Supplemental Material

Supplemental material includes experimental procedures, six tables,

eight figures with legends and references.

Super-enhancer reprogramming

drives a B cell-epithelial transition and high-risk leukemia

Hu et al.

Supplemental Experimental Procedures

Animals, Cell Sorting and Pre-B Cell Culture Studies

Animals

WT, *IkE5*$^{fl/fl}$;CD2-Cre and *IkE5*$^{fl/fl}$*;Rosa26-ERT2-Cre* mouse strains were used for isolation of bone marrow large pre-B cells. All animal experiments were done according to protocols approved by the Subcommittee on Research Animal Care at Massachusetts General Hospital (Charlestown, MA) and the Institutional Animal Use and Care Committee of the University of California, Irvine, and in accordance with the guidelines set forth by the National Institutes of Health.

Cell Sorting

For isolation of large pre-B cells, also described as early pre-B cells that express pre-BCR and are highly proliferative, (reviewed by (Hardy and Hayakawa, 2001)), BM cell preparations were depleted of cells binding to anti-Ter119, anti-Mac-1, anti-Gr-1, anti-IgM, anti-CD3, anti-CD8α, anti-TCRβ and anti-DX5 (Table S1) by removal with magnetic beads conjugated to BioMag goat anti-rat IgG (310107; Qiagen). The cells remaining after depletion were labeled with fluorochrome-conjugated monoclonal antibodies to B cell markers (anti-CD19 (25-0193; eBiosciences), anti-CD43 (553270; BD), anti-BP1 (11-5995; Ebiosciences), anti-CD25 (553075; BD) and anti-CD2 (553112; BD)) and used for flow cytometry. Large pre-B cells from WT, *IkE5*$^{fl/fl}$ CD2-Cre, or *IkE5*$^{fl/fl}$ *ERT2-Cre* were sorted as CD19^{+}CD43^{+}BP1^{+} using a MoFlo-Legacy (Cytomation) cell sorter.

Pre-B Cell Cultures

WT and *IkE5*$^{\Delta/\Delta}$ (IKDN) sorted large pre-B cells were co-cultured with OP9 stroma as previously described (Joshi et al., 2014). Non-adherent pre-B cells were collected by several washes with fresh media. Stroma-adherent large pre-B cells from WT or IKDN

cultures were detached with 0.2% trypsin (EDTA free) for 2 minutes at room temperature. The harvested large pre-B cells were ~ 99% of the cells in suspension.

Deletion of Ikaros in Large Pre-B cells In Vitro

Large pre-B cells were sorted from the BM of *IkE5$^{fl/fl}$;Rosa26-ERT2-Cre* mice, and expanded for 6 days on OP9 stromal cells in the presence of 5ng/ml of IL7. They were then re-plated onto stromal cells with 0.2 μM of 4-OHT (Sigma H7904) or with an equal volume of DMSO as control. Cells were harvested at different time points (1-16 days) of culture in 4-OHT.

Knockdown of Transcription Factors in IKDN Pre-B cells in vitro

Small hairpins targeting *Lhx2*, *Lmo2* and *Yap1* were cloned into the pLKO.1 lentiviral vector. The *Ebf1* shRNA vectors were previously reported by (Griffin et al., 2013). All hairpin sequences are listed in Table S2. shRNA and packaging vectors were co-transfected into HEK 293T/17 cells (ATCC® CRL-11268), and viral supernatants were collected at 48 and 72 hours post-transfection and concentrated by ultracentrifugation through a 20% sucrose cushion. For infection, large adherent pre-B cells were mixed with viruses and spun at 1500rpm for 30 minutes onto Retronectin-coated wells of a 24 well plate and then cultured for 5 hours in the presence of 5 ng/ml IL7 and 4 μg/ml of polybrene (Sigma, H9268-10G). Pre-B cells were then trypsinized and transferred onto a new plate seeded with OP9 stroma. After 24 hours, the media was removed and replaced with fresh media containing 1μg/ml puromycin (Sigma, P8833). Cells were selected for 3-5 days in puromycin and used for further analysis.

Colony Forming Assay

Adherent IKDN pre-B cells infected with shRNAs for *extra-lineage* and B cell transcription factors were selected with puromycin for three days and then re-plated onto OP9 stroma-containing 96-well plates at two fold serial dilutions. Cells at concentrations of 1-16 cells per well and with 16 to 24 replicates per concentration were grown with 5 ng/ml of IL-7. Wells with pre-B cell colonies were scored between days 8-10. The mean frequency of colony forming cells and p-values were calculated by the ELDA method (Hu and Smyth, 2009).

Cell Growth Analysis

IKDN large adherent pre-B cells infected with shRNAs and selected with puromycin were plated in 96 well- (counted on day 1), 48 well- (counted on day 2), 12 well- (counted on day 3) or 6 well- (counted on day 4) plates in the presence of OP9 stroma and 5 ng/ml of IL7. From day 1 to day 4 after re-plating cells were harvested and cell number, cell cycle and apoptosis were determined. Growth curves and statistical analysis on growth were performed with the GraphPad Prism software. P values were calculated by two-tailed unpaired *t*-test with Welch's correction.

Cell-Cycle Analysis

Cells harvested at indicated time points were fixed in 70% ethanol overnight at 4^0C. Fixed cells were stained with propidium iodide (PI) staining buffer (250 μg/ml

RNaseA, 50 μg/ml PI) for 30 min at 37^0C and the DNA content was detected by FACS canto. Data was analyzed with FlowJo software (Tree Star).

Cell Adhesion to Fibronectin

Adhesion assays were performed as previously described (Joshi et al., 2014). Briefly, non-TC treated plates (BD) were coated with 10 μg/ml fibronectin (Invitrogen) overnight at 4^oC. After blocking the plates with 2% BSA for 1 h, equal number of cells were plated and incubated for one hour. At the end of assay, unbound and bound cells were harvested and enumerated. Percent adhesion was calculated by the ratio of bound cells over total cells used in assay.

Chromatin Immunoprecipitation (ChIP) and Analysis

ChIP

ChIP was performed according to previously described protocols (Zhang et al., 2011) with some modifications. Cells were fixed in 1% of fresh formaldehyde for 10 minutes at room temperature, quenched with glycine and washed twice with ice-cold PBS. Cells were re-suspended in lysis buffer 1 (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) for 10 minutes at 4^0C, pelleted, and then re-suspended in lysis buffer 2 (200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris pH8) for 5 minutes at 4 ºC. Cells were finally re-suspended in RIPA buffer (50mM Hepes pH 7.9, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.2% SDS, 0.5 mM PMSF, and 1x Protease inhibitor cocktail (Roche)) to a final concentration of $5x10^7$ cells/ml. Chromatin was sonicated to an average size of 350bp with a Branson sonifier 450 equipped with a micro probe. Chromatin was cleared by centrifugation at 20,000xg for 10 minutes, and then incubated with 5-10μg of antibodies pre-bound to Dynabeads protein G (Life technologies), and rotated overnight at 4^0C. The antibodies used for ChIP are listed in Table S2. Chromatin from 5-$10x10^6$ cells was used for each histone modification, and chromatin from 2-$10x10^7$ cells was used for each transcription factor. ChIPs for each histone modifications and transcription factor were performed in parallel with WT and IKDN pre-B cell chromatins. 3 times more chromatin was used for TEAD ChIPs in WT and E2A ChIPs in IKDN pre-B cells, because of their low respective expression in these cells. After extensive washing on beads, bound chromatin was eluted and de-crosslinked in 300μl of elution buffer (50 mM Tris pH8, 10 mM EDTA, 1% SDS, 0.3M NaCl) overnight at 65^0C. Proteinases K was added to a final concentration of 200 μg/ml and incubated at 45^0C for 2 hours. DNA was ethanol precipitated, re-suspended in TE and purified using the DNA Clean & Concentrator-5 kit (Zymo Research).

ChIP-Seq Libraries, Genome Mapping and Data Analysis

Immunoprecipitated (ChIP) DNA and Input DNA were prepared for sequencing according to a modified version of the Illumina/Solexa Genomic DNA protocol. Briefly, 2.5 to 40ng DNA was end repaired, end adenylated, and then ligated with Illumina Truseq indexed adaptors. The ligated DNA was purified with AMPure XP beads(Beckman Coulter), and then amplified with KAPA HiFi DNA Polymerase (KAPA

Biosystems) for 8 to 13 cycles. After amplification, the library DNA was separated on a 2% agarose gel, and DNA fragments in the 200-500 bp range were purified with a gel DNA recovery kit (Zymo Research). The purified DNA was diluted to 10 nM, and multiplexed for sequencing in a HiSeq2000 sequencer at the Bauer Center Systems Biology Core at Harvard University. Image analysis and base calling was performed using the Illumina Hiseq 2000 software. Raw sequencing data sets were uploaded to DNAnexus, a cloud-based genome informatics & data management platform.

