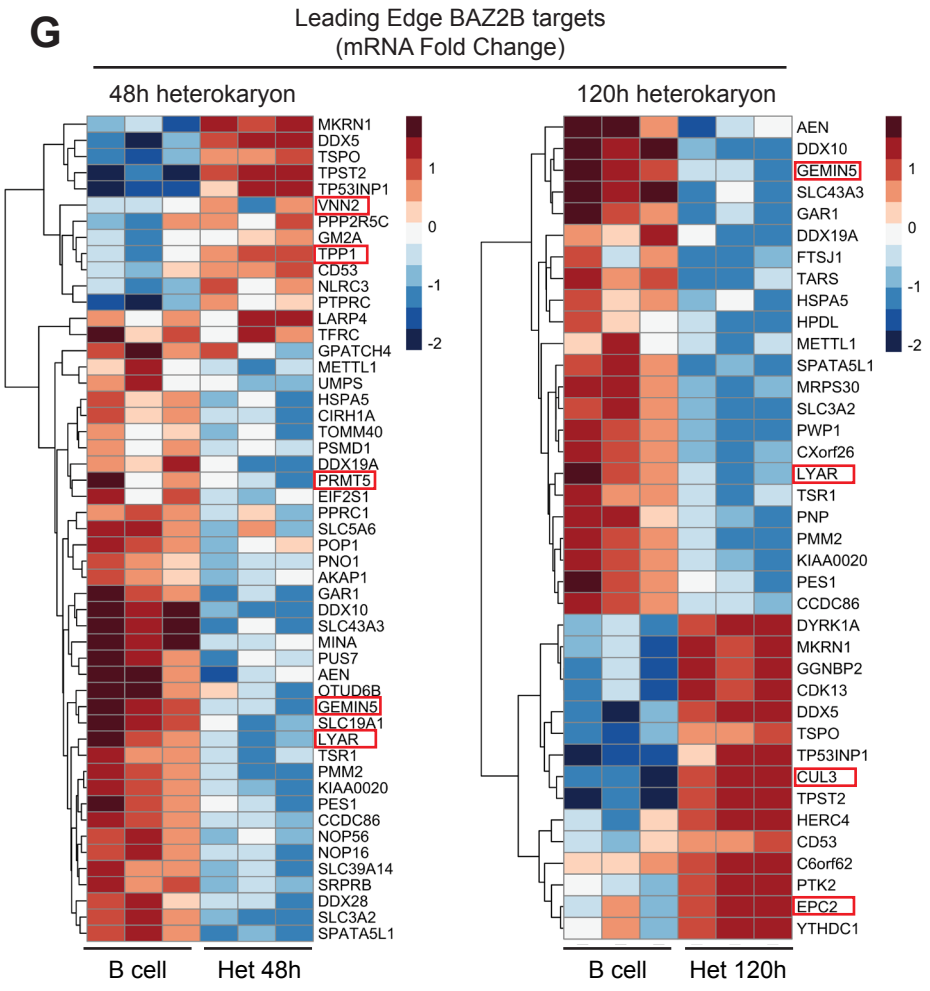

A
5' LTR
TRE2
CMV
cDNA
Ubc
rtTA3
IRES
eGFP
3'-LTR
B
CD34+ cells
Infection with Luc/5TFs
Doxycycline Induction for 14 days
Flow Cytometry Processing
Sort Lineage-GFP+ Cells
LTC-IC
Stromal co-culture for 5 weeks
Methocult for 2 weeks
Lin-GFP+ HSC, MPP Surface Markers Analysis
1° Plating
2° Plating
Methocult for 2 weeks
Methocult for 2 weeks

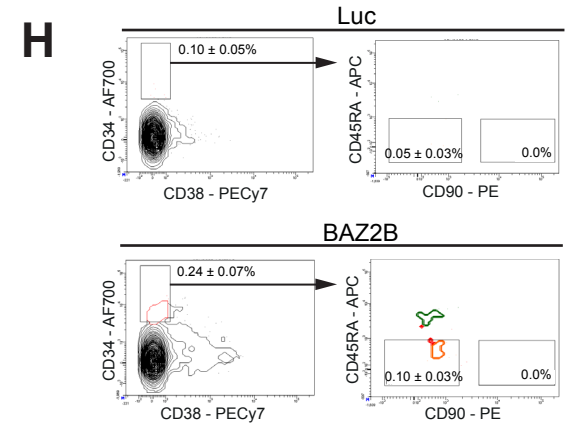

C
Luc
5F
0.13 ± 0.03 %
0.07 ± 0.02%
0.0%
0.33 ± 0.08%
0.14 ± 0.03%
0.0%
CD34 - AF700
CD38 - PECy7
CD45RA - APC
CD90 - PE

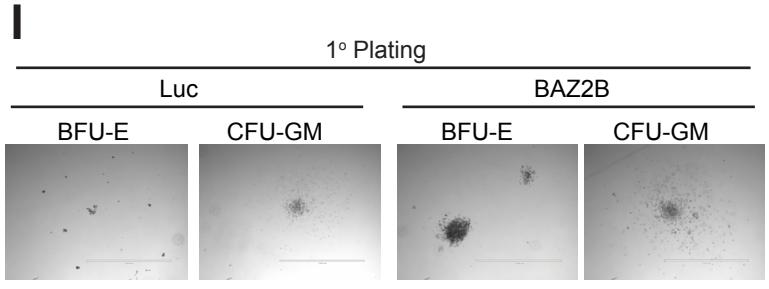

D
1° Plating
Luc
5F
BFU-E
CFU-GM
BFU-E
CFU-GM


E
2° Plating
Luc
5F
CFU-GM


F
LTC-IC
Luc
5F
CFU-GM
CFU-GM


G
Leading Edge BAZ2B targets
(mRNA Fold Change)
48h heterokaryon
120h heterokaryon
MKRN1
DDX5
TSPO
TPST2
TP53INP1
VNN2
PPP2R5C
GM2A
TPP1
CD53
NLRC3
PTPRC
LARP4
TFRC
GPATCH4
METTL1
UMPS
HSPA5
CIRH1A
TOMM40
PSMD1
DDX19A
PRMT5
EIF2S1
PPRC1
SLC5A6
POP1
PNO1
AKAP1
GAR1
DDX10
SLC43A3
MINA
PUS7
AEN
OTUD6B
GEMIN5
SLC19A1
LYAR
TSR1
PMM2
KIAA0020
PES1
CCDC86
NOP56
NOP16
SLC39A14
SRPRB
DDX28
SLC3A2
SPATA5L1
AEN
DDX10
GEMIN5
SLC43A3
GAR1
DDX19A
FTSJ1
TARS
HSPA5
HPDL
METTL1
SPATA5L1
MRPS30
SLC3A2
PWP1
CXorf26
LYAR
TSR1
PNP
PMM2
KIAA0020
PES1
CCDC86
DYRK1A
MKRN1
GGNBP2
CDK13
DDX5
TSPO
TP53INP1
CUL3
TPST2
HERC4
CD53
C6orf62
PTK2
EPC2
YTHDC1
1
0
-1
-2
B cell
Het 48h
B cell
Het 120h


H
Luc
0.10 ± 0.05%
0.05 ± 0.03%
0.0%
BAZ2B
0.24 ± 0.07%
0.10 ± 0.03%
0.0%
CD34 - AF700
CD38 - PECy7
CD45RA - APC
CD90 - PE


