SUPPLEMENTARY MATERIAL

W. J. R. Longabaugh, W. Zeng, J. A. Zhang, H. Hosokawa, C. Jansen, L. Li, M. Romero-Wolf, H. Y. Kueh, P. Liu, A. Mortazavi, and E. V. Rothenberg

1. Supplementary Methods: pp. 2-11
2. Supplementary Text: Detailed review of Gene Regulatory Network model: pp. 12-21
3. Supplementary References: pp. 22-31

4. Supplementary Table Legends (Tables S1-S5) pp. 32-34
5. Supplementary Figure Legends (Figs. S1-S12) pp. 35-44

6. Supplementary Figures S1-S12

1. SUPPLEMENTARY METHODS

Mice and cells

Bcl11b conditional knockout samples were obtained from two different sources: from *Bcl11b*$^{L2/L2}$ homozygously floxed stock repeatedly crossed to a C57BL6/J (B6) background, obtained originally from Mark Leid (1, 2), and from PLBD homozygously floxed *Bcl11b*$^{fl/fl}$ mice with a Cre-ERT2 transgene ubiquitously expressed from the *ROSA26* locus, as previously described (3).The Bcl11b-conditional knockout mice, originally (129xB6)F_1 and then crossed to B6, were mostly but not completely B6 congenic at the time of use. Some of the animals were also crossed with (B6,129)-*ROSA26R-YFP* mice, to introduce a positive fluorescent marker for successful activation of Cre, as already reported elsewhere (2). Animals were bred and maintained under specific pathogen-free conditions in our colony at the California Institute of Technology under protocols approved by the Institutional Animal Care and Use Committee, with regular monitoring for mouse pathogens.

Cell culture conditions

To generate protocol I samples, timed pregnancies were set up and cells were harvested and processed essentially as previously described (1). Typically, E13.5 fetal liver (FL) cell samples from Bcl11b$^{fl/fl}$-ROSA26R-YFP and control C57BL/6-ROSA26R-YFP mouse embryos were first depleted of Ter119$^+$, CD19$^+$ and Gr1$^+$ cells using streptavidin-coated magnetic beads (Miltenyi Biotec), then sorted to purify Kit+CD27+ multi-lineage precursors by fluorescence-activated cell sorting (FACS; using FACSAria IIu). The sorted cells were then retrovirally transduced with a Cre retrovirus and incubated in culture with OP9-DL1 stromal cells in the presence of 5 ng/mL each of Flt3L, IL-7 and SCF for 24 hours. Successfully transduced cells were re-sorted for YFP expression (or for a fluorescent protein in the retroviral vector if *ROSA26R-YFP* was not present in the background genotype) as well as progenitor-enriched phenotype (Kit+CD27+), and

returned to the OP9-DL1 or OP9-DL4 stromal co-culture. After 8.5-14 days of culture as indicated, the cells were harvested and sorted for YFP$^+$ DN2a phenotype (YFP+ Kit^{++}CD45$^+$CD25+Lin$^-$) from *ROSA26-YFP; Bcl11b$^{fl/fl}$* samples, and YFP$^+$ DN2b phenotype (YFP+c-Kit^{int} CD45$^+$CD25$^+$Lin$^-$) from C57BL/6. *ROSA26-YFP* or other Bcl11b-intact controls. Here, "Lin" was detected by Lin cocktail = biotinylated antibodies to Ter-119, CD19, F4/80, Gr-1, NK1.1, CD122, CD11c, TCRγδ, TCRβ, CD3ε, CD8α and by PerCP-Cy5.5-coupled streptavidin (4). Typical profiles of the YFP$^+$ cells generated from Bcl11b-floxed and control cultures are shown in Fig. S1A.

