EBF1 and PAX5 control pro-B cell expansion via opposing regulation of the *Myc* gene.

Somasundaram et al.

Supplementary Information/Materials and Methods

Figure S1

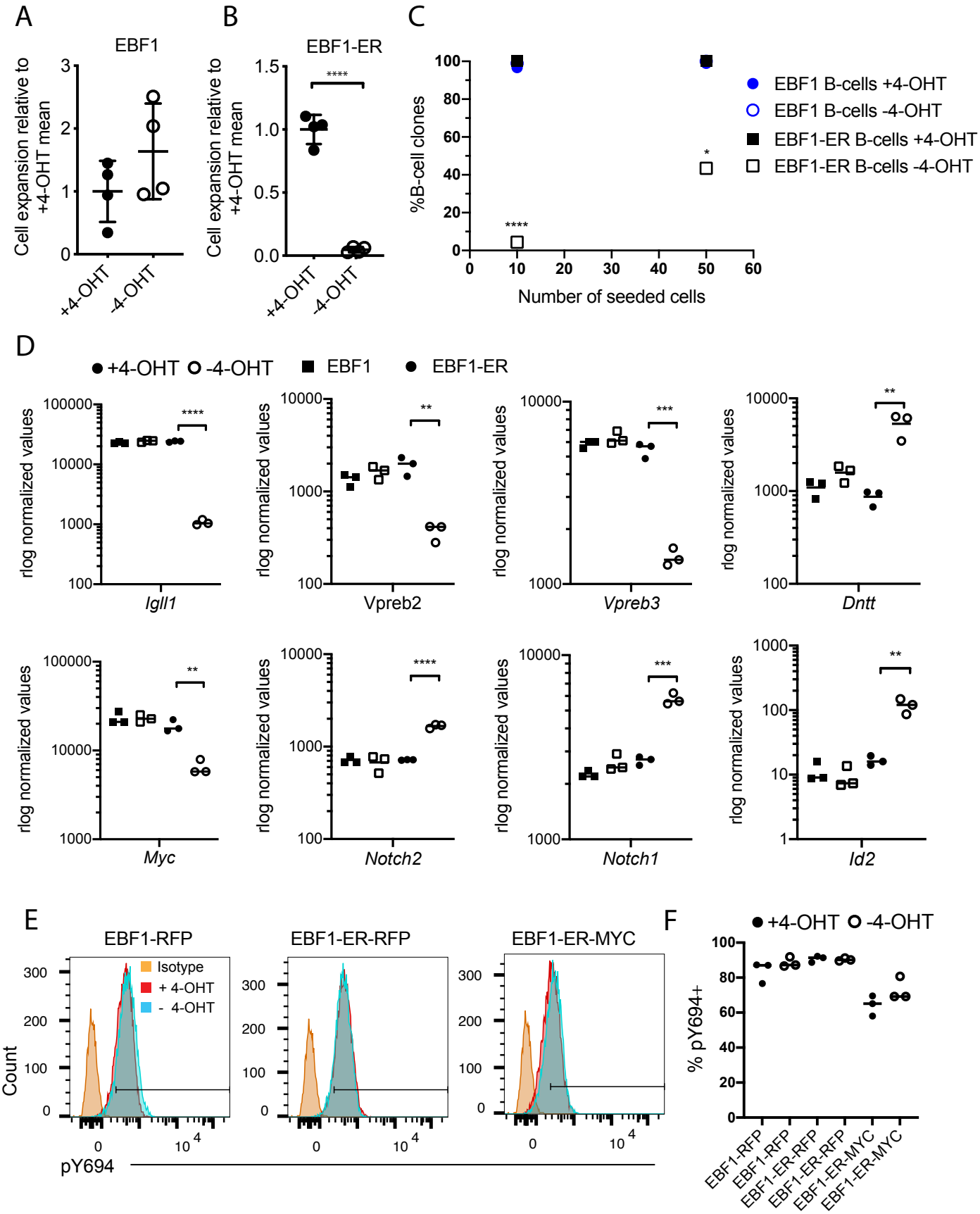
A
EBF1
Cell expansion relative to +4-OHT mean
+4-OHT
-4OHT
B
EBF1-ER
****
C
%B-cell clones
Number of seeded cells
EBF1 B-cells +4-OHT
EBF1 B-cells -4-OHT
EBF1-ER B-cells +4-OHT
EBF1-ER B-cells -4-OHT
D
+4-OHT -4-OHT EBF1 EBF1-ER
rlog normalized values
Igll1
Vpreb2
Vpreb3
Dntt
Myc
Notch2
Notch1
Id2
E
EBF1-RFP
EBF1-ER-RFP
EBF1-ER-MYC
Isotype
+ 4-OHT
- 4-OHT
Count
pY694
F
% pY694+
EBF1-RFP
EBF1-RFP
EBF1-ER-RFP
EBF1-ER-RFP
EBF1-ER-MYC
EBF1-ER-MYC

Figure S1: Loss of nuclear EBF1 in pro-B cells impairs cell growth and expression of B-lineage genes.

(A-B) Expansion of EBF1 or EBF1-ER transduced and 4-OHT treated Ebf1$^{-/-}$ FL cells. Cells were either cultured continuously in the presence (+4-OHT) or in the absence of 4-OHT (-4-OHT) for 72 hours. (A) Expansion of each EBF1 or EBF1-ER (B) sample was related to the average expansion of the +4-OHT groups. Data is presented as mean and SD, n=4, from four experiments. Statistical analysis is based on Student's unpaired t-test. **** $p<0.0001$. (C) Cloning frequency of live EBF1 or EBF1-ER transduced cells after 72h hour of 4-OHT withdrawal (-4-OHT) or cultured continuously with 4-OHT (+4-OHT). 10 or 50 live cells were re-seeded on OP9 stroma cells in the presence of 4-OHT and the fraction of wells with detectable B-cell growth was determined 10 days after seeding. The data is based on 92 and 46 seeded wells with EBF1 and EBF1-ER expressing cells, respectively. The statistical analysis is performed with Fisher's Exact test calculated between 4-OHT treated and untreated samples. * $p<0.05$ and **** $p<0.0001$. (D) RPKM normalized RNA-seq data from EBF1 or EBF1-ER transduced and 4-OHT treated Ebf1$^{-/-}$ BM cells that were either continued on 4-OHT (+4-OHT) or 4-OHT withdrawn (-4-OHT) for 72 hours (GSE136238). Statistical analysis shown is based on Student's unpaired t-test. * $p<0.05$, ** $p<0.01$, *** $p<0.001$ and **** $p<0.0001$. (E-F) Phosphorylated STAT5a (pY694) in CD19^{+}CD45^{+}GFP^{+}RFP^{+} Ebf1$^{-/-}$ BM cells transduced with an EBF1-GFP vector (EBF1) or an ER-fused EBF1 protein (EBF1-ER), serially transduced with a RFP control or a MYC expressing RFP retrovirus cells 48 hours after 4-OHT withdrawal (-4-OHT) or in cells cultured continuously on 4-OHT (+4-OHT). (E) FACS histograms showing representative staining profiles of anti-pY694 and (F) percentage of pY694^{+} cells in triplicate samples. Data is presented as mean and SD, n=3.