I
1° Plating
Luc
BAZ2B
BFU-E
CFU-GM
BFU-E
CFU-GM

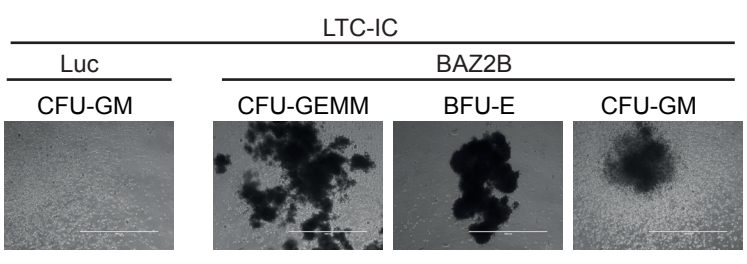
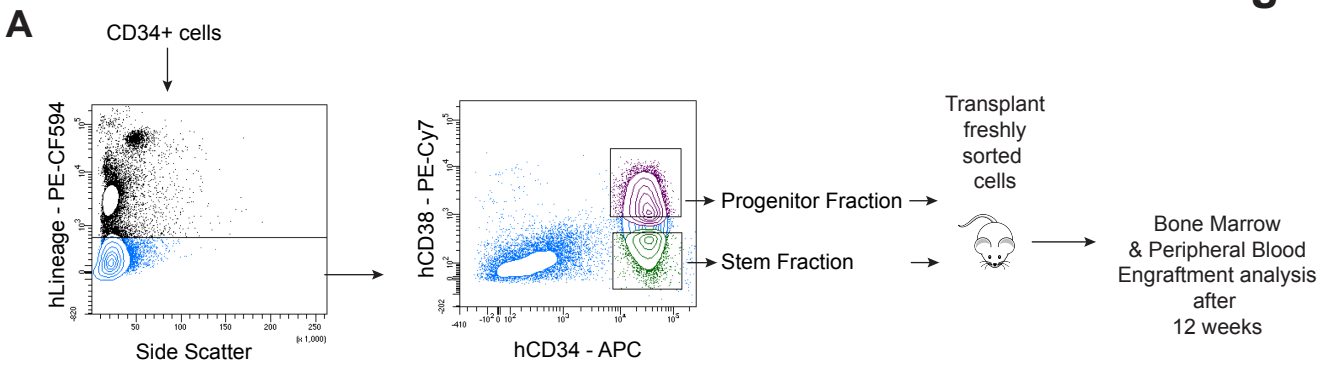
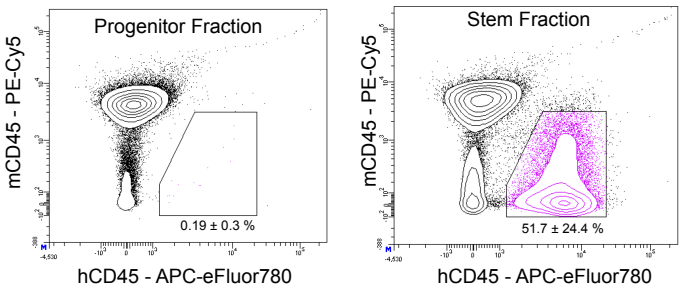
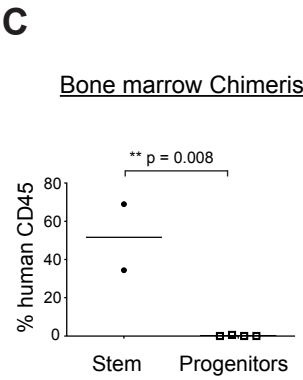

J
LTC-IC
Luc
BAZ2B
CFU-GM
CFU-GEMM
BFU-E
CFU-GM

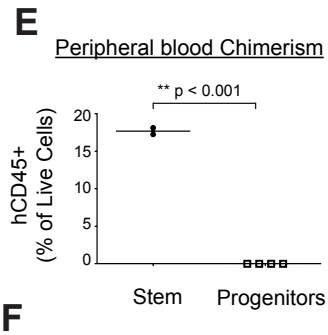
Figure S4. Overexpression of the combination of 5 transcription factors or of *BAZ2B* shows enhanced stemness and clonogenicity of CD34+ hematopoietic progenitors. Related to Figure 4.
(A) Map of the inducible lentiviral vector used for cloning all the cDNAs. **(B)** Schematic showing the experimental workflow of the transduction of CD34+ hematopoietic stem and progenitor cells with Luciferase or 5-MR cocktail or *BAZ2B*. **(C-F)** CD34+ cells transduced with Luciferase or 5MR. **(C)** Representative FACS plots on day 14 for Lin-GFP+CD34+CD38-CD45RA-CD90- MPPs and Lin-GFP+CD34+CD38-CD45RA-CD90+ HSCs. Percentage fractions for each gate normalized to Lineage-GFP+ fraction represented as mean ± SD from N = 5 donors **(D)** Representative colony images of BFU-E and CFU-GM from the primary Methocult plating. Scale bar 0-1000um. **(E)** Representative colony images of CFU-GM from the secondary Methocult plating. Scale bar 0-1000um. **(F)** Representative colony images of CFU-GM from the Methocult colonies of the LTC-IC assay. Scale bar 0-1000um. **(G)** mRNA fold change of the GSEA leading edge targets of BAZ2B in the heterokaryons at 48 hours and 120 hours after fusion. Differential expression was calculated using EdgeR, and shown genes are in the leading edge and have an FDR adjusted pvalue < 0.01. Boxed genes are relevant key factors discussed in the text. **(H-J)** CD34+ cells transduced with Luciferease or BAZ2B **(H)** Representative FACS plots on day 14 for Lin-GFP+CD34+CD38-CD45RA-CD90- MPPs and Lin-GFP+CD34+CD38-CD45RA-CD90+ HSCs. Percentage fractions for each gate normalized to Lineage-GFP+ fraction represented as mean ± SD from N = 5 donors **(I)** Representative colony images of BFU-E and CFU-GM from the primary methocult plating. Scale bar 0-1000um. **(J)** Representative colony images of BFU-E, CFU-GM and CFU-GEMM from the LTC-IC assay. Scale bar 0-1000um.

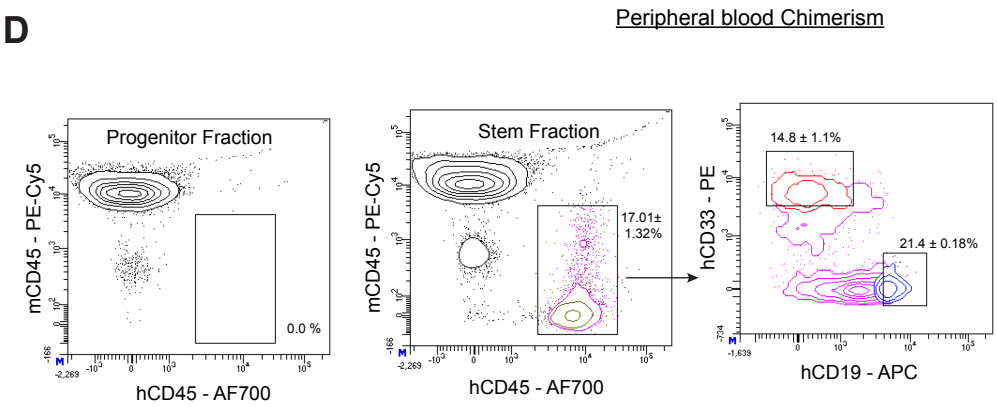
Figure S5

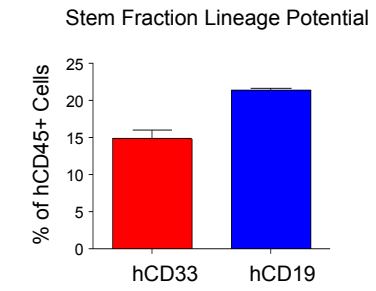

A
CD34+ cells
hLineage - PE-CF594
Side Scatter
hCD38 - PE-Cy7
hCD34 - APC
Progenitor Fraction
Stem Fraction
Transplant freshly sorted cells
Bone Marrow & Peripheral Blood Engraftment analysis after 12 weeks