Some variants of these conditions were used. We noted that in some batches of medium, the controls took longer than normal to progress beyond early DN2b. The cells moved more rapidly to DN3 if the IL-7 level were reduced on re-feeding the cultures. Therefore in some of the earlier experiments we shifted down the IL-7 concentration by omitting it from the medium at the first cell feeding, after about 4 days. However, the quality of the RNA obtained in these conditions was more variable and the survival of control cells was lower than in higher IL-7, so later samples all used 5 ng/ml IL-7 throughout the culture period and compensated by using longer culture times (10-12 days rather than 8-9 days).

In another variation, to test the possibility that Bcl11b deletion mimicked Notch inhibition, we incubated normal B6 fetal liver derived precursors to generate DN2a or DN2b cells as previously described (4), but with addition of 1 μM γ-secretase inhibitor [GSI, InSolution™ γ-Secretase Inhibitor X (EMD Millipore)] for the last 12 hr before harvest. This timepoint was chosen in order to allow inhibition to have an effect but not yet to downregulate cell-surface CD25 expression enough to interfere with subset purification [see (5)]. Two samples each were collected of DN2a cells with and without GSI treatment, and of DN2b cells with and without GSI treatment. However, at this timepoint gene expression changes were minimal (see heat map of effects on SOM metaclusters in Fig. S12, GSI column).

Bcl11b deletion after commitment

For protocol II comparisons between Bcl11b knockout and wildtype samples, we obtained cells that had lost Bcl11b after commitment either in vivo or in vitro. For in vivo deletion, fresh thymocytes were obtained from *Lck-Cre;ROSA26R-eYFP; Bcl11b*$^{fl/fl}$ mice and *Lck-Cre;ROSA26R-eYFP* controls. YFP^{+} thymocytes were primarily DN3 and later in the *Lck-Cre*$^{+}$ *Bcl11b*+ controls, but the *Lck-Cre*$^{+}$ *Bcl11b*$^{fl/fl}$ animals accumulated a substantial population with an altered phenotype, DN2-like with respect to Kit and CD25 (Kit^{++} CD44^{low} CD25^{+}), that represent arrested cells with Bcl11b deleted. This aberrant YFP^{+} population in the mutant and the DN3 YFP^{+} population from the controls were sorted for RNA analysis as described below. For in vitro deletion, we used cells from PLBD mice in which Cre-ERT2 ubiquitously expressed from the *ROSA26* locus could be activated experimentally by treatment with 4-OH Tamoxifen (4-OHT). Normal, unperturbed DN3 thymocytes (Kit^{low}CD44^{low} CD25^{+}) were sorted from PLBD mice and from control mice in which the *ROSA26-Cre-ERT2* gene had been crossed out repeatedly onto a B6 background ("ERT2Cre" mice). The cells were then treated with 100 nM 4-OHT for four days in co-culture with OP9-DL4 stromal cells. As shown in Fig. S1B, the controls progressed from DN3 to DN4 while the mutant cells underwent retrograde-appearing differentiation to a DN2-like phenotype (repeated in 4 independent experiments). These cells were purified from the stroma by sorting as before and used to generate RNA samples for RNA-seq as previously described (4).

RNA-seq data analysis to define "gold standard" differentially expressed genes

RNA-seq data were initially aligned by Bowtie and eXpress using Gencode M1 transcriptome models. Bigwig tracks were generated for individual inspection of RNA-seq sample quality based on relatively consistent representation of reads across 5′ and 3′ exons from genes across the genome. To generate a high-confidence set of differentially expressed genes, higher quality pairs of Bcl11b knockout and control samples were selected and the reads were pooled into three control and matched knockout sample