Read alignment was performed to the mouse mm10 genome using BWA (Li and Durbin, 2009) or a modified version of STAR (Dobin et al., 2013). Picard CollectMultipleMetrics tools (Broad Institute) were used for read mapping analysis. Sequencing depth was tested by analyzing high quality biological replicates and by determining if independent data sets when combined yielded a significant higher number of new peaks.

Transcription factor peaks, histone modification enriched regions and gene body coverage were initially identified using both MACS and findPeaks (HOMER) algorithms (Heinz et al., 2010; Zhang et al., 2008) with input chromatin as control**.** Comparison between the two methods for both transcription factors and histones indicated that 90% of peaks or regions identified by MACS were also identified by the HOMER method. We continued our analysis with the HOMER peak finding and analysis tools. Peaks that were differentially enriched in either WT or IKDN pre-B cell chromatin were obtained by setting one condition as test and the other as input control. Peaks were identified with a threshold of false discovery rate of 0.001 for tag clustering, a 4-fold increase in tag density relative to input control in both local and average distributions (Poisson distribution p-value threshold of 0.0001) and for a 4-fold increase in tag density at peaks relative to a surrounding 10kb region (local Poisson distribution p-value threshold of 0.0001).

Identification of transcription factor super-clusters was performed using the HOMER findPeaks style-super method as described by (Hah et al., 2015) in a highly similar manner to the method described by (Whyte et al., 2013). Hierarchical and K-means clustering of ChIP-seq data sets was performed using NGS.PLOT (Shen et al., 2014) or the Homer AnnotatePeaks.pl program (Heinz et al., 2010) combined with Cluster 3.0 (Michael Eisen) and Java Tree View (Alok J. Saldanha) analysis. Histograms of tag densities for histone modifications and transcription factors over regulatory sites or gene bodies were plotted using the NGS.PLOT software (Shen et al., 2014).

The HOMER *de novo* motif discovery algorithm was used to analyze binding sites for frequently occurring DNA sequence (SQ) motifs. For data visualization, primary sequencing data were sorted indexed and uploaded onto a local IGV or a cloud based-Biodalliance genome browser (www.biodalliance.org). Venn diagrams were generated using the on line based Venny Tool (Juan Carlos Oliveros). The R Bioconductor packages implemented through RStudio was used for statistical analysis and for extraction and plotting of data from NGS-generated data files.

Statistical Analysis

The R Bioconductor packages implemented through RStudio was used for statistical analysis and for extraction and plotting of data from NGS-generated data files. The statistical significance between two array groups was determined by Student's *t*-test.

In order to determine whether variances between two groups were equal or unequal, F-test was performed.

Gene expression analysis

Construction of RNA-Seq Libraries, Gene Expression and Pathway Analysis
RNA was extracted with Trizol (Invitrogen) and purified using the PureLink RNA mini kit (Ambion). The Truseq RNA sample prep kit was used for construction of cDNA libraries for RNA-sequencing (Illumina). The cDNA libraries were ligated with indexed primers and amplified by PCR for 10 cycles. The amplified libraries were multiplexed and sequenced at the Systems Biology Lab, Harvard University. Alignment to the mouse mm10 assembly was conducted by STAR (Dobin et al., 2013). Normalization and differential expression was performed using DeSEQ2 through a HOMER implementation of the package in R. Heatmaps of normalized tags for gene subsets across WT and IKDN pre-B cell populations were generated with the Cluster and Java Tree View software. Gene Ontology analysis of deregulated genes in IKDN pre-B cells was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID).

Supplemental Tables

Supplemental Table S1. Genes down-regulated in IKDN preB cells with and without superenhancers.

Supplemental Table S2. Genes up-regulated in IKDN preB cells with and without superenhancers.

Supplemental Table S3. Transcription factor knockdown in IKDN preB cells and gene expression changes.

Supplemental Table S4. Gene expression changes upon *in vitro* induction of IkE5 deletion in WT adherent pre-B cells.

Supplemental Table S5: Hairpin sequences, Related to Figure 6

shRNA	Hairpin Sequence
Lmo2-sh1	CCATAGTAACTGACAAGATTA
Lmo2-sh2	CGGGTGAAAGACAAAGTGTAT
Lhx2-sh1	TTCACATGCACAACGTGTAAC
Lhx2-sh2	GGACAATGAAGTCTTACTTTG
Ebf1-sh1	AAGTCACACTGTCGTACAAGT
Ebf1-sh2	GCGCGACTGTGATCATCATAG
Yap-sh1	GCAGACAGATTCCTTTGTTAA
Yap-sh2	GAAGCGCTGAGTTCCGAAATC
shGFP	GTCGAGCTGGACGGCGACGTA

Supplemental Table S6: Antibodies, Related to Figures 1-7

Antibody	Vendor	Catalog number
H3K4me1	Active Motif	AM39297
H3K4me2	EMD Millipore	07-030
H3K4me3	Active Motif	AM39155
H3K27ac	Abcam	ab4729
H3K27me3	EMD Millipore	07-449
H3K36m3	Abcam	ab9050
RNA POLIIS5p	Abcam	Ab5131
LEF1	Santa Cruz Biotech. Inc	sc-8591x

ETS1	Santa Cruz Biotech. Inc	sc-350x
LHX2	Santa Cruz Biotech. Inc	sc-19344x
LMO2	Novus Biologicals	NB110-78626
SUZ12	Cell Signaling Technology	3737s
EZH2	Active Motif	AM39639
PAX5	Santa Cruz Biotech. Inc	sc-1974x
TEAD1	BD Transduction Laboratories	610923
Pan-TEAD	Cell Signaling Technology	13295s
RUNX1	Abcam	Ab23980
E2A	Santa Cruz Biotech. Inc	sc-349x
EBF1	Abnova	H00001879-M01
IRF4	Santa Cruz Biotech. Inc	sc-6059x
TBX19	Jacques Drouin Lab (IRCM,Canada)	Home-made
IKAROS	EMD Millipore	MABE912 & MABE913
AIOLOS	9D10	Homemade mouse monoclonal
MED1	Bethyl Laboratories	A300-793A

Supplemental References

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Griffin, M.J., Zhou, Y., Kang, S., Zhang, X., Mikkelsen, T.S., and Rosen, E.D. (2013). Early B-cell factor-1 (EBF1) is a key regulator of metabolic and inflammatory signaling pathways in mature adipocytes. The Journal of biological chemistry *288*, 35925-35939.

Hah, N., Benner, C., Chong, L.W., Yu, R.T., Downes, M., and Evans, R.M. (2015). Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs. Proceedings of the National Academy of Sciences U.S.A., E297-302.

Hardy, R.R., and Hayakawa, K. (2001). B cell development pathways. Annual review of immunology *19*, 595-621.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular Cell *38*, 576-589.

Hu, Y., and Smyth, G.K. (2009). ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. Journal of immunological methods *347*, 70-78.

Joshi, I., Yoshida, T., Jena, N., Qi, X., Zhang, J., Van Etten, R.A., and Georgopoulos, K. (2014). Loss of Ikaros DNA-binding function confers integrin-dependent survival on pre-B cells and progression to acute lymphoblastic leukemia. Nature immunology *15*, 294-304.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC genomics *15*, 284.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. In Cell (Elsevier Inc.), pp. 307-319.

Zhang, J., Jackson, A.F., Naito, T., Dose, M., Seavitt, J., Liu, F., Heller, E.J., Kashiwagi, M., Yoshida, T., Gounari, F.*, et al.* (2011). Harnessing of the nucleosome-remodeling-

deacetylase complex controls lymphocyte development and prevents leukemogenesis. Nature immunology *13*, 86-94.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Supplemental Figure S1. Histone modification and transcription factor enrichment in WT and IkE5$^{\Delta/\Delta}$ stromal-adherent large pre-B cells, related to Figure 1. (*A*-*B*) Number of enriched regions or peaks from ChIPs of histone modifications or transcription factors obtained with the Homer findPeaks algorithm (-style histone or factor) using WT or *IkE5*$^{\Delta/\Delta}$ (IKDN) input chromatin as control (Heinz et al. 2010). Differentially enriched peaks or regions between WT and IKDN pre-B cells were obtained using the alternate ChIP sample, e.g. IKDN or WT as the input chromatin control. (*C*-*D*) The approximate ChIP efficiency for histone modifications or transcription factors was calculated by the Homer findPeaks algorithm as the fraction of reads found in enrichment regions compared to random enrichment at genomic regions of similar size.