Figure S2

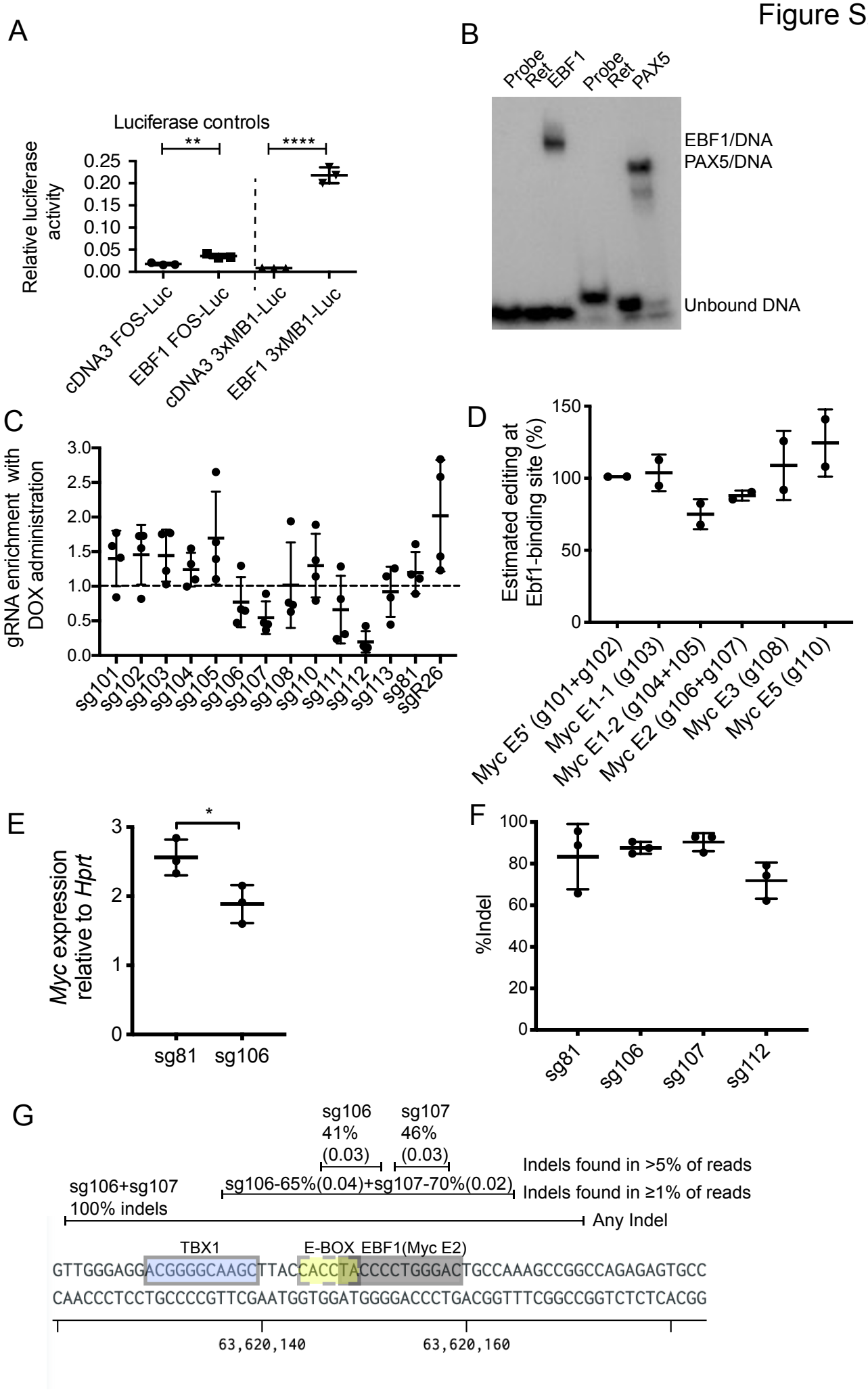
A
Luciferase controls
Relative luciferase activity
0.25
0.20
0.15
0.10
0.05
0.00
**
****
cDNA3 FOS-Luc
EBF1 FOS-Luc
cDNA3 3xMB1-Luc
EBF1 3xMB1-Luc
B
Probe
Ret
EBF1
Probe
Ret
PAX5
EBF1/DNA
PAX5/DNA
Unbound DNA
C
gRNA enrichment with DOX administration
3.0
2.5
2.0
1.5
1.0
0.5
0.0
sg101
sg102
sg103
sg104
sg105
sg106
sg107
sg108
sg110
sg111
sg112
sg113
sg81
sgR26
D
Estimated editing at Ebf1-binding site (%)
150
100
50
0
Myc E5' (g101+g102)
Myc E1-1 (g103)
Myc E1-2 (g104+105)
Myc E2 (g106+g107)
Myc E3 (g108)
Myc E5 (g110)
E
Myc expression relative to Hprt
3
2
1
0
*
sg81
sg106
F
%Indel
100
80
60
40
20
0
sg81
sg106
sg107
sg112
G
sg106
41%
(0.03)
sg107
46%
(0.03)
Indels found in >5% of reads
sg106+sg107
100% indels
sg106-65%(0.04)+sg107-70%(0.02)
Indels found in ≥1% of reads
Any Indel
TBX1
E-BOX
EBF1(Myc E2)
GTTGGGAGGACGGGGCAAGCTTACCACCTACCCCTGGGACTGCCAAAGCCGGCCAGAGAGTGCC
CAACCCTCCTGCCCCGTTCGAATGGTGGATGGGGACCCTGACGGTTTCGGCCGGTCTCTCACGG
63,620,140
63,620,160

Figure S2: The mouse *Myc* comprises regulatory elements including binding sites for EBF1.

(A) Positive and negative luciferase controls for Figure 3C. As a negative control we used a reporter carrying a minimal Fos promoter [1] and as positive control for functional EBF1 expression we used a construct positioning 3 mb-1 (Cd79a) promoter EBF1 sites in front of a TATA-box [2]. Each dot represents one transfection and the statistical analysis is based on Student's unpaired t-test. ** $p<0.01$ and **** $p<0.0001$. (B) Autoradiogram displaying EMSA analysis of *in vitro* translated EBF1 or PAX5 bound to the *Cd79a* promoter EBF1 site or the *Cd19* promoter PAX5 site. Un-programmed reticulocyte lysate is indicated as Ret. (C) The distribution of gRNAs within each sample of cultured iCas9 BM cells infected with a mix of 14 gRNA viruses as a post-DOX (Cas9 expression):pre-DOX (no Cas9 expression) ratio. Sg101-108, and sg110 represent gRNAs targeting a total of 6 different EBF1-motifs in putative *Myc* enhancer elements, sg111-113 are positive control gRNAs with the theoretical ability to generate Myc KO cells and sg81 and sgR26 are negative control gRNAs with no known ability to interfere with Myc expression. The line represents an unchanged ratio of gRNA distributions after Cas9 induction (y=1). Data is presented as mean with SD, n=4, from 2 independent experiments. (D) The graph shows the estimated percent of editing at each EBF1 binding site in the EBF1 binding site screen 6 days after Cas9 induction. To estimate the CRISPR editing at each site, DNA from iCas9 BM cells infected with viruses containing the 14 gRNA constructs were harvested and the DNA containing each binding site were amplified and sequenced. The editing efficiencies were estimated as the percentage of editing at each TF binding site compared to the distribution of the gRNA(s) targeting each site (mean with SD, n=2, 2 independent samples). (E) *Myc* Q-RT-PCR data from $CD19^+$ iCas9 BM cells transduced with CRISPR guide 81 (control) or 106 (*Myc* E2) and subsequently treated with DOX for 48 hours