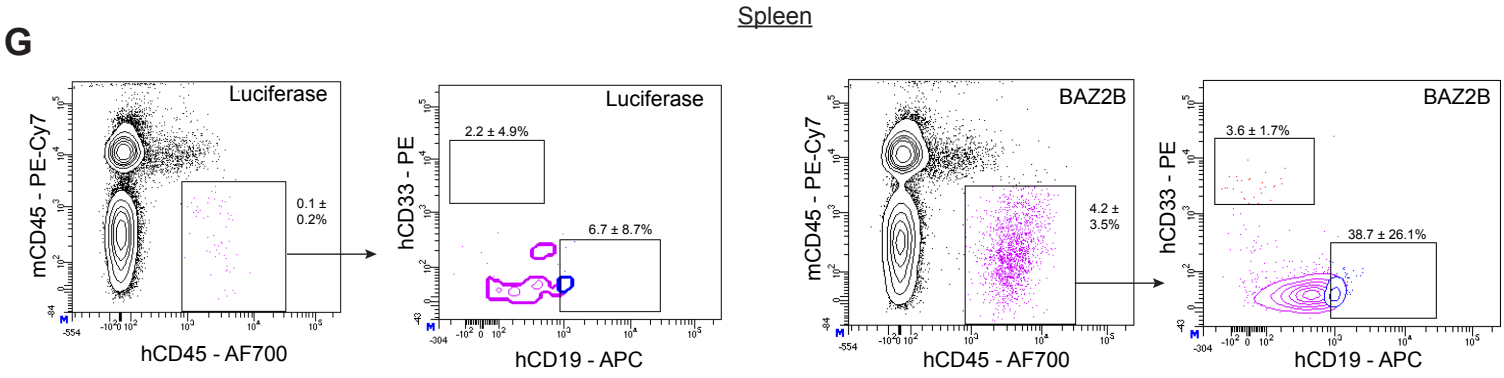

B
Bone marrow Chimerism
Progenitor Fraction
0.19 ± 0.3 %
Stem Fraction
51.7 ± 24.4 %
mCD45 - PE-Cy5
hCD45 - APC-eFluor780


C
Bone marrow Chimerism
** p = 0.008
% human CD45
Stem
Progenitors


E
Peripheral blood Chimerism
** p < 0.001
hCD45+ (% of Live Cells)
Stem
Progenitors


D
Peripheral blood Chimerism
Progenitor Fraction
0.0 %
Stem Fraction
17.01± 1.32%
14.8 ± 1.1%
21.4 ± 0.18%
mCD45 - PE-Cy5
hCD45 - AF700
hCD33 - PE
hCD19 - APC


F
Stem Fraction Lineage Potential
% of hCD45+ Cells
hCD33
hCD19


G
Spleen
Luciferase
0.1 ± 0.2%
2.2 ± 4.9%
6.7 ± 8.7%
BAZ2B
4.2 ± 3.5%
3.6 ± 1.7%
38.7 ± 26.1%
mCD45 - PE-Cy7
hCD45 - AF700
hCD33 - PE
hCD19 - APC

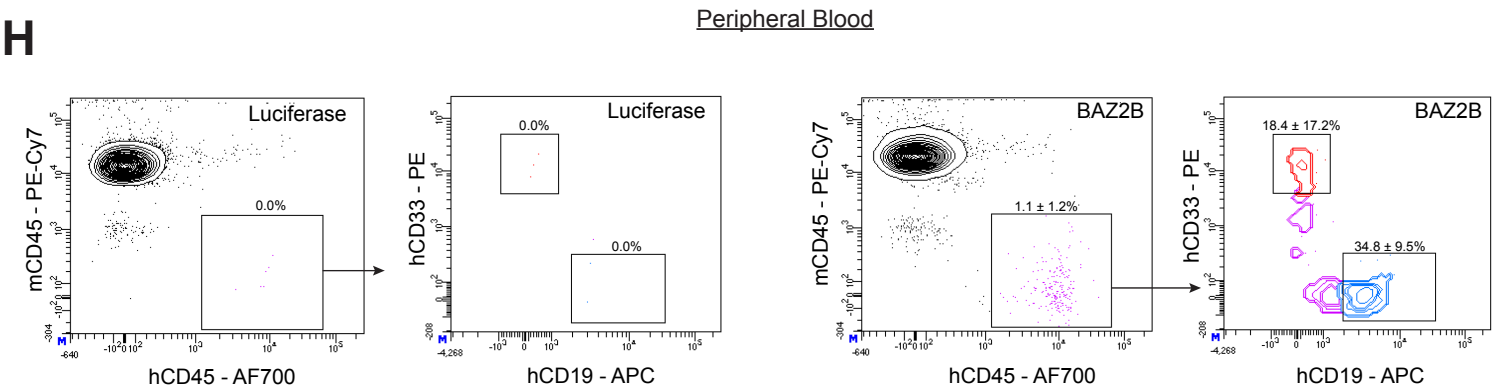

H
Peripheral Blood
Luciferase
0.0%
0.0%
0.0%
BAZ2B
1.1 ± 1.2%
18.4 ± 17.2%
34.8 ± 9.5%
mCD45 - PE-Cy7
hCD45 - AF700
hCD33 - PE
hCD19 - APC

Figure S5. Long-term engraftment potential of uncultured stem and progenitor fractions. Related to Figure 5 and 6.

(A) Lin-CD34+CD38- stem fraction and the Lin-CD34+CD38+ progenitor fraction were freshly sorted and transplanted intra-femorally in NSG mice. Human hematopoietic engraftment was analysed after 12 weeks. (**B**) Representative FACS plot for bone marrow engraftment of human CD45+ hematopoietic cells. Percentage fractions for hCD45 gate normalized to live cells represented as mean ± SD. N = 1 donor, 2 mice for stem fraction, 4 mice for progenitor fraction. (**C**) Quantification of the percentage of human CD45+ cells in the bone marrow of all transplanted mice. Two tailed unpaired t-test **$p<0.01$, *$p<0.05$. (**D**) Representative FACS plots showing human CD45 chimerism in the peripheral blood. Percentage fractions for hCD45 gate normalized to live cells represented as mean ± SD. CD33+ myeloid and CD19+ lymphoid gates show the lineage potential for the human CD45+ cells. Percentage fractions for hCD33/hCD19 gate normalized to hCD45+ cells represented as mean ± SD. N = 1 donor, 2 mice for stem fraction, 4 mice for progenitor fraction. (**E**) Quantification of the peripheral blood chimerism based on the fraction of human CD45+ cells within the live cells. Two tailed unpaired t-test **$p<0.01$, *$p<0.05$. **(F)** lineage potential with respect to the total engrafted human CD45+ cells in the mice transplanted with the Lin-CD34+CD38- stem fraction. **(G-H)** Representative FACS plots showing human chimerism for Luciferase or *BAZ2B* transduced cells indicating the CD33-myeloid and CD19-lymphoid potential. Human CD45 population normalized to live cells represented as mean ± SD. CD33+ myeloid and CD19+ lymphoid gates show the lineage potential for the human CD45+ cells. Human CD33/CD19 populations normalized to human CD45+ cells is represented as mean ± SD. N = 4 donors, 2-3 mice per donor. **(G)** Spleen and **(H)** Peripheral blood.