Supplemental Figure S2. Super-clusters of individual B cell transcription factors in WT and association with down-regulated genes in IKDN stromal-adherent large pre-B cells , related to Figure 2. (*A*) Clusters of chromatin enrichment peaks for individual B cell transcription factors, defined as Super-clusters (SC), were identified using a Rank order super-enhancer algorithm described by (Hah et al. 2015). Transcription factor enrichment peaks were first identified and then stiched together into 12.5 kB regions, normalized to input chromatin, sorted by read score for each region and plotted in ascending order. SC were identified as regions that fall beyond the point where the slope differential is greater than 1. The number of SC over the total number of binding sites for each transcription factor is shown. (*B*) Circular histogram depicting the relative frequency of association of 127 down-regulated genes in IKDN pre-B cells with

individual transcription factor (IKAROS, EBF1,PAX5, E2A, IRF4) SCs. The relative percentages of genes with different SC and overlap are listed. The frequency of genes with SC defined by IKAROS (71/127; 56%) and other transcription factors is indicated by the green box. (*C*) Frequency of individual transcription factor occupancy of the 127 down-regulated genes with SC through their own SC (SC) or by association with another factor's SC (BS in SC) is shown in concentric circles on the left side. Frequency of transcription factor occupancy at the remaining 483 down-regulated genes without SC is shown on the right side. The order of circle appearance (inner to outer) indicates the frequency of occupancy in descending order.

Supplemental Figure S3. The active chromatin state of pre-B cell enhancers is IKAROS-dependent, related to Figure 3. (*A*) Heatmap of expression of down-regulated genes associated with IKAROS super-enhancers in ex vivo (Prim) and cultured (Cult) WT and IKDN pre-B cells. Genes key to pre-B cell signaling and differentiation are highlighted. (*B-D*) Comparative analysis of histone modifications and transcription factor enrichment at transcription factor bindings sites associated with down-regulated genes with regular enhancers (R-BS), as described in Figure 2E. (*B*) Read density (count, reads per million mapped reads) for H3K4me2, H3K27Ac, RNApII and MED1 in WT and IKDN pre-B cells is shown in the vicinity of R-BS (±15 kb from 5' and 3' end of BS). (*C*) Heatmaps of K-means clustering of reads from ChIPs for transcription factors at R-BS (±15 kb from 5' and 3' end of BS). (*D*) Read density (count, reads per million mapped reads) for B cell transcription factor enrichment in WT and IKDN pre-B cells is shown in the vicinity of R-BS (±15kb from 5' and 3' end of BS).

Supplemental Figure S4. Loss of AIOLOS chromatin enrichment in IKDN pre-B cells. (*A*) Read density (count, reads per million mapped reads) and (*B*) heatmaps generated by K-means clustering of reads from ChIPs for AIOLOS and IKAROS transcription factor enrichment in WT and IKDN pre-B cells shown in the vicinity of AIOLOS enrichment sites identified in WT large pre-B cells (±2 kb from AIOLOS peak center). Note that the difference in IKAROS and AIOLOS enrichment is partly due to their difference in protein expression levels with AIOLOS strongly induced after transition from the large to the small pre-B cell stage. (*C*) Venn diagram depicting overlap of genes bound by IKAROS, AIOLOS, or both factors in WT large pre-B cells.

Supplemental Figure S5. Loss of IKAROS causes induction of extra-lineage transcription factors and activation of a feed-forward regulatory loop that inhibits Polycomb-mediated repression, related to Figure 4. (*A-B*) Genome browser tracks at the *Lmo2 and Lhx2* genes are shown for ChIP-seq data sets for IKAROS, LHX2, LMO2, TEAD (1/2), EBF1, H3K4me1, H3K36me3 and H3K27me3 in WT and IKDN pre-B cells. Binding of the *extra-lineage* transcription factor genes by their own protein products and by other members of this group of transcription factors indicates the induction of an integrated feed-forward loop of regulation that is also supported by studies in Figures 6-7. Increase in gene occupancy by these factors in IKDN pre-B cells correlated with local acquisition of enhancer chromatin marks (H3K4me1), loss in PRC2 activity (H3K27me3) from the promoter and gene body and an increase in transcriptional elongation (H3K36me3). (*C-D*) Heatmaps were generated by K-means clustering of

reads from ChIPs for histone modifications (H3K4me1, H3K4me2, H3K4me3, H3K27Ac) and transcription factors (RNApII, IKAROS, LHX2, LMO2, TEAD, PAX5, EBF1 and IRF4) centered (arrow) at *de novo* LMO2 (*C*) or LHX2 (*D*) peaks (C1-C5). *Extra-lineage* and B cell transcription factors are marked respectively in red and blue respectively. Upward arrows on the left side of the heatmaps indicate clusters that show increase in active enhancer histone marks in IKDN preB cells with arrow weight indicating strength of induction.

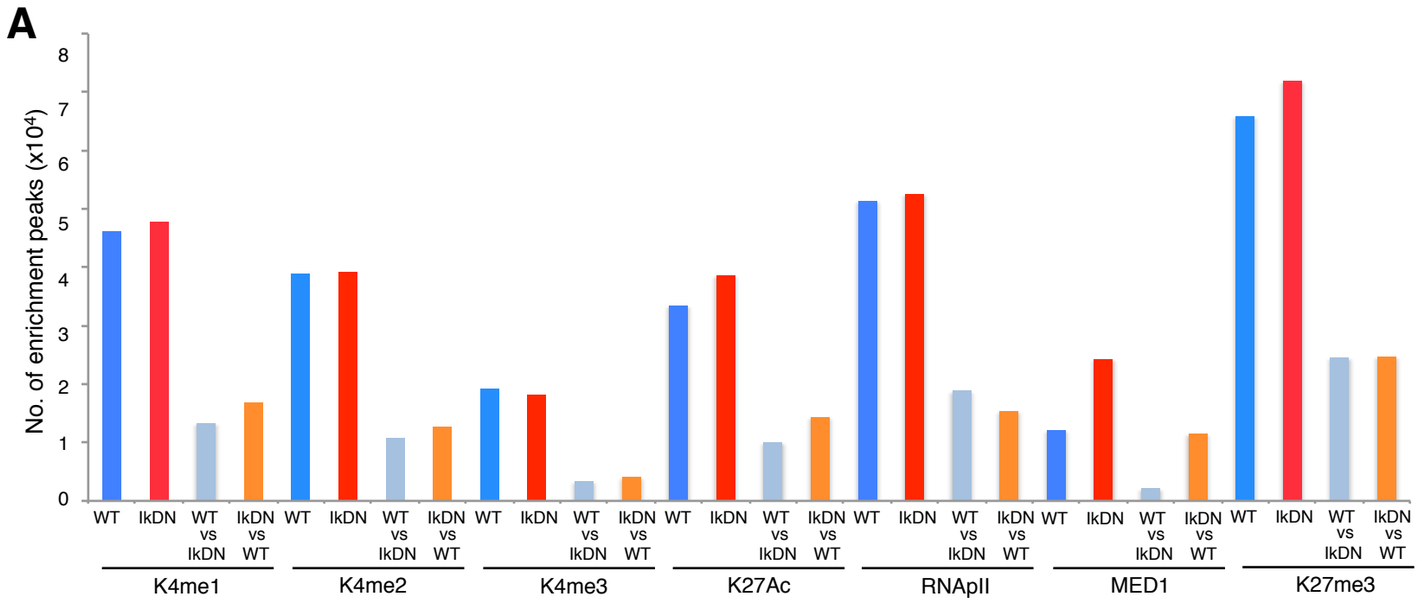
Supplemental Figure S6. Induction of de novo enhancers and super-enhancers in IKAROS-deficient stromal-adherent large pre-B cells, related to Figure 5. (*A*) Super-enhancers (SE) for non-lymphoid and B cell transcription factors were identified in IKDN pre-B cells using a Rank order super-enhancer algorithm as described in Figure S2. The number of super-enhancers over the total number of binding sites for each transcription factor is provided. (*B*) 312 unique up-regulated genes in IKDN pre-B cells were associated with SC defined by at least one of the six transcription factors (TEAD, EBF1, LMO2, LHX2, PAX5, IRF4) tested here. A circular histogram of the frequency of gene occupancy by SC defined by three major contributors (i.e. TEAD, EBF1, LMO2) and overlap are shown. The relative percentages are listed. (*C*) Frequency of individual transcription factor occupancy at the 312 up-regulated genes with SC through their own SC (SC) or through association with another factor's SC (BS in SC) is shown in concentric circles on the left side of this panel. Frequency of transcription factor occupancy at the remaining 1285 up-regulated genes without SC is shown on the right side. (*D*) Box plot depicting the distribution of transcription factor binding sites at