(mean and SD shown, n=3, from 3 individual samples from different mice). The statistical analysis is based on Student's unpaired t-test. * $p<0.05$. (F) The plot shows the percentage of indel (insertion or deletions) found at CRISPR sites in the DNA amplicons of BM samples infected with one gRNA construct (Figure 4E) (mean with SD, n=3 from 3 independent samples/mice). (G) Regions edited by g106 and g107 obtained from the NGS data in Figure S2F. The genomic regions that encompass all indels, indels found in ≥1% and >5% of sequenced amplicons are indicated. The percentages describe the fraction of indels that fall within a region and are shown as mean (SD), n=3. The EBF1 binding site targeted by guides 106 and 107, a potential binding site for TBX1 and an E-BOX are indicated by colored boxes.

Figure S3

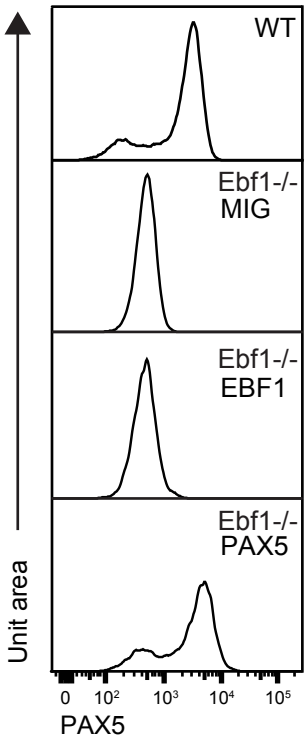
WT
Ebf1-/-
MIG
Ebf1-/-
EBF1
Ebf1-/-
PAX5
Unit area
0
10²
10³
10⁴
10⁵
PAX5

Figure S3: Ectopic expression of PAX5 results in expression levels comparable to Wt pro-B cells.

Representative histograms showing the PAX5 protein levels in *Wt* or *Ebf1*$^{-/-}$ FL cells transduced with either MIG-control, EBF1 or PAX5 encoding virus 24 hours after transduction.

Supplementary data sheet.

Supplementary data sheet 1-2: RNA-seq analysis displaying differential gene expression 72 hours after 4-OHT removal from *Ebf1*$^{-/-}$ pro-B cells rescued to the CD19^{+} pre-B cell stage by ectopic expression of either a conventional EBF1 or an 4-OHT responsive EBF1-ER fusion protein (GSE136238).The presented data are RPKM normalized as described in Materials and methods.

Supplemetary data sheet 3-4: Sequences of oligonucleotides used for EMSA or production of Guide RNAs or amplification of edited elements.

Supplementary data sheet 5-6: Crispr-Dav analysis of NGS data describing indels generated using gRNAs 106 and 107.

Supplementary Materials and methods

Cells and Cell culture.

Bone marrow B-cell precursors were generated by infection of primary *Ebf1*$^{-/-}$ BM progenitor cells with retroviruses (MIGR1 [3]) either GFP and a normal EBF1 protein or an EBF1 protein fused to a 4-OHT responsive ER ligand binding domain (EBF1-ER) [4] as described in [5]. Fetal liver (FL) pro-B progenitor cells were generated by sorting Lineage$^-$Sca1$^+$KIT$^+$ cells from *Ebf1*$^{-/-}$ FLs followed by *in vitro* differentiation on OP9 stroma cells and subsequent transduced with EBF1 expression vectors (above). Cells were maintained in B-cell media (B-cell media; Opti-MEM (ThermoFisher Scientific, Waltham, MA) supplemented with 10% heat-inactivated fetal calf serum (FCS) (HyClone, GE Healthcare, Chicago, IL), 25mM HEPES (ThermoFisher Scientific), 50μg/ml Gentamicin (ThermoFisher Scientific), 50μM β-mercaptoethanol (β-ME) (Sigma-Aldrich, St. Louis, MO), 10ng/ml KIT ligand, 10ng/ml FLT3 ligand, and 10ng/ml Interleukin-(IL-)7). All mouse cytokines were purchased from PeproTech (Rocky Hill, NJ). Antibodies used for cell sorting were from BioLegend unless otherwise indicated. Lineage markers were PECY5 conjugated and as follows: GR-1 (RB6-8C5), MAC-1 (M1/70), TER119 (TER-119), CD3 (145-2C11, B220 (RA3-6B2), CD19 (1D3), NK1.1 (PK136), CD11C (N418). In addition to lineage markers, cells were stained with SCA-1-FITC (E13-161.7) and KIT-APC (2B8, eBioscience).

For EBF1 loss of function experiments, cells were grown in B-cell media supplemented with 1uM 4-hydroxy-tamoxifen (4-OHT). The cells were then washed and reseeded in medium as above except that we used charcoaled treated serum (Fisher Scientific) was included either in the presence or absence of 4-OHT as indicated. Cell recovery was determined by FACS by

recording the number of $CD45^+CD19^+$ events in each well or when cells were grown in stroma free conditions by live cell counting using a Biorad TC10.

Transduction of Ebf1$^{-/-}$ FL cells.

Retroviral supernatants were produced in Platinum-E cells (seeded in 10 cm Petri dishes 1 day prior to transfection) with XtremeGene 9 (Roche). Supernatants were collected 48h post transfection and 500 ul was used to transduce EBF1-ER and EBF1-expressing BM-pro-B cells by spin infection (MIG-control (MIGR1), EBF1 (MIG-R1 with a full length EBF1 encoding cDNA insert) or PAX5 [MIG-R1 virus encoding a full-length mouse PAX5]) encoding virus in plates coated with retronectin (25 μg/ml) at 4°C overnight. Prior to transfection, the retronectin coated plates were blocked with 2%BSA in PBS for 30 min at RT, washed once in PBS and incubated with viral supernatant at 37°C for 5h. Transduction was performed by centrifugation of the *Ebf1$^{-/-}$* cells with fresh viral supernatant at 1800 × g for 90 min at RT, followed by overnight incubation at 37°C.