Figure S6

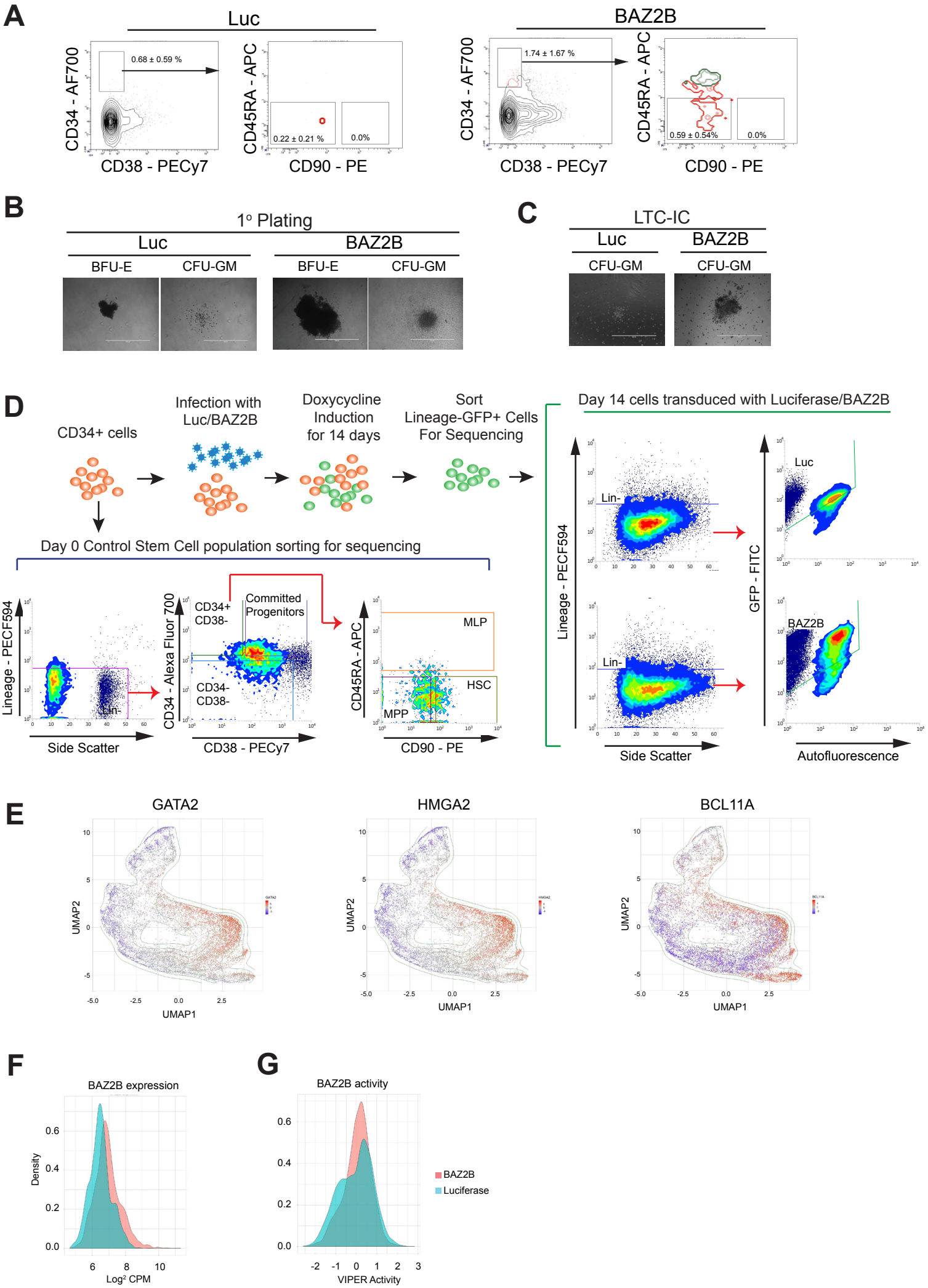

A
Luc
BAZ2B
CD34 - AF700
CD38 - PECy7
0.68 ± 0.59 %
1.74 ± 1.67 %
CD45RA - APC
CD90 - PE
0.22 ± 0.21 %
0.0%
0.59 ± 0.54%
B
1° Plating
Luc
BAZ2B
BFU-E
CFU-GM
C
LTC-IC
D
CD34+ cells
Infection with Luc/BAZ2B
Doxycycline Induction for 14 days
Sort Lineage-GFP+ Cells For Sequencing
Day 14 cells transduced with Luciferase/BAZ2B
Day 0 Control Stem Cell population sorting for sequencing
Lineage - PECF594
Lin-
Side Scatter
CD34 - Alexa Fluor 700
CD34+ CD38-
Committed Progenitors
CD34- CD38-
MLP
HSC
MPP
GFP - FITC
Autofluorescence
E
GATA2
HMGA2
BCL11A
UMAP1
UMAP2
F
BAZ2B expression
Density
Log² CPM
G
BAZ2B activity
VIPER Activity
BAZ2B
Luciferase

Figure S6. *BAZ2B*-induced reprogramming of committed progenitors to multipotent hematopoietic progenitors. Related to figure 6.

(A) Representative FACS plot on day 14 showing the progenitor Lin-GFP+CD34+CD38-, MPP Lin-GFP+CD34+CD38-CD45RA-CD90- and HSC Lin-GFP+CD34+CD38-CD45RA-CD90+. Percentage fractions for each gate normalized to Lineage-GFP+ fraction represented as mean ± SD from N = 4 donors. **(B)** Representative colony images of BFU-E and CFU-GM from the primary Methocult plating. Scale bar 0-1000um. **(C)** Representative colony images of BFU-E, CFU-GEMM and CFU-GM from the LTC-IC methocult plating. Scale bar 0-1000um. **(D)** Top-left panel - Experimental schematic of the single-cell sequencing experiment. Bottom-Left Panel - FACS plots showing control cell populations sorted on day 0 for sequencing: HSC (Lin-CD34+CD38-CD45RA-CD90+), MPP (Lin-CD34+CD38-CD45RA-CD90-), MLP (Lin-CD34+CD38-CD45RA+) enriched population, and Lineage committed Progenitors (Lin-CD34+CD38+). Right panel - FACS plots showing the CD34+ cells transduced with Luciferase and *BAZ2B* for 14 days. Lineage-GFP+ cells were sorted for single-cell sequencing. **(E)** UMAP plot showing VIPER activity of transcription factor markers in reference single-cell populations, including HSCs, MPPs, MLPs, and Lineage committed progenitors. **(F,G)** Lineage-committed progenitors transduced with Luciferase/*BAZ2B*. **(F)** mRNA expression of *BAZ2B* **(G)** VIPER-predicted activity of *BAZ2B*.