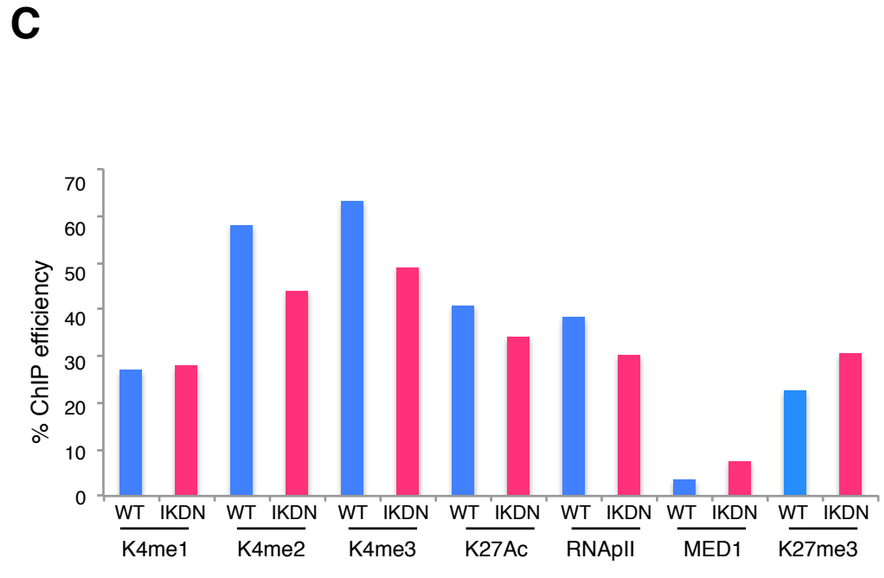
subsets of up-regulated genes with super-clusters (SC, BS in SC) or without (R-BS, as detailed in Figure 5C. Box plot whiskers extend to 1.5x the interquartile range. The average number of binding sites per gene subset is shown below the plot. P-values for differential distributions in the SC or BS in SC group relative to the R-BS group were obtained using an unpaired two-tailed *t*-test with variance determined by an F test. Statistical significance compared to the R-BS subset are shown as p-values; ****, $p<10^{-19}$; *** , $p<10^{-15}$; **, $p<10^{-11}$; *, $p<10^{-7}$. (*E*) PRC2 and transcriptional activities (read count per million mapped reads) over the gene body of 205 up-regulated genes associated with PRC2 repression and IKAROS binding in WT pre-B cells and displaying super-enhancer activity in IKDN pre-B cells (IK_ K27_SUPER). This subset of genes was further subdivided according to IKAROS location at enhancers (Enh 133) or at promoters and enhancers (Pr-Enh 66). Read density distributions for H3K27me3, H3K36me3 and IKAROS in WT and IKDN pre-B cells are plotted.

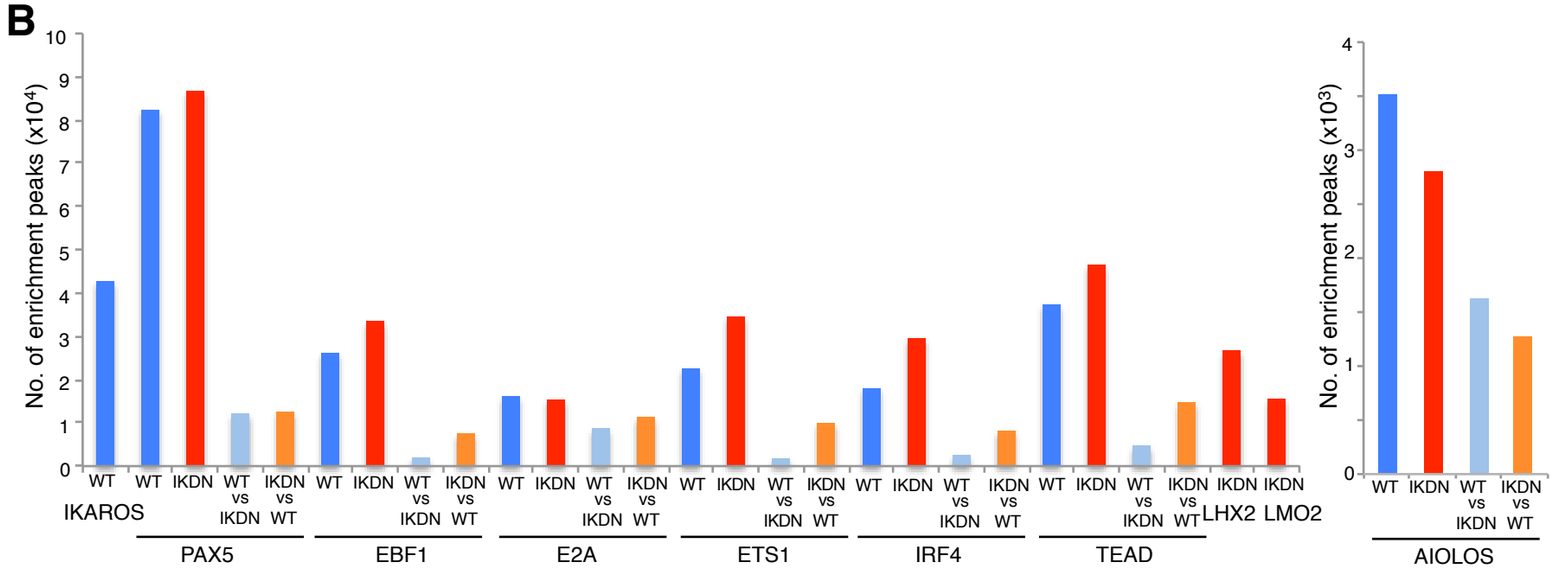
Supplemental Figure S7. Activation of de novo super-enhancers by the extra-lineage-B cell transcription factor network causes PRC2 eviction and gene induction, related to Figure 5. (*A-B*) Super-enhancers for TEAD (14.3 kb) at the *Amotl2* gene (A), for TEAD, EBF1, PAX5 and LMO2 (21.1 kb) at the *Rgs10* gene and for EBF1 (0.8 kb) at the *Plxnb1* gene (*B*) are shown. Genome browser tracks for ChIP-seq of IKAROS (WT), EBF1 (WT, IKDN), TEAD (WT, IKDN), LMO2 (IKDN), H3K4me1 (WT, IKDN), H3K27Ac (WT, IKDN) and H3K27me3 (WT, IKDN). Super-enhancer regions are marked with blue lines. Note that all three genes exhibit PRC2 repression in WT and show induction of super-enhancers in IKDN pre-B cells. Whereas *Amotl2* is not a target of Ikaros, *Rgs10* and

Plxnb1 are bound by IKAROS at poised enhancers in WT pre-B cells. These sites, normally occupied by IKAROS in WT pre-B cells, become part of *de novo* super-enhancers in IKDN pre-B cells, occupied by *extra-lineage* and B cell transcriptional regulators.

Supplemental Figure S8. (A) Genes dependent on IKAROS for expression in WT pre-B cells are not regulated by extra-lineage transcription factors.
Venn diagrams depicting overlap between genes up-regulated (UP) by factor-specific knockdowns (KD) in IKDN pre-B cells, genes associated with *de novo* sites for these factors in IKDN pre-B cells (IKDN BS), all genes down-regulated in IKDN pre-B cells (Down IKDN) and the subset of down-regulated genes associated with super-enhancers (Down IKDN Super). Limited overlap is seen between genes dependent on EBF1, LMO2, LHX2 and YAP1 for expression in IKDN pre-B cells and genes dependent on Ikaros expression in WT pre-B cells. (B) Effect of inactivation of transcription factors on the cell cycle properties of IKDN pre-B cells. Histograms depicting frequency of cells in S-G2-M after 4 days of infection by lentiviral vectors expressing factor-specific shRNAs or control is shown. The combined cell cycle data from two independent shRNA vector infections performed 1-3 times each is shown. P-value are shown for shLhx2=0.0011 (**) and for shEBF1=0.0254 (*).