Intracellular analysis of cell cycle status, phosphorylated Stat5 and PAX5 or MYC expression levels by FACS

To investigate the cell cycle status, flow cytometry was performed on fixed cells using Ki-67 and DAPI. Prior to fixation, the cells were stained with anti-CD19 antibody (ID3, eBiosciences) in PBS/5%FCS for 15 min at 4°C. Cells were fixed and permeabilized by incubation in 250 μl BD Cytofix/Cytoperm (BD Biosciences, cat no. 554722) for 30 min on ice, followed by 2× wash in BD Perm/Wash buffer. Intracellular Ki-67 staining was performed by incubation in Perm/Wash buffer with anti-Ki-67 antibody (B56, BD Biosciences) for 1h 45min at 4°C. Prior to analysis, the cells were washed in PBS/5%FCS, stained with 200 μl 0.5 μg/ml DAPI (Invitrogen, lot 633921) at RT for 20 min, followed by another washing step in PBS/5%FCS.

Analysis of cell cycle status based on the levels of Ki-67 and DAPI was performed on GFP^+ cells using FACSAriaIIu (BD Biosciences). Events with a sub-G0/G1 DNA content were gated out to exclude apoptotic cells from the cell cycle analysis. For investigation of phosphorylated Stat5, cells were stained as above with the exceptions that anti-CD45 (30-F11, BioLegend) was added to the surface staining with anti-CD19 and anti-pY694 (47/Stat5, BD Biosciences) or isotype control (MOPC-21), BD Biosciences) was added to the intracellular staining. The analysis was performed on a BD LSRFortessa X20 flow cytometer (BD Biosciences).

For analysis of MYC and PAX5 protein levels by intracellular staining, anti-MYC (9E10, R&D systems) or anti-PAX5 (1H9, BioLegend) antibodies were added to the cells during Ki-67 staining. For intracellular staining of EBF1-ER cells, charcoal-stripped FCS was used in all washing steps.

Intracellular staining for PAX5 alone was performed using the Transcription factor buffer set (BD Biosciences Cat.no562574) according to manufacturer's instructions.

FlowJo software was used for flow cytometry data analysis, and all gatings are based on FMO controls.

AnnexinV staining

To investigate cell death, cells were washed once in PBS and once in binding buffer (556454, BD Biosciences) and thereafter stained for AnnexinV-APC for 15 minutes at RT. Binding buffer was added and samples were incubated for 5 minutes at RT followed by a spin at 400 × g for 5 minutes. Cells were resuspended in DAPI containing binding buffer for 15 minutes and analyzed by flow cytometry within the hour. AnnexinV and DAPI staining was investigated on FSC-A vs. FSC-H singlet cell events where small events (fragments) as well large events

(stromal cells) were excluded. The analysis was performed on a BDLSRFortessa X20 flow cytometer (BD Biosciences).

Transplantation.

Ebf1$^{-/-}$ FL cells were transduced with GFP tagged EBF1 expression viruses alone or in combination with RFP tagged MYC encoding viruses. Single GFP^{+} or double positive GFP^{+}RFP^{+} cells were sorted by FACS and expanded *in vitro* in B-cell media. 1 million cells were resuspended in 150 μL PBS/1%FCS and transplanted by intravenous injection via the tail vein to 9-11 weeks old pre-conditioned (2.25 Grey) CD45.1 mice. Three to four weeks after transplantation BM was collected as described in [6] and analyzed for the presence of GFP^{+}CD45.2^{+} or GFP^{+}RFP^{+}CD45.2^{+} cells by FACS using the following antibodies. CD45.1-BV605 (A20), CD45.2-AlexaFluor700 (104), CD19-APC (ID3), IgM-APC-Cy7 (RMM-1). All antibodies were from BioLegend.

Ebf1 ChIP-sequencing in 230-238 Pre-B cells

20 million 230-238 cells were fixed at RT in 1 mg/ml DSG (ThermoFisher Scientific) in PBS for 30 min followed by an additional 10 min after addition of formaldehyde to a 1% final concentration. The reactions were quenched by addition of 1/10 volume of 0.125M glycine and the cells were washed in PBS. Nuclei were isolated by 10 min incubation in Nuclei Isolation buffer (50 mM Tris-pH 8.0, 60 mM KCl, 0.5% NP40) + protease inhibitor cocktail (PIC) (Roche) on ice. Pelleted nuclei were dissolved in Lysis buffer (0.5% SDS, 10 mM EDTA, 0.5 mM EGTA, 50 mM Tris-HCl pH 8.0) + PIC followed by 18 rounds of sonication (30s max intensity followed by 30s rest) on a Bioruptor (Diagenode). Sonication was followed by pelleting of debris and the supernatant was transferred to a new tube and chromatin was diluted three times in 1×HBSS (Lonza, Verviersa, Belgium) +PIC followed by another dilution with

2×RIPA buffer (20 mM Tris–HCl, pH 7.5, 2 mM EDTA, 2% Triton X-100, 0.1% SDS, 0.2% Sodiumdeoxycholate, 200 mM NaCl + PIC), followed by the removal of input material. Ten µg per 10^7 cells of antibody Rabbit anti-Ebf1 polyclonal IgG (Millipore, ABE1294) was hybridized to 70µl Protein-G Dynabeads (Life Technologies). ChIP was performed over night at 4°C, and subsequently washed (1 time with 500 µl Low Salt Immune Complex Wash Buffer, 1 time with 200 µl High Salt Immune Complex Wash Buffer, 1 time with 200 µl LiCl Immune Complex Wash Buffer, 2 times with 200 µl TE buffer) and eluted for 6 h at 65°C (20 mM Tris-HCl, pH 7.5, 5 mM EDTA 50 mM NaCl, 1% SDS, 100 µg RNase A and 50 µg proteinase K) treated and finally cleaned up using Zymo ChIP DNA Clean & Concentrator before ChIP-seq library preparation using NEXTflex DNA barcodes (BIOO scientific). 76 bp single read sequencing was performed on an Illumina NextSeq500. The data are deposited in the GEO database (GSE159957).

Ebf1 ChIP-seq data-analysis

ChIP-seq reads from high throughput sequencing were aligned to mouse reference genome mm10 using Bowtie2 [7]. Ebf1 peaks were identified with *findPeaks.pl* against a matched input control using the settings "*-P .1 -LP .1 -poisson .1 -style factor*". Transcription factor peak reproducibility of the duplicate experiments was determined by a HOMER adaptation of the IDR (Irreproducibility Discovery Rate) package [8] (Karmel A. 2015. homer-idr: Second pass updated) according to (https://sites.google.com/site/anshulkundaje/projects/idr).

Reanalysis of public PLAC-seq, ChIP, ATAC-seq and Cut&Run data.

ChIP, ATAC-seq and Cut&Run data was retrieved from the NCBI Gene expression Omnibus GEO (https://www.ncbi.nlm.nih.gov/gds/) (GSE92434, GSE69227, GSE126375, GSE126300, GSE162858 and GSE126300) [9-12] and mapped to reference genome mm10 using Bowtie2

(2.3.4.3) [7,13]. Downstream analyses were performed using the HOMER platform (v4.8-4.10) [14]. BedGraph files were generated using *makeUCSCfile* and resulting files were visualized on the UCSC Genome Browser [15]. FL Wt H3K4me2, H3K4me3 ChIP-seq as well as H3K27ac Cut&Run was processed as described in (GSE162858). BigWig files were generated using *bamCoverage* (v.3.3.1) from the deepTools suite and visualized on the UCSC Genome Browser as described above.