groups for analysis of differentially expressed genes, using EdgeR. Data from samples shown in Table S5 were combined as follows. Groups A and B consisted of Cre+ samples generated through a "protocol I" strategy, starting with Cre-transduced fetal liver hematopoietic precursors and harvesting at 8.5-12 days of culture. Group A contained Bcl11b KO samples 11970, 12497 and 12501 and the respective, paired Cre-treated controls 11969, 12499, and 12502. Group B contained Bcl11b KO samples 11785, 13102, and 13149 and their paired Cre-treated controls 11787, 13100, and 13148. Group C contained representative samples from two "protocol II" comparisons using postnatal thymocytes acutely exposed to Cre to delete Bcl11b after commitment. One in vivo comparison pair used Lck-Cre to delete in vivo and consisted of Bcl11b KO sample 12217 and Lck-Cre control 12886. The other close sample pair in group C consisted of PLBD (*Bcl11b*$^{fl/fl}$; *ROSA26-CreERT2*) DN3 cells freshly sorted as controls (13218) and after 4 days of tamoxifen treatment to delete *Bcl11b* and OP9-DL4 coculture (13219). Not included in the gold standard sets but also analyzed to validate the wildtype initial phenotype of the PLBD cells were supporting controls from *ROSA26-CreERT2* controls (ERT2Cre) both freshly sorted as DN3 and after parallel tamoxifen treatment and culture (13220 and 13221, respectively). EdgeR analysis was used to identify genes differentially expressed between knockouts and controls with FDR <0.05 in each group. As reported in Table S1, the numbers of differentially expressed genes identified were different in the three groups and likely underestimated the number of truly regulated genes. Venn diagrams were used to identify transcripts that scored as significant (FDR <0.05) in at least two of the three groups as shown in Table S1.

Multiple additional samples made by similar or matched protocols were also used for RNA-seq analysis but either lacked a sufficient quality control sample or had somewhat lower quality read distributions or lower deletion efficiency for *Bcl11b* in the mutants. These samples were not used to define the gold standard differentially expressed genes but were included in the larger gene list to construct the SOM.

RNA-seq data alignment and gene expression quantitation for SOM

RNA-seq read alignment and gene expression quantification were done using Rsem (v1.2.12) (6). The Rsem transcriptome reference file was generated with the mm9 reference genome and mouse GENCODE M1 gene annotation file. For those samples expressing obligate repressor PU.1-Engrailed (PU.1-ENG), the PU.1-ENG sequence and gene annotation information were included in the reference file to accurately quantify native endogenous PU.1 levels(*Spi1*) (7). Following gene expression quantification, PU.1-ENG levels were ignored. Gene level FPKM values were merged into data matrix for further processing. Small RNAs such as rRNA, tRNA, snRNA, snoRNA as well as miRNA were removed from the matrix and genes with < 1 FPKM in all the samples were also filtered out. The data matrix was further batch corrected with ComBat (8).

Self-organizing map (SOM)

A pseudocount of 1.0 was added to the input data matrix after ComBat adjustment to enable logarithmic transformation. The $\log_2$-transformed matrix used as input to build a SOM with a toroid topology with size 40x60 with 12 million time steps using the SOMatic package (v.1.0) (this package can be downloaded from <https://github.com/csjansen/SOMatic>); the best of three tries based on fitting error was selected. The SOMatic website for this mapping can be viewed at <http://crick.bio.uci.edu/Bcl11b_SOM/>.

SOMs of this size can be further clustered into supersets of similar units called metaclusters (9) by first performing a k-means clustering on the vector values of the units to determine centroids. Metaclusters were built around these centroids so that each cluster is contiguous on the SOM topology. After trying a range of different K values, a metaclustering with K=300 was chosen. R scripts for generating metacluster reports are provided in the SOMatic package.

Metacluster/Trait correlation and hypothesis testing analysis

To determine which metaclusters captured changes in gene expression due to experiments with similar conditions, we assigned each dataset used to train the SOM with a number of traits related to different developmental stages or perturbation conditions. A correlation matrix is then built using the Pearson correlation between the columns of this trait matrix and the columns of a matrix containing the average signal in each metacluster in each experiment map. A p-value can then be calculated for each of these correlation coefficients. These p-values are then corrected for Type 1 error using the Bonferroni method and reported in a trait heatmap.