A
No. of enrichment peaks (x10⁴)
0
1
2
3
4
5
6
7
8
WT
IkDN
WT vs IkDN
IkDN vs WT
K4me1
K4me2
K4me3
K27Ac
RNApII
MED1
K27me3


C
% ChIP efficiency
0
10
20
30
40
50
60
70
WT
IKDN
K4me1
K4me2
K4me3
K27Ac
RNApII
MED1
K27me3


B
No. of enrichment peaks (x10⁴)
0
1
2
3
4
5
6
7
8
9
10
WT
IKDN
WT vs IKDN
IKDN vs WT
IKAROS
PAX5
EBF1
E2A
ETS1
IRF4
TEAD
LHX2
LMO2

No. of enrichment peaks (x10³)
0
1
2
3
4
WT
IKDN
WT vs IKDN
IKDN vs WT
AIOLOS

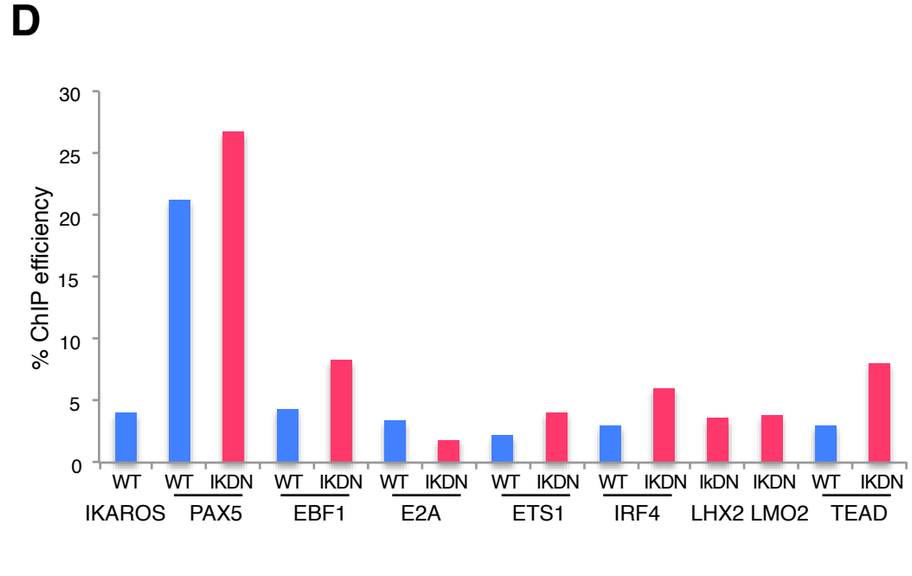

D
% ChIP efficiency
0
5
10
15
20
25
30
WT
IKDN
IKAROS
PAX5
EBF1
E2A
ETS1
IRF4
LHX2
LMO2
TEAD

Figure S1

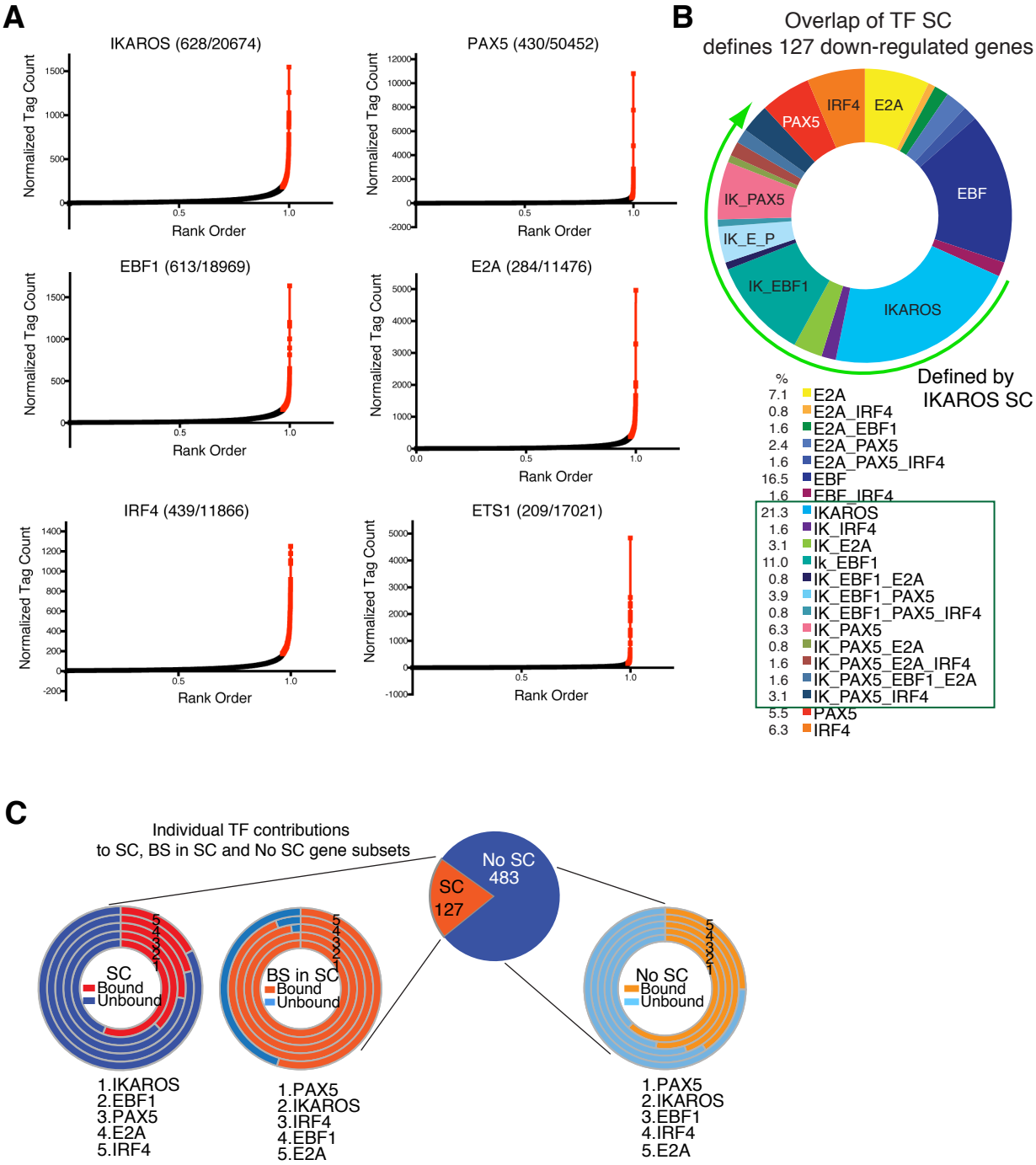

A
IKAROS (628/20674)
PAX5 (430/50452)
EBF1 (613/18969)
E2A (284/11476)
IRF4 (439/11866)
ETS1 (209/17021)
Normalized Tag Count
Rank Order
B
Overlap of TF SC
defines 127 down-regulated genes
IRF4
E2A
PAX5
EBF
IK_PAX5
IK_E_P
IK_EBF1
IKAROS
Defined by
IKAROS SC
% E2A
0.8 E2A_IRF4
1.6 E2A_EBF1
2.4 E2A_PAX5
1.6 E2A_PAX5_IRF4
16.5 EBF
1.6 EBF_IRF4
21.3 IKAROS
1.6 IK_IRF4
3.1 IK_E2A
11.0 Ik_EBF1
0.8 IK_EBF1_E2A
3.9 IK_EBF1_PAX5
0.8 IK_EBF1_PAX5_IRF4
6.3 IK_PAX5
0.8 IK_PAX5_E2A
1.6 IK_PAX5_E2A_IRF4
1.6 IK_PAX5_EBF1_E2A
3.1 IK_PAX5_IRF4
5.5 PAX5
6.3 IRF4
C
Individual TF contributions
to SC, BS in SC and No SC gene subsets
No SC
483
SC
127
SC
Bound
Unbound
BS in SC
Bound
Unbound
No SC
Bound
Unbound
1.IKAROS
2.EBF1
3.PAX5
4.E2A
5.IRF4
1.PAX5
2.IKAROS
3.IRF4
4.EBF1
5.E2A
1.PAX5
2.IKAROS
3.EBF1
4.IRF4
5.E2A