Reanalysis of human B-ALL cell line NALM6 PLAC-seq, ChIP-seq and ATAC-seq data (GSE126300) was performed as described in [12]. The resulting BigWigs and PLAC-seq interactions were uploaded and visualized on the WashU EpiGenome Browser.

Proximity Ligation-Assisted ChIP (PLAC)-sequencing

PLAC-seq was carried out similar to previously reported [12] with minor modifications. Fifteen million FL derived Wt or *Ebf1*$^{-/-}$ pro-B cells were cross-linked with 1% formaldehyde for 5 min, thawed on ice for 5 minutes, re-suspended in ice cold lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.2% NP-40 and Roche protease inhibitors – 11697498001 (PIC)) and incubated on ice for 15 min. Samples were spun for 5 min at 4°C at 2500×g and cell pellet was washed in ice-cold lysis buffer + PIC. The pellet was resuspended in 0.5% SDS and incubated at 62°C for 10 min. The SDS reaction was diluted by the addition of water and 10% Triton X-100 followed by 15 min incubation at 37°C. The samples were digested with 40U MboI restriction enzyme with the addition of 25 μl NEBuffer2 followed by 2h incubation at 37°C with shaking (900 RPM). After restriction enzyme digestion the samples were incubated for 20 min at 62°C to inactivate MboI followed by a cool-down to RT. A fill-in reaction was conducted using 0.3 mM Biotin-14-dATP (ThermoFisher, 19524016), 0.3 mM dCTP, 0.3 mM dTTP, 0.3 mM dGTP and 40U Klenow (NEB, M0210) at 37°C for 1.5h with shaking (900 RPM). Next a Ligation master

mix (1× T4 ligation buffer (NEB, B0202), 1% Triton-X 100, 120 ug BSA (NEB, B9000), 4000 U T4 DNA ligase (NEB, M0202)) was added and samples were incubated at RT with rotation for 2h. The samples were centrifugated 2500×g 5min at 4°C, supernatant was removed and cell pellet resuspended in 250 μl ChIP SDS-lysis buffer (0.5% SDS, 10 mM EDTA, 0.5 mM EGTA, 50 mM Tris-HCl (pH 8.0)) + PIC followed by sonication on a Covaris ME220 for 6 min (Peak power = 75, cycles per burst = 1000, Duty Factor = 15%). Sheared chromatin was centrifuged for 10 min at 4°C, 13000 rpm and supernatants were transferred to new tubes and diluted in 750 μl 1×HBSS + 1 ml 2× RIPA (20 mM Tris–HCl, pH 7.5, 2 mM EDTA, 2% Triton X-100, 0.1% SDS, 0.2% sodium deoxycholate, 200 mM NaCl) + PIC. 1% input was removed from supernatant, and 10 μg H3K4me3 (Millipore, 07-473) antibody pre-adsorbed to 60 μl Protein-G dynabeads in PBS/0.5% BSA were added to the remaining supernatant and incubated at 4°C overnight with rotation. Samples were washed as follows: 2 times with 1 ml Low Salt Immune Complex Wash Buffer (0.1% SDS, 1 % Triton X-100, 2 mM EDTA, 50 mM Tris-HCl pH8, 150 mM NaCl), 2 times with 1 ml High Salt Immune Complex Wash Buffer (0.1% SDS, 1 % Triton X-100, 2 mM EDTA, 50 mM Tris-HCl pH 8.0, 500 mM NaCl), 1 time with 1 ml LiCl Immune Complex Wash Buffer (0.25 M LiCl, 1% Igepal-CA630, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.0), 2 times with 1 ml TE buffer (10 mM Tris–HCl, pH 8.0, 10 mM EDTA) followed by elution of chromatin from magnetic beads with two rounds of 100 μl elution buffer (1%SDS, 100 mM $NaHCO_3$) with shaking (1500 RPM) at RT. Supernatants were transferred to new tubes and chromatin complexes were reverse cross-linked overnight at 65°C with the addition of 250 mM NaCl, 100 µg RNase A (ThermoFisher, EN0531) and 50 µg proteinase K (ThermoFisher, AM2546) followed by Zymo Research ChIP DNA Clean & Concentrator (BIOSITE-D5205) clean up. 25 µL of Streptavidin T1 beads (Thermo Fisher, 65601) were washed with Tween Wash Buffer (5 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween-20) then resuspended in 50 µL of 2× Biotin Binding Buffer (10 mM Tris-

HCl pH 7.5, 1 mM EDTA, 2M NaCl). Beads were added to the samples and incubated at RT for 15 minutes with shaking. After capture, beads were recovered using a magnet and supernatant was discarded. Samples were washed twice by adding 500 µL of Tween Wash Buffer and incubated at 55°C for 2 minutes shaking followed by washing once in 100 μl 1× T4 DNA ligation buffer. Beads were collected on a magnet and resuspended on 100 μl end-repair mastermix (0.5 mM dNTPs (VWR, E636-40UMOLE), 12U T4 DNA Polymerase (NEB, M0203), 50U T4 Polynucleotide Kinase (NEB, M0201), 5U Klenow (NEB, M0210), 1× NEB T4 DNA ligase buffer) followed by incubation at RT for 30 min with shaking (900 RPM). 300 μl Tween Wash Buffer was added, beads were recovered using a magnet and supernatant was discarded. Samples were washed twice by adding 500 µL of Tween Wash Buffer and incubated at 55°C for 2 minutes shaking followed by washing once in 100 μl 1× NEB2 buffer. Beads were collected on a magnet and resuspended in 100 μl A-tailing mastermix (0.5 mM dATP (ThermoFisher), 25U Klenow Exo- (NEB, M0212), 1× NEB2 buffer) and incubated at 37°C for 30 min with shaking (900 RPM). 300 μl Tween Wash Buffer was added and beads were recovered using a magnet and supernatant was discarded. Samples were washed twice by adding 500 µL of Tween Wash Buffer and incubated at 55°C for 2 minutes shaking followed by washing once in 100 μl 1× Fast-Link ligation buffer (Epicenter, LK0750H) followed by NEXTflex DNA barcode ligation (BIOO scientific) using the Fast-link ligation kit (Epicenter, LK0750H). 300 μl Tween Wash Buffer was added and beads were placed on a magnet and supernatant was discarded. Samples were washed twice by adding 500 µL of Tween Wash Buffer and incubated at 55°C for 2 minutes shaking followed by washing once in 100 μl 10mM Tris-HCl pH 8.0, resuspended in 45 μl and a 1:1000 dilution was made for qPCR determination of number cycles needed for final PCR amplification. After final PCR, T1 streptavidin beads were collected on a magnet, supernatant transferred to new tubes and Ampure XP beads (×0.8 sample volume) were used to clean up libraries. PLAC-seq libraries were subject to 2× 75 cycles

of paired-end sequencing on a NextSeq500. The data is deposited to the GEO database (GSE159957).