Figure S7

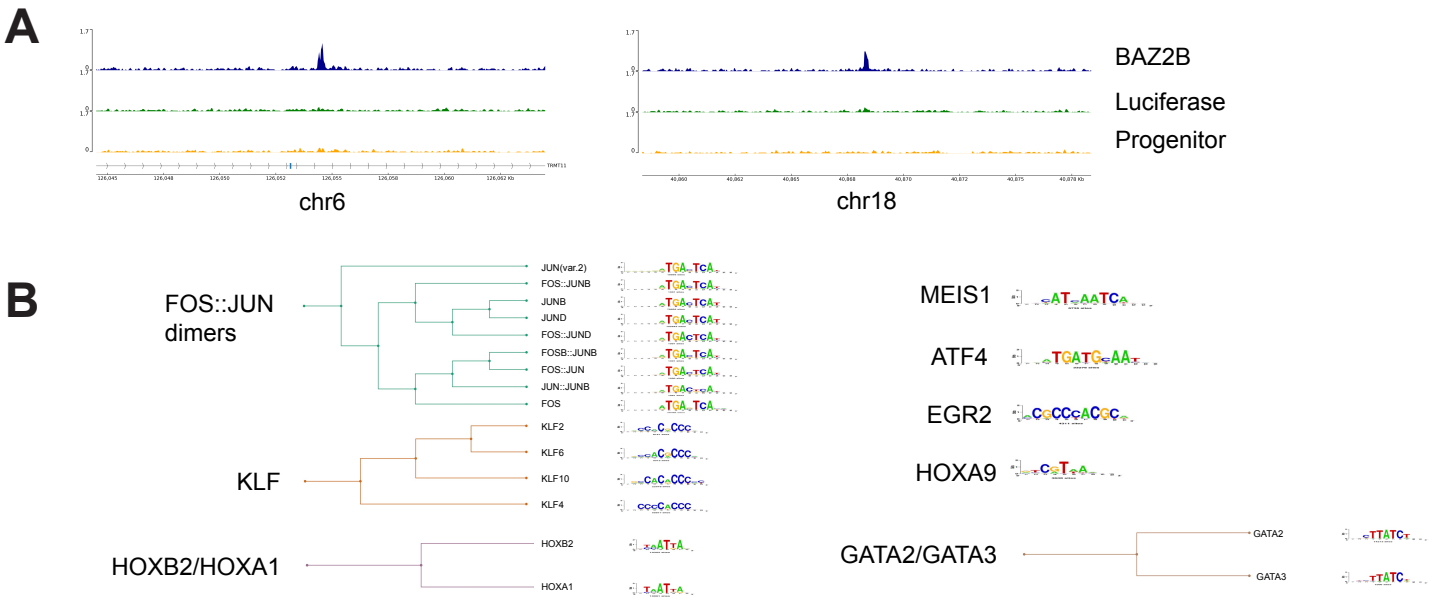

A
chr6
chr18
BAZ2B
Luciferase
Progenitor
B
FOS::JUN dimers
JUN(var.2)
FOS::JUNB
JUNB
JUND
FOS::JUND
FOSB::JUNB
FOS::JUN
JUN::JUNB
FOS
KLF
KLF2
KLF6
KLF10
KLF4
HOXB2/HOXA1
HOXB2
HOXA1
MEIS1
ATF4
EGR2
HOXA9
GATA2/GATA3
GATA2
GATA3
C
Committed Progenitors
Luciferase
BAZ2B
ZNF148
TFE3
TAL1
SP1
POU2F1
PBX2
NRF1
NFIL3
NFE2L1
MAFK
KLF9
KLF5
KLF3
KLF16
KLF11
HOXB5
HOXB4
HOXB3
HOXA7
HOXA6
HOXA4
HLF
FOSL2
FOSL1
EGR1
DBP
CUX1
CEBPG
CEBPA
BATF
BACH2
-5.0
-2.5
0
2.5
5.0
Activity
D
Spleen
Luciferase
0.0%
mCD45 - PE-Cy7
hCD33 - PE
BAZ2B
0.2 ± 0.4%
0.9 ± 1.6%
3.2 ± 7.1%
hCD45 - AF700
hCD19 - APC
E
Peripheral Blood
0.1 ± 0.1%
1.7 ± 3.3%
3.3 ± 5.8%

Figure S7. *BAZ2B*-induced chromatin accessibility and lineage potential of long-term engrafted cells during reprogramming of committed progenitors to multipotent hematopoietic progenitors. Related to Figure 6 and 7.

(A) Representative ATAC-Seq peaks unique to *BAZ2B*-transduced progenitors compared to Luciferase-transduced or untransduced committed progenitors. **(B)** Motif clustering by RSAT for consensus motifs of 17 MRs sub-grouped into 9 TF families with enriched motif-binding sites in *BAZ2B*-induced nucleosome-free regions and VIPER protein activity > 1.0. **(C)** Ridgeline density plots showing the VIPER protein activity of 31 MRs with < 1.0 activity score. **(D-E)** Lin-CD34+CD38- stem fraction and the Lin-CD34+CD38+ progenitor fraction were freshly sorted and transplanted intra-femorally in NSG mice. Human hematopoietic engraftment was analysed after 12 weeks. Representative FACS plots showing engrafted human CD45+ cells in NSG mice indicating the CD33-myeloid and CD19-lymphoid potential. Human CD45 population normalized to live cells represented as mean ± SD. Human CD33/CD19 populations normalized to hCD45+ population represented as mean ± SD. N = 3 donors with 2-3 mice per donor. **(D)** Spleen and **(E)** Peripheral blood.

Table S1. For each sample, the table shows the total number of mapped reads, the number of reads that mapped uniquely to the human and mouse genomes, multi-mapping reads that mapped to both genomes, reads that mapped to either the human (hg19) or mouse (mm10) genomes after multi-mapping reads were assigned to either the mouse or human genome using the "fair-split" method, number of human/mouse genes based on the mapped reads after "fair-split" correction. *Related to Figure 1*.

	Total Reads	**Human (Unique)**	**Mouse (Unique)**	**Multi-mapping**	**Human (Post fair-split)**	**Mouse (Post fair-split)**	**Human Genes > 0 reads**	**Human Genes > 1 read**	**Human Genes > 5 reads**	**Mouse Genes > 0 reads**	**Mouse Genes > 1 read**	**Mouse Genes > 5 reads**
Bcell	28,273,008	27,680,415	1,765	590,828	28,271,243	1,765	17,029	16,082	14,418	134	71	34
Bcell	23,071,287	22,450,851	1,402	619,034	23,069,885	1,402	17,025	15,975	14,338	139	70	35
Bcell	25,572,699	25,020,566	1,461	550,672	25,571,238	1,461	17,057	16,048	14,443	126	68	30
Het4h	26,385,518	8,357,365	17,498,299	529,854	8,670,944	17,714,574	15,071	14,054	12,474	17,730	16,998	15,608
Het4h	33,172,158	9,984,775	22,655,854	531,529	10,296,287	22,875,871	15,281	14,362	12,827	17,882	17,236	15,895
Het4h	83,758,127	41,070,282	41,062,055	1,625,790	41,925,026	41,833,101	17,188	16,435	15,020	17,612	16,937	15,625
Het12h	71,139,709	32,242,448	37,825,136	1,072,125	32,842,567	38,297,142	17,214	16,561	15,084	17,708	17,154	15,887
Het12h	75,805,230	37,458,398	36,854,827	1,492,005	38,254,324	37,550,906	16,215	15,473	14,064	17,282	16,667	15,404
Het12h	78,824,375	39,022,278	38,117,926	1,684,171	39,931,191	38,893,184	16,456	15,653	14,239	17,388	16,776	15,448
Het48h	51,280,183	3,271,949	47,294,808	713,426	3,674,748	47,605,435	14,138	13,054	11,518	17,600	16,866	15,618
Het48h	51,742,113	3,034,489	47,850,548	857,076	3,495,871	48,246,242	14,192	13,072	11,451	17,763	17,132	15,993
Het48h	48,586,528	8,151,493	39,044,433	1,390,602	8,858,769	39,727,759	15,884	14,814	12,977	18,167	17,442	16,196
Het120h	63,139,366	1,115,469	60,562,409	1,461,488	1,897,669	61,241,697	13,410	12,028	9,817	18,247	17,605	16,492
Het120h	53,906,586	798,759	51,915,397	1,192,430	1,506,793	52,399,793	11,921	10,741	8,317	18,184	17,711	16,494
Het120h	67,176,421	754,427	65,038,187	1,383,807	1,488,917	65,687,504	11,211	9,536	6,359	18,296	17,682	16,567

Table S2. 105 TFs that were significantly associated (P-val < 0.05) with Principal Component 1 of SVD analysis. *Related to Figure 2*.