Figure S2

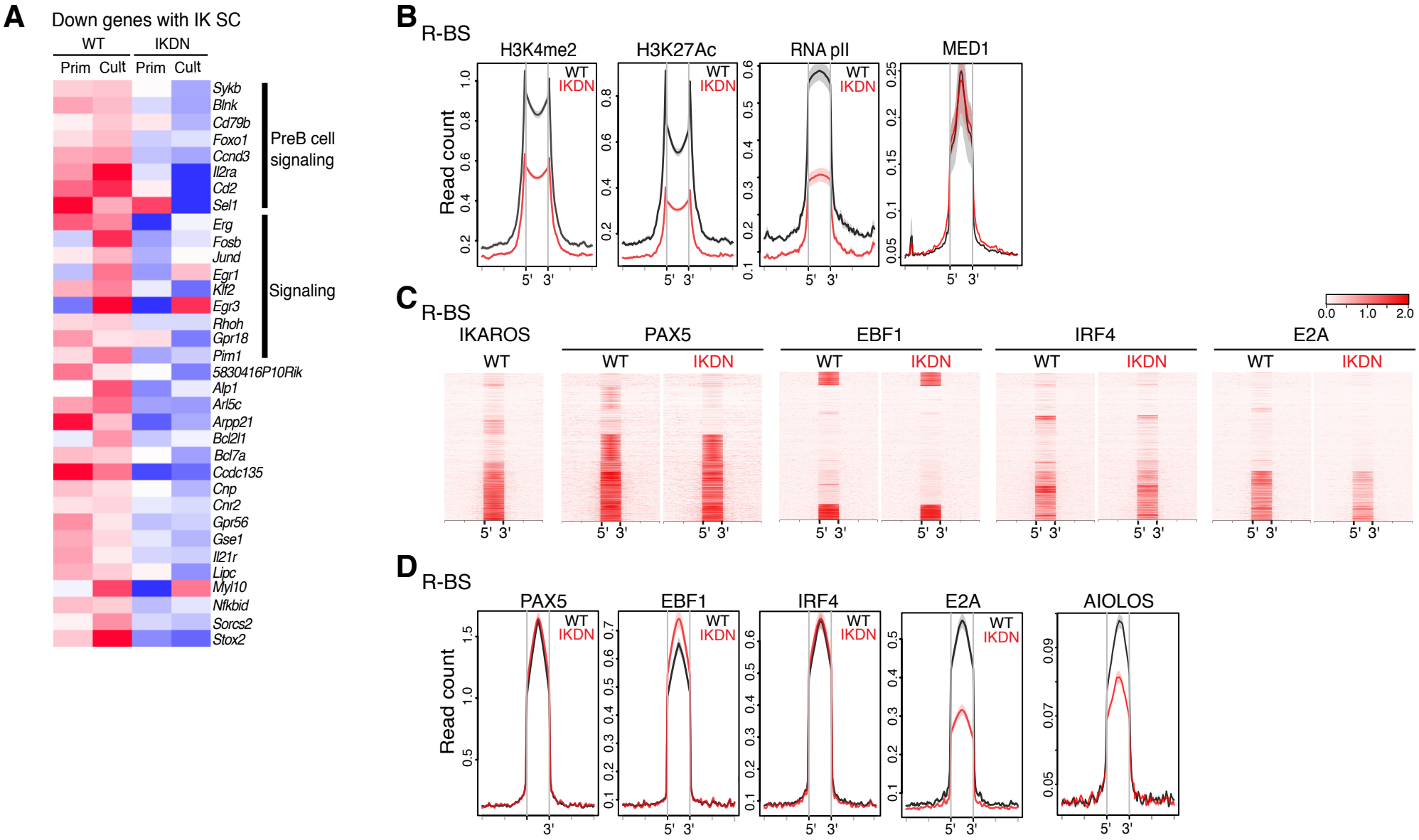

A
Down genes with IK SC
WT
IKDN
Prim
Cult
Prim
Cult
Sykb
Blnk
Cd79b
Foxo1
Ccnd3
Il2ra
Cd2
Sel1
Erg
Fosb
Jund
Egr1
Klf2
Egr3
Rhoh
Gpr18
Pim1
5830416P10Rik
Alp1
Arl5c
Arpp21
Bcl2l1
Bcl7a
Ccdc135
Cnp
Cnr2
Gpr56
Gse1
Il21r
Lipc
Myl10
Nfkbid
Sorcs2
Stox2
PreB cell signaling
Signaling
B
R-BS
H3K4me2
H3K27Ac
RNA pII
MED1
Read count
WT
IKDN
5'
3'
C
R-BS
IKAROS
PAX5
EBF1
IRF4
E2A
WT
IKDN
0.0
1.0
2.0
D
R-BS
PAX5
EBF1
IRF4
E2A
AIOLOS
Read count
WT
IKDN

Figure S3

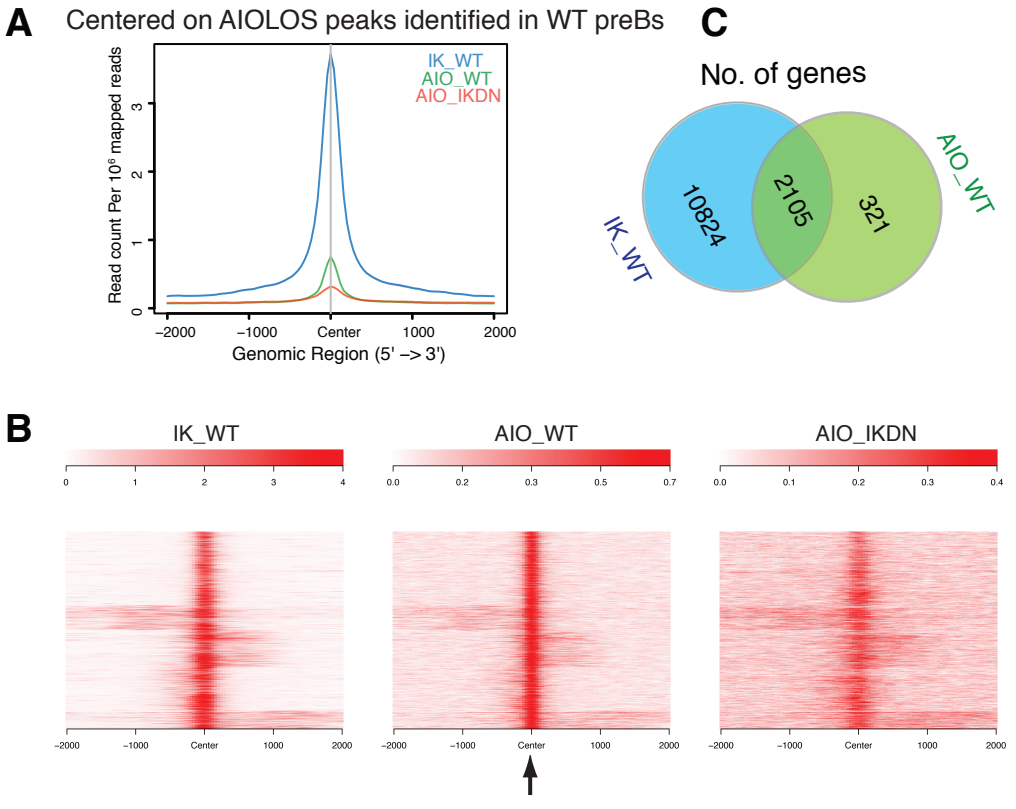

A
Centered on AIOLOS peaks identified in WT preBs
IK_WT
AIO_WT
AIO_IKDN
Read count Per 10^6 mapped reads
0
1
2
3
−2000
−1000
Center
1000
2000
Genomic Region (5' –> 3')
C
No. of genes
IK_WT
AIO_WT
10824
2105
321
B
IK_WT
AIO_WT
AIO_IKDN
0
1
2
3
4
0.0
0.2
0.3
0.5
0.7
0.0
0.1
0.2
0.3
0.4
−2000
−1000
Center
1000
2000