Processing of PLAC-sequence data and virtual 4C analysis

FL *Wt* pro-B and FL *Ebf1*$^{-/-}$ Pro-B H3K4me3 PLAC-seq experiments were carried out in triplicates and quadruplicates, respectively. To maximize read-depth prior to analysis, fastq-files containing both read 1 and read 2 with their counterparts from the replicates were combined for a total of 517M Trim Galore trimmed paired-end reads per sample. Reads were trimmed in pair-end mode with Trim Galore (0.6.4)(--paired --fastqc --clip_R1 10 --clip_R2 10 --three_prime_clip_R1 3 --three_prime_clip_R2 3) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Trimmed paired fastq reads were preprocessed through the HiC-Pro pipeline (v2.11.1) [16], using Bowtie2 (2.3.5) [7,13] for alignment against mm10 (GRCm38) reference genome, assigning mapped reads to a mm10 MboI restriction map followed by removal of PCR duplicates with Picard tools MarkDuplicates (2.21.4) (https://broadinstitute.github.io/picard/) and subsequently the generation of a list of valid interaction read pairs. Approximately, 230M (Wt pro-B) and 250M (*Ebf1*$^{-/-}$ pro-B) valid interaction pairs were generated.

The valid interaction files, the mm10 MboI restriction digestion map and a bed file containing the mm10 coordinates for the *Myc* gene's TSS extended 2500 bp in both directions, were used as an input to the *make_viewpoints.py* script from the HiC-Pro pipeline. The resulting *Myc*-centric virtual 4C bedGraph files were sorted with the *bedtools sort* command and uploaded to the UCSC genome browser for visualization.

Identification of putative Ebf1 and Pax5 targeted Myc cis-regulatory elements

Visual inspection of the virtual 4C genome browser tracks revealed two regions with and increased interaction frequency with the *Myc* TSS. Together with the regions just upstream of the TSS, the coordinates in these regions were subtracted and used for the downstream analysis. The coordinates (mm10 reference genome) are as follows: TSS; chr15:61,983,046-61,988,606 (5'E, TSS), proximal putative CRE; chr15:62,690,111-62,856,784 (E1), distal putative CRE; chr15:63,566,791-63,791,994 (E2-5). To narrow down the search-space for EBF1 and PAX5 target regions, these regions were further filtered for 230-238 pre-B EBF1 (this paper) and PAX5 (GEO: GSE126375 [12]) binding residing in FL *Wt* and/or *Ebf1*$^{-/-}$ ATAC-accessible regions (GEO: GSE92434 [10]), using *mergePeaks.pl* from the HOMER package. This resulted in a total of 6 regions subjected for further analysis (see *Transient transfections and reporter assays*).

Transient transfections and reporter assays.

50,000 HeLa cells/well, were seeded into a 24-well culture plate. After an overnight incubation the cells were transfected as in [2]. Briefly, plasmids were mixed with 75 µL serum-free medium Optimem (Gibco, Invitrogen), 4 µL PLUS reagent and 1 µL LIPOFECTIN reagent (Invitrogen Life Technologies, Carlsbad, CA). After replacement of cell growth medium with 300 µL of serum-free medium, the DNA-PLUS-LIPOFECTIN mixture was added to the cells. Plates were then incubated at 37°C in 5% CO_2 for 3 hours, after which each well was supplemented with 1 mL RPMI. Cells were harvested 48 hours after transfection and protein extracts were prepared by adding 80 µL cell lysis buffer to each well.

Each transfection included 100 ng of pGL3-based Luciferase reporter plasmids. As a negative control we used a reporter carrying a minimal Fos promoter [1] and as positive control for functional EBF1 expression we used a construct positioning 3 Mb1 (*Cd79a*) promoter EBF1 sites upstream of a TATA-box [2]. Putative EBF1 responsive enhancer elements, (5'E; chr

15:61,983,762-61,984,085, E1; 15:62,698,675-62,699,077, E2; 15:63,619,902-63,620,403, E3; 15:63,636,392-63,637,093, E4; 15:63,657,807-63,658,209, E5; 15:63,678,859-63,679,261) were cloned upstream of the basal Fos promoter. As an internal control and for normalization of the luciferase activities, all transfections included 50 ng Renilla luciferase encoding pRL-0 (Promega) plasmid.

The luciferase assays for all transfections were performed with a Dual-Luciferase Reporter Assay System (Promega) using 20 μL of the total protein extract.

Electrophoretic Mobility Shift Assay (EMSA).

In vitro translation was performed using TNT- combined in vitro transcription translation system (Promega). Oligonucleotides were labeled with γ[^{32}P] adenosine triphosphate (ATP; Amersham Biosciences, Buckinghamshire, United Kingdom) by incubation with T4 polynucleotide kinase (Roche Diagnostics, Mannheim, Germany), annealed and purified on a mini-Quick Spin Oligo Column (Roche Diagnostics, Mannheim, Germany). Nuclear extracts were incubated with labeled probe (20,000 cpm, 3 fmol) for 30 minutes at RT in binding buffer (10 mM HEPES, pH 7.9, 70 mM KCl, 1 mM dithiothreitol, 1 mM EDTA [ethylenediaminetetraacetic acid], 2.5 mM $MgCl_2$, 4% glycerol) with 0.75 μg Poly(dI/dC) (Amersham Pharmacia Biotech, Buckinghamshire, United Kingdom). When EBF binding was investigated, 1 mM $ZnCl_2$ was supplemented in the mixture. The samples were electrophoresed on a 6% polyacrylamide Tris boric acid EDTA (TBE) gel, which was dried and subjected to autoradiography.

Sequences of oligonucleotides used for electrophoretic mobility shift assays (EMSAs) are presented in Supplementary data sheet 3.

Identification, cloning and virus production for gRNA for CRISPR/Cas9 mediated editing of Ebf1-binding sites in putative Myc enhancer elements.

CRISPR guides were identified by overlaying the genomic coordinates of 6 putative Ebf-1 binding sites with potential CRISPR/Cas9 guides in Benchling (Biology Software. 2020) using the single guide, mm10 genome and 3' NGG PAM settings. Guides which promote Cas9 mediated cutting within the Ebf1 motifs would be selected for cloning into pAW13.lentiGuide-emCherry plasmid (lentiguide-mCherry, Addgene plasmid # 104375; http://n2t.net/addgene:104375 ; RRID:Addgene_104375, a gift from Richard Young). Lentiviruses were produced by transfecting 293T-HEK cells with lentiguide-mCherry plasmids as well as psPax2 and pMD2G packaging plasmids (originally from the Didier Trono Lab) together with X-tremeGENE HP DNA Transfection Reagent (Sigma) according to the manufacturer's instructions. The resulting virus was harvested after 54-64 hours and concentrated using the Lenti-X concentrator according to the manufacturer's instructions (Takara Bio).

Ebf1 binding site screen.