Name	Het4h	Het12h	Het48h	Het120h
AATF	4.543	1.518	-2.372	-1.320
ARHGAP35	2.172	2.575	1.013	1.570
ASH2L	1.805	3.364	2.638	2.154
ATAD2	4.171	6.900	2.323	2.134
BUD31	4.872	5.853	1.518	-1.323
CARHSP1	3.324	4.407	6.150	2.710
CBX2	4.972	4.176	1.070	0.305
CCRN4L	3.885	3.627	-1.524	-2.300
CHAMP1	2.708	3.371	0.966	0.815
CREB3	2.299	2.437	2.537	0.031
CREM	1.123	3.672	-1.452	-2.367
DMAP1	2.799	2.849	1.741	-0.550
E2F1	5.530	4.551	1.604	-0.545
E2F2	6.573	6.011	2.385	1.818
E2F6	2.763	3.417	-5.651	-4.576
E2F7	7.408	8.296	5.779	2.270
E2F8	8.996	10.467	6.720	4.615
ECSIT	5.210	3.108	1.368	-1.422
ENO1	4.656	2.521	0.072	0.354
FOXM1	10.100	8.889	5.362	2.396
FUBP1	4.084	4.318	1.374	3.333
GABPB1	2.542	4.371	-2.339	-2.235
GTF2H4	4.341	3.410	2.516	0.312
GTF2IRD1	3.562	3.600	0.011	1.606
GZF1	1.104	4.556	-0.266	-0.158
HDAC1	1.816	1.956	4.663	3.900
HDGF	5.749	5.919	0.481	0.368
HLTF	3.882	6.754	4.518	3.706
HMGA1	5.444	2.161	2.314	0.207
HMGB1	2.907	5.386	3.323	3.378
HMGB2	6.350	8.586	5.777	4.188
HMGN5	2.850	4.127	1.600	-0.007
HNRNPAB	8.661	8.957	2.849	2.259
HSF2	1.512	3.356	-1.888	-0.530
IKZF5	2.913	3.714	-2.045	0.500
ILF2	5.922	5.817	-0.782	-0.813
ILF3	5.285	2.222	-1.339	0.691
KDM1A	5.808	3.711	2.958	2.299

MAZ	6.143	1.368	0.058	-0.150
MNX1	2.082	2.890	1.703	1.037
MRPL28	8.491	7.121	3.800	0.993
MRRF	3.248	3.290	-4.592	-4.436
MTA2	4.257	3.924	-0.785	-0.291
MXD3	2.041	2.545	3.225	2.004
MYBL2	8.945	8.727	6.145	3.542
NAT14	4.136	2.389	2.491	-0.073
NFYA	2.557	1.687	0.165	1.012
NR2F6	6.098	5.415	0.910	2.198
PA2G4	5.216	3.137	-4.946	-5.014
PHF5A	3.685	4.654	0.761	-0.201
PITX1	6.277	4.879	-0.461	-0.559
PLAGL2	4.208	2.498	-2.272	-2.731
PREB	3.970	4.164	1.290	-0.158
PRKDC	6.401	6.080	1.973	4.338
PSMC3IP	5.789	7.087	0.451	0.063
PTTG1	8.772	10.245	8.582	4.624
PURB	3.587	5.602	-0.988	-2.163
RFXANK	6.241	7.156	3.805	0.354
RFXAP	2.581	3.701	1.413	1.770
SARNP	5.199	7.567	5.038	1.473
SCMH1	6.303	3.742	3.047	3.084
SMARCA4	4.092	1.527	1.458	1.221
TADA2A	2.930	2.637	-1.924	-1.838
TAF12	1.608	4.715	3.383	1.100
TAF1B	3.813	5.687	-0.455	-0.813
TAF5	4.787	6.774	0.146	-0.914
TAF9	4.098	7.709	0.480	-0.473
TAF9B	1.886	3.234	0.295	0.821
TCF19	5.102	5.239	4.705	2.881
TCF3	3.128	1.439	2.462	1.211
TFAM	4.050	6.567	-1.645	-0.656
TFAP4	3.056	3.616	-0.240	0.241
TFDP1	6.924	8.627	6.414	4.172
TFDP2	2.218	3.585	-0.497	-0.559
THAP7	3.727	1.292	-0.877	-1.835
TOP2A	8.939	10.126	6.888	4.756
UHRF1	9.938	9.326	6.130	4.017
WHSC1	8.589	7.152	4.670	2.999
YBX1	6.092	6.491	2.652	2.485

YEATS4	6.043	8.922	3.776	1.309
ZBED4	4.087	2.913	-2.311	-0.847
ZC3HC1	3.289	2.379	-2.882	-2.686
ZCCHC4	4.086	3.843	0.173	0.062
ZCRB1	0.782	4.113	1.620	-0.301
ZDBF2	1.580	2.932	1.923	1.015
ZDHHC16	4.238	4.345	-0.287	-1.737
ZFPL1	3.992	2.256	0.266	-1.976
ZMAT5	3.073	2.474	2.787	-0.095
ZMYND19	6.231	3.468	-3.816	-3.495
ZNF174	4.455	4.718	-2.413	-3.180
ZNF207	6.071	6.775	0.949	2.146
ZNF250	2.321	2.238	1.425	0.497
ZNF259	3.954	3.774	-5.315	-5.148
ZNF282	4.617	1.833	0.481	-1.582
ZNF326	2.248	3.761	-3.014	-1.342
ZNF367	4.541	6.792	2.838	1.980
ZNF511	2.662	2.844	-1.308	-2.913
ZNF552	3.091	2.930	-1.417	-1.576
ZNF576	3.832	3.012	-2.835	-2.514
ZNF593	5.758	5.692	-2.610	-3.423
ZNF653	4.467	2.150	0.926	0.096
ZNF670	3.003	3.627	-3.435	-1.389
ZNF747	2.505	3.142	-0.561	-1.823
ZNHIT1	3.722	4.847	4.393	1.094
ZNHIT3	1.735	3.778	3.575	0.996

Table S3. 64 TFs that were significantly associated (P-val < 0.05) with Principal Component 2 of SVD analysis. *Related to Figure 2.*

Name	Het4h	Het12h	Het48h	Het120h
AKNA	-2.825	-6.503	3.724	2.997
ALS2CR8	-2.326	0.183	2.050	2.418
ASH1L	-4.155	-3.424	1.498	3.693
BAZ2B	-2.148	-2.705	4.588	5.026
BCL11A	-2.983	-2.277	2.987	2.243
CBFA2T2	-1.138	-3.141	2.202	2.715
CBFA2T3	0.220	-1.581	2.593	2.103
CGGBP1	-2.611	0.289	1.996	2.655
CNOT8	-3.093	-0.733	4.056	4.675
CREB1	-4.687	-2.142	1.573	2.923
CREBZF	-2.112	-1.460	1.054	3.279
DMTF1	-2.816	-2.914	3.015	3.585
ELF1	-2.012	-0.312	1.495	3.683
FLI1	-2.634	-0.762	3.686	3.128
FOXN3	-2.498	-2.778	3.013	2.615
FOXO1	-0.955	-0.864	2.850	2.852
GATA1	-3.216	-3.715	2.861	1.394
GTF2I	0.778	0.112	1.755	3.522
HBP1	-5.360	-3.934	2.962	2.065
HCLS1	-1.455	-2.715	5.913	3.405
HDAC1	1.816	1.956	4.663	3.900
IFI16	-3.810	-3.397	1.842	1.461
IRF2	-1.918	-2.976	4.728	2.944
IRF8	-3.875	-3.686	4.545	2.318
IRF9	-5.574	-7.158	2.826	1.340
KLF12	-2.864	-1.214	4.590	3.620
LYL1	-0.725	-1.550	5.069	2.773
MEF2C	-2.820	-0.276	3.604	3.567
MSL3	-3.131	-1.231	2.765	2.200
MYBL1	0.010	1.085	4.805	3.233
NR2C2	-4.493	-3.318	0.579	2.528
POU2F2	-1.903	-4.024	2.881	1.368
RBL2	-6.015	-5.206	2.421	2.534
RERE	-0.358	-3.939	2.291	3.288
RFX5	-2.632	-2.376	3.778	2.704
SP140	-3.840	-3.315	3.101	0.956
SPI1	1.393	-1.663	3.842	1.262
SPIB	0.005	-0.088	4.447	2.099