Figure S4

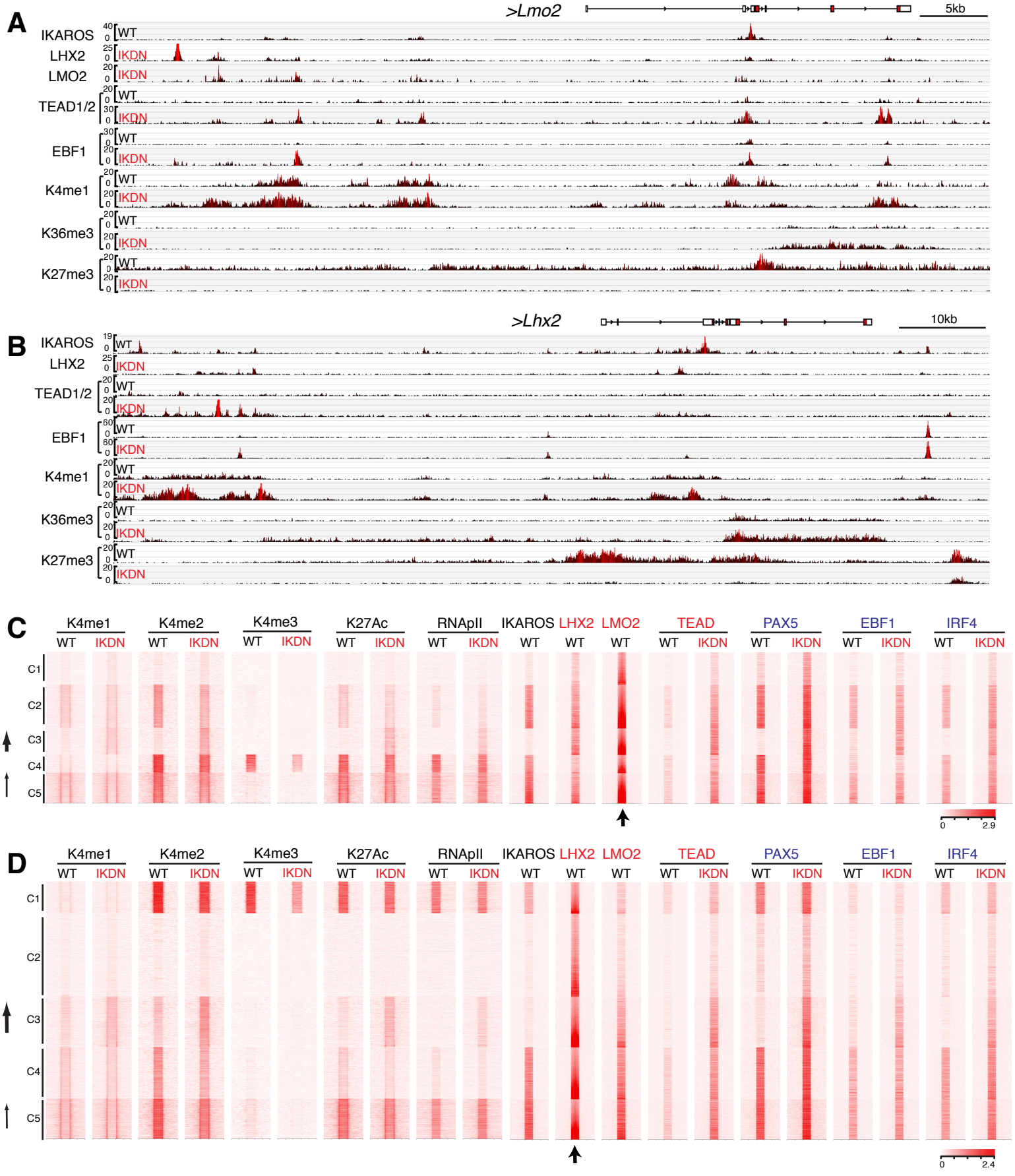

A
>Lmo2
5kb
IKAROS
LHX2
LMO2
TEAD1/2
EBF1
K4me1
K36me3
K27me3
WT
IKDN
B
>Lhx2
10kb
C
D
K4me1
K4me2
K4me3
K27Ac
RNApII
IKAROS
LHX2
LMO2
TEAD
PAX5
EBF1
IRF4
C1
C2
C3
C4
C5
2.9
2.4

Figure S5

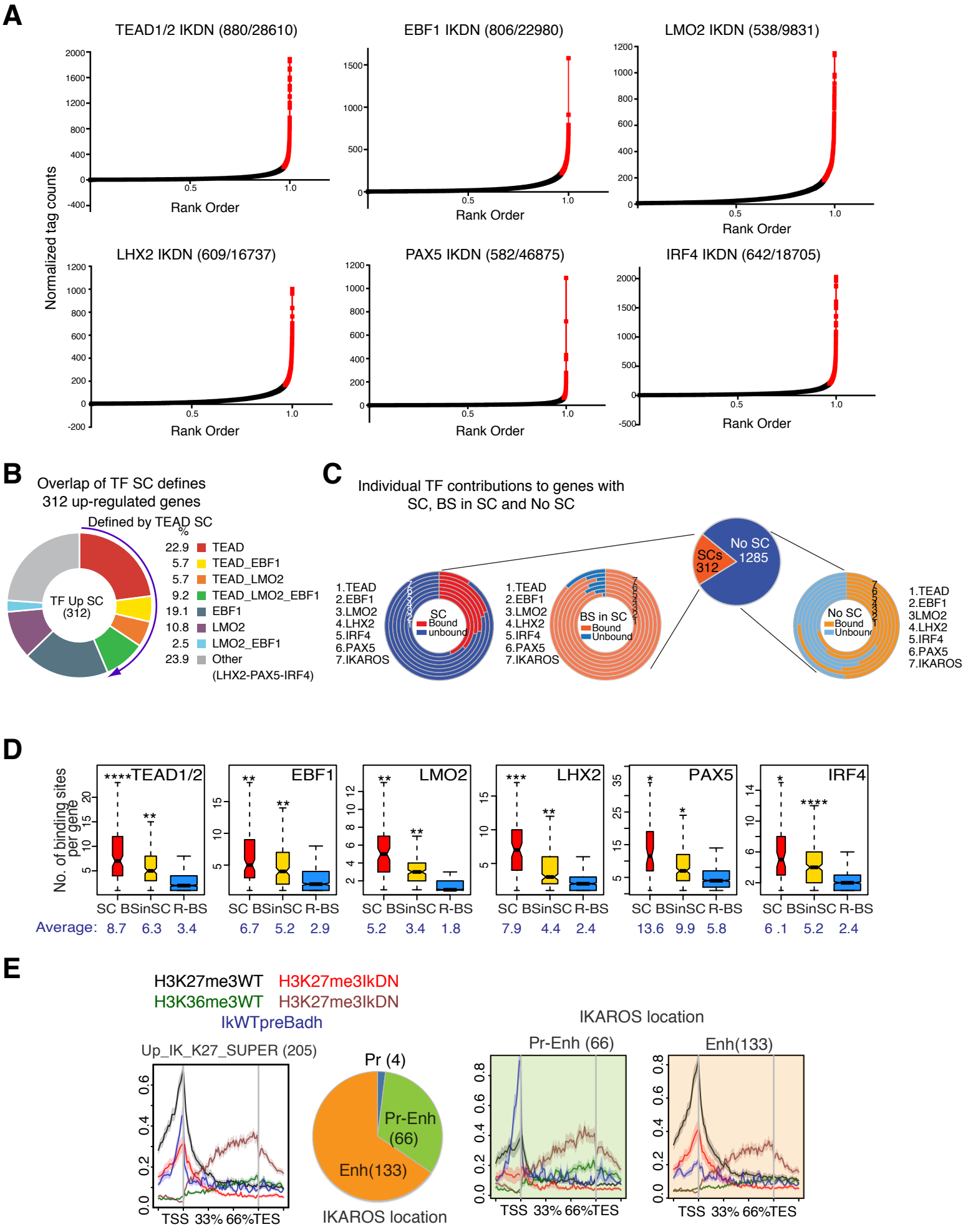

A
TEAD1/2 IKDN (880/28610)
EBF1 IKDN (806/22980)
LMO2 IKDN (538/9831)
LHX2 IKDN (609/16737)
PAX5 IKDN (582/46875)
IRF4 IKDN (642/18705)
Normalized tag counts
Rank Order
B
Overlap of TF SC defines
312 up-regulated genes
Defined by TEAD SC
TF Up SC
(312)
% 22.9 TEAD
5.7 TEAD_EBF1
5.7 TEAD_LMO2
9.2 TEAD_LMO2_EBF1
19.1 EBF1
10.8 LMO2
2.5 LMO2_EBF1
23.9 Other
(LHX2-PAX5-IRF4)
C
Individual TF contributions to genes with
SC, BS in SC and No SC
SCs
312
No SC
1285
1.TEAD
2.EBF1
3.LMO2
4.LHX2
5.IRF4
6.PAX5
7.IKAROS
SC
Bound
unbound
BS in SC
Bound
Unbound
No SC
Bound
Unboun
D
No. of binding sites
per gene
TEAD1/2
EBF1
LMO2
LHX2
PAX5
IRF4
SC BSinSC R-BS
Average: 8.7 6.3 3.4
6.7 5.2 2.9
5.2 3.4 1.8
7.9 4.4 2.4
13.6 9.9 5.8
6 .1 5.2 2.4
E
H3K27me3WT H3K27me3IkDN
H3K36me3WT H3K27me3IkDN
IkWTpreBadh
Up_IK_K27_SUPER (205)
TSS 33% 66%TES
Pr (4)
Pr-Enh
(66)
Enh(133)
IKAROS location
IKAROS location
Pr-Enh (66)
Enh(133)
TSS 33% 66%TES
TSS 33% 66%TES