Hind bones from four individual iCas9 mice aged 14-21 weeks were crushed in a mortar and the cell suspension passed through a 50um filter. KIT^+ cells and subsequently purified by magnet-activated cell sorting column using anti-CD117 immunomagnetic beads (Miltenyi Biotec). 2 million iCas9 KIT^+ BM cells were transduced with a mix of lentiviruses carrying gRNAs targeting 6 Ebf1-motifs as well as positive and negative control gRNAs (Supplementary data sheet 4). Briefly, non-tissue coated plates were coated with 40 µg/ml retronectin (Takara Bio) overnight at 4°C, blocked with a 2% BSA solution at RT for 30 minutes and washed with PBS. Thereafter, viruses were added, and plates spun at 2000×g and 32°C for 2 hours. The remaining viral supernatant was aspirated, and wells were washed with PBS. Cells were then

added to the virus-coated wells and plates spun at 300×g and 32°C for five minutes. After an overnight culture, transduced BM cells were put in co-culture with OP9 stromal cells. At day two, cells were harvested and stained with CD45-AF700 (30-F11, BioLegend) and CD19-PECy7 (eBio1D3, eBioscience). Thereafter, $CD45^{+}mCherry^{+}CD19^{+}$ cells were sorted on a FACSAriaIII (BD) and each sample split in two for DNA preparation (QIAamp DNA micro Kit, Qiagen) as well as for continued co-culture with OP9 stromal cells under Doxycyclin (DOX, Sigma) containing conditions to induce Cas9 expression (0.1 μg/mL DOX every two days). Following 6 days of DOX treatment, $CD45^{+}mCherry^{+}CD19^{+}$ cells were stained and sorted as above for DNA isolation (QIAamp DNA micro Kit (Qiagen). DNA samples pre-DOX (day2) and post-DOX (day8) were sequenced for distribution of gRNAs (n=4) and a duplicate of samples sequenced for CRISPR editing at each targeted Ebf1 binding site. Throughout the cell culture, cells were kept in OptiMEM media supplemented with 10% fetal calf serum, 50μg/mL Gentamycin, 50μM β-mercaptoethanol, 10-50ng/mL KIT ligand, 10-50ng/mL Fms-like tyrosine kinase 3 ligand (FLT3L) and 10-50ng/mL interleukin-7 (IL-7). All cytokines were obtained from Peprotech.

Investigation of Ebf1 binding site Myc E2 by Crispr editing.

BM cells from 3 individual 11-week-old iCas9 mice were infected overnight on virus-coated plates (as described in the "Ebf1 binding site screen" section) with sgRNA lentiviruses targeting Ebf1 binding site Myc E2 (or controls) and kept in OptiMEM media supplemented with 10% heat-inactivated fetal calf serum, 50μg/mL Gentamycin, and 50μM β-mercaptoethanol, 50ng/mL KIT ligand, FLT3L and IL-7. After the overnight culture, transduced BM cells were transferred to co-culture with OP9 stromal cells and $CD45^{+}mCherry^{+}CD19^{+}$ cells were sorted at day two. Each sorted sample was split in 3 stromal free cultures of 300-500 cells each using the same media as above. After an additional 24, 48 and 96 hours, one set of cultures was

counted, and RNA samples isolated for Q-PCR quantification of Myc transcripts. Furthermore, at 96 hours DNA was isolated (QIAamp DNA micro Kit (Qiagen) for evaluation of CRISPR editing efficiencies by NGS sequencing.

Library production and sequencing to investigate targeting efficiencies.

Sequencing libraries were constructed using a nested PCR method in which regions spanning integrated gRNA protospacers or CRISPR edited sites were amplified using primers containing sequencing adaptors followed by a nested PCR to add Nextera XT indices (Illumina, Supplementary data 2). Briefly, DNA regions were amplified using the NEB Next UltraII Q5 Master mix (NEB) using the cycle conditions 98°C 30s, (98°C 10s, 58-60°C 25s, 72°C 30s) × 20 repeats, 72°C 10 mins for the first PCR and conditions 72°C 3 mins, 98°C 1 min (98°C 10s, 58-60°C 25s, 72°C 30s) × 10 repeats, 72°C 10 mins for the second PCR. PCRs were followed by at least two rounds of 1× AMPure bead purification according the manufactures instructions for a left sided size selection (Beckman Coulter). Library quality was verified with a HS DNA chip run on a Bioanalyzer (Agilent) and the DNA concentrations measured with a Qubit (ThermoFisher). 4 pM libraries with 20% PhiX (Illumina) were sequenced on a MiSeq using a MiSeq Nano v2 500 cycle kit (Illumina). CRISPR editing efficiencies in the PCR amplicons were analyzed using Crispr-Dav (Wang, Bioinformatics, 2017). Briefly, the reads must span the sgRNA sequence region. Reads that only overlap with the sgRNA sequence partially were ignored. Wild type reads refer to reads that have no insertions or deletions (indel) in the sgRNA region. Indel reads have insertion and/or deletion of at least one base inside the sgRNA region. The deleted sequence can extend continuously beyond the sgRNA region. For the analysis of the type and distribution of indels generated at the Myc E2 site using gRNAs 106 and 107 (Figure S2G, Supplemental Data Sheet 5-6), the Crispr-Dav analysis was used to define the regions that encompassed all, ≥1% and >5% of indels in the sequenced amplicons. These indel

regions were transferred to genomic coordinates and a TF motif search was performed (see "*Enhancer E2 motif scanning"* section. The impact of gRNA guided editing in different regions was investigated by establishing the proportion of indels at the different regions (all, ≥1% and >5% of indels, Figure S2G) as mean (SD) of all indels found in the NGS data. In order to look for potential functional EBF1 binding sites (Figure S2D), the distribution of library gRNAs was calculated as the frequency of sequences found containing each gRNA pre DOX induction and related to how the distribution changed with Cas9 induction.

Enhancer E2 motif scanning

To identify other potential TF binding sites targeted by the E2 enhancer CRISPR/Cas9 perturbation (chr15:63,620,136-63,620,164), we performed motif scanning of this region with *gimme scan* from the GimmeMotifs suite (v. 0.15.2) [17]. G*imme scan* was run with default settings, using mm10 as the reference genome, against the JASPAR2020 CORE vertebrate database (http://jaspar.genereg.net/download/CORE/JASPAR2020_CORE_vertebrates_non-redundant_pfms_transfac.txt). Putative binding sites with an FDR of 0.01 were identified and the expression of the relevant transcription factor in B-cell progenitors was explored using RNA-seq data (GSE136238). Only sites for factors expressed in B-cell progenitors are indicated in Figure S2G.

CRISPR/Cas9 mediated knock-out of Pax5 in ProB cells.