STAT2	-1.082	-2.467	4.916	3.165
STAT5B	-1.264	-0.988	2.801	2.955
STAT6	-1.631	-5.465	2.694	1.077
TADA3	1.164	0.214	4.372	2.056
TAF10	-1.452	-2.289	5.065	2.162
TCEAL1	-2.477	-0.754	2.264	1.686
TFEB	-1.576	-2.273	2.782	1.377
ZBTB20	-4.185	-3.492	3.746	3.861
ZC3H6	-2.718	-1.188	1.829	2.028
ZCCHC11	-3.312	-3.614	2.095	2.060
ZCCHC6	-2.229	-0.822	2.599	2.390
ZDHHC17	-3.765	-3.623	0.806	3.139
ZFC3H1	-3.625	-3.812	0.490	3.763
ZFP161	-2.205	0.608	3.708	3.302
ZFP90	-4.046	-4.074	2.092	1.380
ZFYVE21	-0.369	-0.867	2.805	1.681
ZHX2	-1.316	-1.112	4.998	3.970
ZMAT1	-4.296	-3.626	2.493	3.840
ZMYM2	-3.633	-1.092	1.146	3.018
ZNF292	-3.893	-2.265	1.987	2.974
ZNF395	0.752	-0.306	3.015	2.023
ZNF512	-0.888	-2.147	1.529	3.258
ZNF581	1.701	0.588	3.640	1.530
ZNF611	-2.437	-2.978	1.940	2.693
ZNF652	-2.529	-0.104	1.973	2.730
ZNF763	-2.319	-2.089	2.378	1.490
ZSWIM6	-2.708	-2.172	1.556	2.817

Table S4. MRs ranked by statistical significance. Genes in red font with white background used in the 8-factor cocktail. Genes in red font with yellow background used in the 5-factor cocktail. *Related to Figure 4*.

Rank	Gene	Het120h	Het120h	Gene	HSC	HSC	Gene	MPP	MPP	Gene	MLP	MLP
		NES	FDR		NES	FDR		NES	FDR		NES	FDR
1	BAZ2B	5.02562	1.50E-05	BAZ2B	4.24258	0.000461	CNOT8	5.6119	9.19E-07	BAZ2B	6.0133	1.05E-07
2	CNOT8	4.67461	7.03E-05	CNOT8	4.19081	0.000559	ZDHHC17	5.00837	1.81E-05	ZBTB20	5.90527	1.88E-07
3	ZHX2	3.9697	0.001049506	HBP1	4.11598	0.000747	KLF12	4.19347	0.000554	ASH1L	5.23584	6.22E-06
4	ZBTB20	3.86105	0.001517053	ZBTB20	3.85393	0.001884	CREB1	4.09086	0.000818	CNOT8	5.16886	8.65E-06
5	ZMAT1	3.83977	0.001632292	KLF12	3.58386	0.004582	DMTF1	4.0104	0.001095	CREB1	4.80127	4.71E-05
6	ZFC3H1	3.76325	0.002099259	DMTF1	3.18559	0.014848	HBP1	3.87893	0.001726	DMTF1	4.39867	0.000251
7	ASH1L	3.69267	0.002618486	ZDHHC17	3.02631	0.022934	BAZ2B	3.71004	0.003054	KLF12	4.32559	0.000335
8	ELF1	3.68283	0.002689561	ASH1L	3.01203	0.023839	ZBTB20	3.55604	0.00496	HBP1	4.21674	0.000509
9	KLF12	3.62014	0.00330098	ELF1	2.73989	0.04708	ASH1L	3.54716	0.005103	ZMAT1	4.13246	0.000704
10	DMTF1	3.58468	0.003683325	CREB1	2.62948	0.060037	CREBZF	3.54655	0.005108	HCLS1	3.95528	0.001335
11	MEF2C	3.56698	0.003898188	ZMAT1	2.55368	0.071013	ZFC3H1	3.5215	0.005535	CREBZF	3.95449	0.001337
12	HCLS1	3.40531	0.006547864	ZNF512	2.48523	0.083021	ELF1	3.51595	0.005609	ZHX2	3.73065	0.002853
13	RERE	3.28813	0.009285318	HCLS1	2.48457	0.083121	ZNF512	2.96383	0.027129	ELF1	3.69605	0.003213
14	CREBZF	3.27927	0.009564171	ZHX2	2.3553	0.109479	ZMAT1	2.82563	0.03837	ZDHHC17	3.65817	0.00359
15	ZNF512	3.2576	0.010213069	IRF2	2.35267	0.109808	ZNF292	2.62743	0.060315	IRF2	3.24122	0.012628
16	STAT2	3.16453	0.013592443	CREBZF	2.25786	0.132438	IRF2	2.61643	0.061855	FOXN3	3.16584	0.015669
17	ZDHHC17	3.1391	0.01449553	ZFC3H1	2.07549	0.183242	ZMYM2	2.47409	0.08516	ZMYM2	3.14054	0.016729
18	FLI1	3.12843	0.014805925	ZNF292	2.05693	0.190129	ZHX2	2.37114	0.105274	FLI1	3.13468	0.016998
19	ZMYM2	3.01767	0.020021959	STAT2	2.04817	0.192896	FLI1	2.22344	0.141716	MSL3	2.98771	0.02544
20	ZNF292	2.97411	0.022281624	MSL3	1.93049	0.23424	MEF2C	2.13014	0.167618	IRF8	2.9831	0.025757
21	IRF2	2.94383	0.024163623	ZMYM2	1.92281	0.236724	IRF8	2.07561	0.183242	STAT2	2.98122	0.025876
22	CREB1	2.92263	0.025535747	RERE	1.87354	0.255841	RERE	1.96734	0.220294	RERE	2.97198	0.026538
23	FOXN3	2.61504	0.053207671	MEF2C	1.78313	0.293293	MSL3	1.88486	0.250913	ZFC3H1	2.94799	0.028254
24	IRF8	2.31811	0.102091602	IRF8	1.77609	0.296536	HCLS1	1.46311	0.449435	ZNF512	2.90785	0.031122
25	MSL3	2.20003	0.12811302	FLI1	1.41279	0.476419	STAT2	1.16474	0.607859	MEF2C	2.87547	0.033744
26	HBP1	2.06533	0.166044095	FOXN3	1.2932	0.537395	FOXN3	1.10859	0.634883	ZNF292	2.83477	0.037473

Table S5. Genes that were identified as classifiers by the random forest model to distinguish HSCs, MPPs, MLPs and the committed progenitors. *Related to Figure 6*.