Figure S6

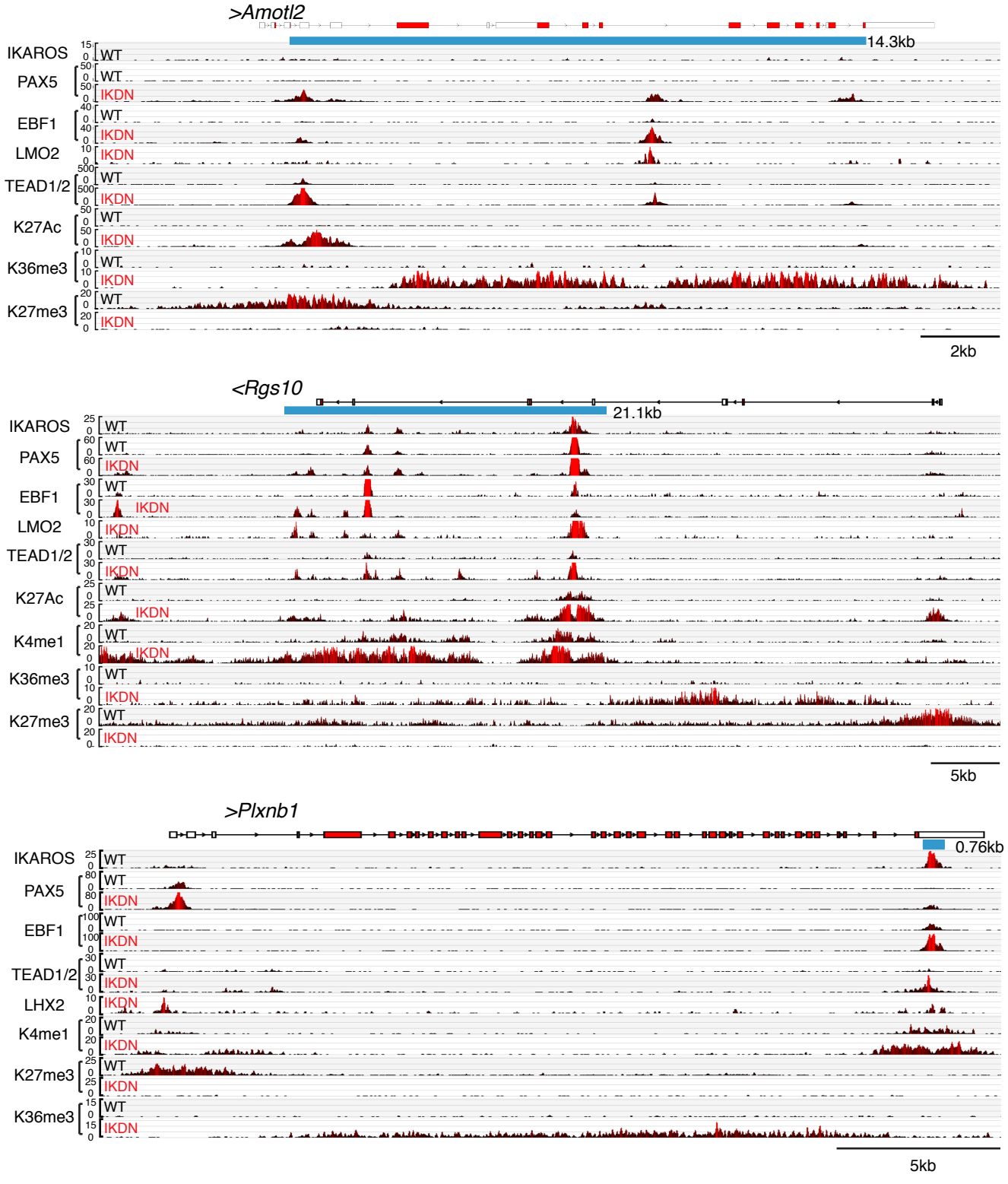
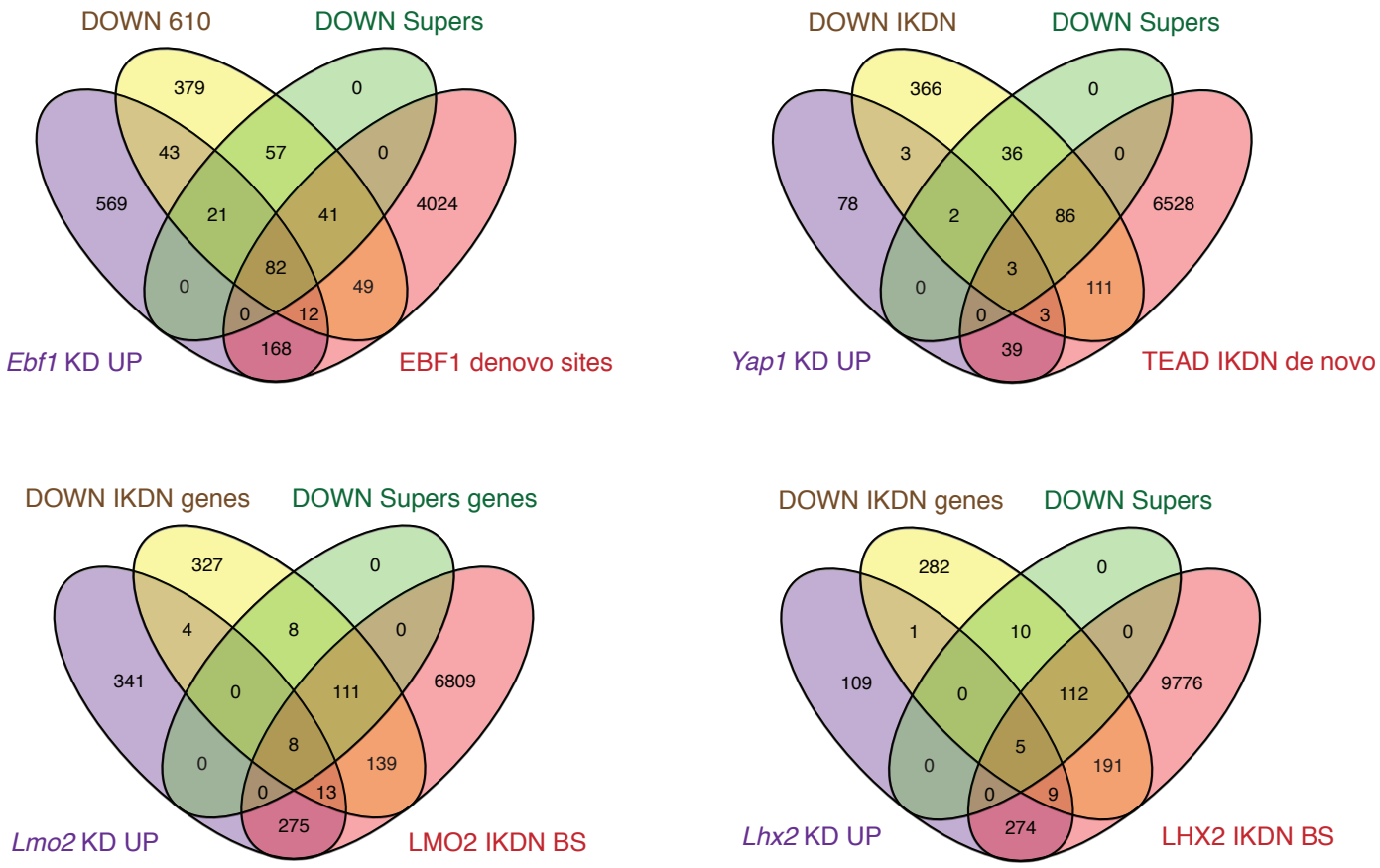

>Amotl2
14.3kb
IKAROS
PAX5
EBF1
LMO2
TEAD1/2
K27Ac
K36me3
K27me3
WT
IKDN
2kb
<Rgs10
21.1kb
IKAROS
PAX5
EBF1
LMO2
TEAD1/2
K27Ac
K4me1
K36me3
K27me3
WT
IKDN
5kb
>Plxnb1
0.76kb
IKAROS
PAX5
EBF1
TEAD1/2
LHX2
K4me1
K27me3
K36me3
WT
IKDN
5kb

Figure S7

A


DOWN 610
DOWN Supers
379
0
43
57
0
569
21
41
4024
82
0
49
0
12
168
Ebf1 KD UP
EBF1 denovo sites
DOWN IKDN
DOWN Supers
366
0
3
36
0
78
2
86
6528
3
0
111
0
3
39
Yap1 KD UP
TEAD IKDN de novo
DOWN IKDN genes
DOWN Supers genes
327
0
4
8
0
341
0
111
6809
8
0
139
0
13
275
Lmo2 KD UP
LMO2 IKDN BS
DOWN IKDN genes
DOWN Supers
282
0
1
10
0
109
0
112
9776
5
0
191
0
9
274
Lhx2 KD UP
LHX2 IKDN BS

B

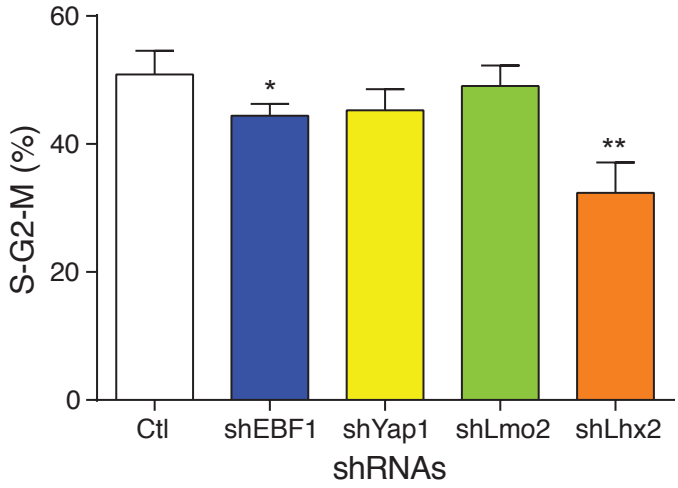

60
40
20
0
S-G2-M (%)
*
**
Ctl
shEBF1
shYap1
shLmo2
shLhx2
shRNAs

Figure S8