MSCV-Cas9-GFP mediated gene inactivation was performed similar to [18], with the difference that a modified version of the pSuper vector was used for cloning of protospacer sequences. The Retroviral "pSuper-sgRNA-mTurquoise2" expression vector was made as follows: for the sgRNA expression cassette, human U6 promoter (Addgene ID 52961), two BbsI restriction sites separated with a 1.1kb stuffer DNA sequence and optimized long stem loop sgRNA

scaffold (Addgene ID 51024) were assembled by Gibson cloning. For fluorescent protein marker expression cassette, EF1alpha shorter promoter (Addgene ID 52961), Tet-on RTA activator (not used in this study), P2A self-cleavage peptide and mTurquoise2 fluorescent protein cDNA were assembled by Gibson cloning. Retroviral shRNA expression vector "pSuper retro puro Scr shRNA" (Addgene ID 30520) was digested with XhoI and DraIII to remove the original H1-shRNA and puromycin expression cassette, followed by insertion of the sgRNA and fluorescent protein expression cassettes above are into the pSuper backbone by Gibson cloning. Protospacer sequences targeting murine PAX5, were ligated into BbsI digested pSuper (generated CACC and AAAC overhangs added to the 5'-end of the sense and antisense sequence, respectively). The following sequences were used 1) GATGGAGTATGAGGAGCCCG 2) GAGTGCTGTCTCTCAAACACG 3) GAGAGTAGCTGCCCTGTCCAG 4) GTTCTTGGCAGGTAAAGGCAC 5) GTCGGTGAGCACCGACTCCGC 6) GCTGCTCCCGATGTCAGCGG. MSCV-Cas9-GFP and pSuper retroviral supernatants were produced in Platinum-E cells as described above (see Transduction of *Ebf1*$^{-/-}$ FL cells). Supernatants were collected 48h post transfection and 500 µl was used to transduce EBF1-ER and Ebf1-expressing BM-pro-B cells by spin infection [6]. Puro resistance Cas9 expressing cells were generated by 4 days of selection in puromycin (Sigma-Aldrich) followed by secondary infection with *Pax5* guides. After 4 days CD19^{-} cells were sorted and expanded.

RNA-sequencing and data analysis from cultured pro-B cells.

Total RNA was isolated using RNAeasy Micro Kit (Qiagen, Hilden, Germany) according to manufacturer's recommendations. Libraries were constructed using NuGEN's Ovation Ultralow Library systems (NuGEN Technologies, San Calros, CA) and were subsequently subjected to 76 cycles of NextSeq500 sequencing (Illumina, San Diego, CA). For analysis of

RNA-Seq experiments the reads were aligned to mouse reference genome (mm10 / GRCm38) using STAR (2.6.0b-1) [19]. For analysis of statistically significance among differently expressed genes the data was analyzed using the HOMER platform (v4.8) [14]. For differential expression the commands *analyzeRepeats.pl* with the *–noadj* option followed by the *getDiffExpression.pl* command using DESeq2 [20] was used for differential expression analysis. Cut-off for differential expression was set to a fold change of 2 (*-log2fold 1*) and an adjusted p-value of equal to or less than 0.05 (*-fdr 0.05*).

Gene set enrichment analysis (GSEA)

Gene set enrichment analysis was performed using GSEA software (v. 4.0.1) with default settings except for that Permutation type was set to *gene_set* and Collapse/Remap to gene symbols set to *No Collapse*. Reads Per Kilobase of transcript per Million (RPKM) normalized reads were extracted from FL *Ebf1*$^{-/-}$ and FL *Ebf1*$^{-/-}$+ PAX5-pMIG HOMER tag directories using *analyzeRepeats.pl* (settings: *-count exons -condenseGenes -rpkm*) used to create a GSEA expression dataset. This dataset was tested against the MSigDB gene set HALLMARK_MYC_TARGETS_V1 (https://www.gsea-msigdb.org/gsea/msigdb/genesets.jsp?collection=H), which was accessed and downloaded on the 13th of December 2019.

Data availability.

The RNA-seq analysis of B-cells lacking EBF1 activity has been deposited in GEO with the accession number GSE136238. PLAC-seq data, EBF1 Chip Seq data and RNA seq data from transduced cells are deposited in GSE159957.

References.

1. Sigvardsson M, O'Riordan M, Grosschedl R. EBF and E47 collaborate to induce expression of the endogenous immunoglobulin surrogate light chain genes. *Immunity*. 1997;7(1):25-36.

2. Smith EM, Akerblad P, Kadesch T, Axelson H, Sigvardsson M. Inhibition of EBF function by active Notch signaling reveals a novel regulatory pathway in early B-cell development. *Blood*. 2005.

3. Pear WS, Miller JP, Xu L, et al. Efficient and rapid induction of a chronic myelogenous leukemia-like myeloproliferative disease in mice receiving P210 bcr/abl-transduced bone marrow. *Blood*. 1998;92(10):3780-3792.

4. Gao H, Lukin K, Ramirez J, Fields S, Lopez D, Hagman J. Opposing effects of SWI/SNF and Mi-2/NuRD chromatin remodeling complexes on epigenetic reprogramming by EBF and Pax5. *Proc Natl Acad Sci U S A*. 2009;106(27):11258-11263.

5. Ahsberg J, Ungerback J, Strid T, et al. Early B-cell factor 1 regulates the expansion of B-cell progenitors in a dose-dependent manner. *J Biol Chem*. 2013;288(46):33449-33461.

6. Somasundaram R, Ahsberg J, Okuyama K, et al. Clonal conversion of B lymphoid leukemia reveals cross-lineage transfer of malignant states. *Genes Dev*. 2016;30(22):2486-2499.

7. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25.

8. Li QB JB, Huang H, Bickel PJ. Measuring reproducibility of high-throughput experiments. . *Ann Appl Stat* 2011;5(5):1752-1779.

9. Revilla IDR, Bilic I, Vilagos B, et al. The B-cell identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis. *Embo J*. 2012;31(14):3130-3146.

10. Jensen CT, Ahsberg J, Sommarin MNE, et al. Dissection of progenitor compartments resolves developmental trajectories in B-lymphopoiesis. *J Exp Med*. 2018;215(7):1947-1963.

11. Ungerback J, Ahsberg J, Strid T, Somasundaram R, Sigvardsson M. Combined heterozygous loss of Ebf1 and Pax5 allows for T-lineage conversion of B cell progenitors. *J Exp Med*. 2015;212(7):1109-1123.

12. Okuyama K, Strid T, Kuruvilla J, et al. PAX5 is part of a functional transcription factor network targeted in lymphoid leukemia. *PLoS Genet*. 2019;15(8):e1008280.

13. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359.

14. 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. *Mol Cell*. 2010;38(4):576-589.

15. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996-1006.

16. Servant N, Varoquaux N, Lajoie BR, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol*. 2015;16:259.

17. van Heeringen SJ, Veenstra GJ. GimmeMotifs: a de novo motif prediction pipeline for ChIP-sequencing experiments. *Bioinformatics*. 2011;27(2):270-271.

18. Hosokawa H, Ungerback J, Wang X, et al. Transcription Factor PU.1 Represses and Activates Gene Expression in Early T Cells by Redirecting Partner Transcription Factor Binding. *Immunity*. 2018;48(6):1119-1134 e1117.

19. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.

20. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.