	Gene		**Gene**
1	CD74	24	CD53
2	HSPA5	25	PNRC1
3	DNTTIP2	26	HHEX
4	IRAK3	27	RHOC
5	PIK3IP1	28	RBPMS
6	XBP1	29	PDZD8
7	RIPK2	30	STX3
8	EEF1D	31	LAIR1
9	BIRC2	32	UBB
10	SELENOK	33	FOS
11	TMEM59	34	MLLT3
12	DUSP1	35	DDIT3
13	SOCS2	36	TAF7
14	FOSB	37	HLA-DQB1
15	ZFP36	38	GATA2
16	BST2	39	MAFF
17	JUND	40	HLA-DRB1
18	RNF138	41	EVL
19	LMO2	42	SVIP
20	ITM2B	43	LRBA
21	GYPC	44	GPSM3
22	APP	45	VAMP2
23	PRKACB	46	HLA-DPB1

Table S6. RSAT clustering of enrriched motifs with the average p-values of motif enrichment within each cluster. p-values for individual motif enrichment were first calculated using AME. RSAT was then used to cluster the motifs based on consensus sequences, and the average of individual p-values for motifs within each group were calculated. *Related to Figure 6*.

	Motif Clusters	No of Motifs	p val	adj pval
1	JUN(var.2),BACH2,FOSL1,JUNB,JUND,FOSL1::JUNB,FOSL1::JUN,FOS::JUND,FOSL2::JUN,FOS::JUNB,JDP2,NFE2,FOS,BATF3,BATF,BATF::JUN,FOSL2,JUN::JUNB,FOSL1::JUND,FOS::JUN,FOSL2::JUND,FOSB::JUNB,FOSL2::JUNB,NFE2L2,MAF::NFE2,BACH1,MAFK,NFE2L1	28	1.21E-08	5.23E-05
2	LHX2,LHX6,HOXA6,MEOX2,MEOX1,EVX1,EVX2,GSX1,HOXB3,MNX1,EN2,LBX2,ESX1,MIXL1,ALX3,NKX6-3,HOXC8,HOXA1,HOXB5,NKX6-1, NKX6-2,HOXA4,HOXB4,HOXC4,HOXD4,VAX1,VAX2,GSX2,HOXB2,PDX1,LHX1,EN1,DRGX,SHOX,UNCX,RAX2,DLX6,MSX2,MSX1,BARX1,BSX,HOXB6,HOXA7,HOXB8,HOXD8	45	2.20E-07	0.00148087
3	NFYB,NFYA,NFYC	3	5.10E-07	0.00358
4	FOS::JUN(var.2),FOSL1::JUND(var.2),FOSB::JUNB(var.2),FOSL1::JUN(var.2),JUN::JUNB(var.2),FOSL2::JUN(var.2),FOSB::JUN	7	4.71E-07	0.00227121
5	POU5F1B,POU3F4,POU2F1,POU1F1,POU3F3,POU3F2,POU3F1	7	1.76E-07	0.000655467
6	ONECUT1,CUX1,CUX2	3	2.87E-06	0.00530333
7	KLF5,KLF15,KLF3,KLF2,KLF6,KLF14,KLF10,KLF16,KLF11,SP3,SP8,SP9	12	1.37E-16	1.57E-12
8	MEIS2(var.2),PBX2,MEIS1(var.2)	3	3.10E-07	0.00192567
9	ATF4,CEBPG(var.2)	2	3.01E-11	1.29E-07
10	CEBPA,HLF,NFIL3	3	1.57E-07	0.000993667
11	KLF4,MAZ,ZNF148	3	2.51E-07	0.0022833
12	GATA1,GATA6,GATA2,GATA4,GATA3,GATA5	6	1.72E-21	1.87E-17
13	EGR2,EGR3,EGR1	3	3.78E-07	0.00108122
14	SP2,SP4,SP1	3	5.37E-14	2.23E-10
15	POU4F3,POU4F1	2	1.34E-07	0.000576595
16	DMRTA2,DMRT3	2	7.19E-07	0.00641905
17	GSC2,DPRX	2	4.35E-08	0.0004275
18	PHOX2B,PHOX2A,DUXA	3	4.96E-13	2.12E-09
19	KLF9	1	1.91E-07	0.00233
20	CEBPE,DBP	2	4.96E-06	0.010275
21	POU6F1(var.2)	1	3.40E-09	2.64E-05
22	FOXB1	1	4.13E-09	2.99E-05
23	ZBTB14	1	1.42E-12	1.20E-08
24	CTCF,CTCFL	2	1.89E-11	2.58E-07
25	BARX2	1	3.50E-10	2.93E-06
26	NRF1	1	3.24E-21	1.54E-17
27	NKX3-1	1	2.57E-08	0.000123
28	MITF,TFE3	2	1.12E-06	0.005905
29	HOXA9	1	2.42E-06	0.0143
30	GATA1::TAL1	1	1.73E-10	5.37E-07
	Total No. of Motifs	152		

Table S7. VIPER predicted activites for TFs with enriched binding motifs in BAZ2B induced nucleosome-free regions. TF activities with low statistical significance are marked in red. *Related to Figure 6*.

	Gene	Average VIPER activity	p.val
1	MEIS1	4.165717478	<2.23e-308
2	HOXB2	2.824254346	<2.23e-308
3	KLF10	2.611233037	<2.23e-308
4	GATA2	2.231472878	<2.23e-308
5	KLF6	1.947992437	<2.23e-308
6	JUN	1.838219735	<2.23e-308
7	GATA3	1.802026924	<2.23e-308
8	ATF4	1.656327626	<2.23e-308
9	FOSB	1.46695911	<2.23e-308
10	HOXA9	1.300555709	<2.23e-308
11	FOS	1.235620929	<2.23e-308
12	JUNB	1.085880464	<2.23e-308
13	JUND	1.058128589	<2.23e-308
14	EGR2	3.062017975	4.82E-248
15	HOXA1	3.075411193	4.63E-247
16	KLF2	1.28735717	1.16E-232
17	KLF4	1.100806785	1.25E-218
18	HLF	0.744134719	5.03E-215
19	CUX1	0.466408185	3.70E-202
20	EGR1	0.905845912	1.26E-185
21	KLF5	0.74390973	2.29E-143
22	BACH2	0.303543334	2.81E-123
23	TAL1	0.284711423	1.02E-120
24	BATF	0.2910327	7.43E-114
25	NRF1	0.413293005	4.74E-106
26	CEBPA	-0.089865163	3.77E-100
27	TFE3	0.590631286	5.54E-100
28	NFIL3	0.257470563	5.27E-87
29	FOSL1	0.824222327	9.80E-86
30	HOXB3	0.255682528	1.07E-84
31	KLF3	0.082421261	2.64E-52
32	HOXB5	0.15557354	6.16E-52
33	KLF9	0.160851708	3.96E-50
34	HOXA6	0.33992665	2.93E-47
35	KLF11	0.230972851	7.16E-43
36	HOXB4	0.465682352	4.97E-37
37	HOXA7	0.063953445	1.80E-36
38	POU2F1	0.217705195	3.39E-31
39	HOXA4	0.097971819	1.77E-28
40	MAFK	0.154196678	1.10E-25
41	ZNF148	0.026730097	4.06E-18
42	KLF16	0.00495424	1.05E-17
43	CEBPG	0.022724506	2.46E-16
44	SP1	-0.108364718	5.20E-15
45	FOSL2	0.029001083	5.43E-08
46	DBP	-0.116160488	1.20E-06
47	PBX2	-0.166428132	9.43E-06
48	NFE2L1	-0.101142318	0.004423076
49	NFYB	-0.13381193	0.800102118
50	NFE2	-0.244520637	0.994630511
51	NFE2L2	-0.02996747	0.997284523
52	SP3	-0.103008973	0.999631991
53	CTCF	-0.133484911	0.999958385
54	NFYC	-0.018464373	0.999999953
55	SP2	-0.173251666	1
56	MAZ	-0.541410864	1
57	GATA1	-1.923387595	1