To build the hypothesis maps, we used a similar process where we did a Wilcoxon test to see if the difference between two sets of samples is significant compared to the differences within each set. Those metaclusters that show a significant increase between the two sets of samples are displayed in red, and those that are significantly depleted are shown in blue.
R scripts for these processes are provided in the SOMatic package.

Additional datasets used for comparison

We obtained RNA-seq data for wildtype cells developing in vivo or in vitro in essentially the same OP9-Delta culture conditions as our experimental samples, taken from ref. (4). RNA-seq data from DN2a-like cells developing in vitro after less than 24 hr of transduction with empty vector or PU.1-engrailed obligate repressor were obtained from ref. (7)(Table S3). Microarray data were previously reported for Bcl11b knockout and wildtype DP cells various lengths of time after Bcl11b deletion in vitro or in vivo, from ref. (3) and ref. (10), and we extracted lists containing the genes showing at least twofold increases or decreases in expression from the supplementary tables in these papers (see Table S1E,F). Microarray data from ref. (11) were a source of lists of genes with fold increase or decrease ≥1.5 between E2A knockout and wildtype DN2 cells, for the analyses described for Fig. 3 and Table S4. Lists of transcription factors,

transcriptional cofactors, and chromatin modifying factors were obtained from AnimalTFDB (12).

Chromatin Immunoprecipitation (ChIP)

Ten million BM-derived DN3 cells, generated by culture on OP9-DL1, were fixed with 1 mg/ml DSG (Thermo Scientific) in PBS for 30 min at RT followed by additional 10 min with addition of formaldehyde up to 1%. The reaction was quenched by addition of 1/10 volume of 0.125M glycine and the cells were washed with HBSS (Gibco). Pelleted nuclei were dissolved in lysis buffer (0.5% SDS, 10 mM EDTA, 0.5 mM EGTA, 50 mM Tris-HCl (pH 8) and PIC) and sonicated on a Bioruptor (Diagenode) for 18 cycles of 30sec sonication followed by 30sec rest, with max power. Six µg per 10^7 cells of anti-Bcl11b Abs [a mixture of A300-383A (Bethyl), A300-385A (Bethyl), ab18465 (Abcam) and 12120 (Cell Signaling Technologies)] were adsorbed to Dynabeads Protein A/G (Invitrogen) and added to the diluted chromatin complex. They were incubated over night at 4°C, then washed and eluted for 6hr at 65°C in ChIP elution buffer (20 mM Tris-HCl, pH 7.5, 5 mM EDTA 50 mM NaCl, 1% SDS, and 50 µg proteinase K). Precipitated chromatin fragments were cleaned up using Zymo ChIP DNA Clean & Concentrator and subjected to quantitative PCR on a 7900HT Fast Real-Time PCR System (Applied Biosystems) with SYBR® GreenER™ qPCR SuperMix (Invitrogen).

Primers used for ChIP-PCR:

Zbtb16 intron 2

FW; 5′-ggtgctgcaggagtaagca-3′

RV; 5′-ttgacaggagcaaaccagatt-3′

Tnni1 intron 6

FW; 5′-aaaatgggggtcccagtatg-3′

RV; 5′-ccacaatgaagggctgagtc-3′

Dntt intron 3

FW; 5′-cttttcagttcagtgacagagtcc-3′

RV; 5′-gcctttggagtcaggatcaa-3′

Cd3d exon5

FW; 5′-ttgtcttagcttggctgagttct-3′

RV; 5′-gaagctgtccctcgtactgc-3′

Igκ 3′UTR

FW; 5′-ggcacatctgttgctttcgc-3′

RV; 5′-ggggtagggagcaggtgtat-3′

Gapdh promoter

FW; 5′-gcagcaagggttgcattt-3′

RV; 5′-ttgtctgtactgggctgtttacat-3′

Il4 promoter

FW; 5′-tgactctggggtcttacctcat-3′

RV; 5′-tctccaaagaccacaaacttgtaa-3′

Chromatin Immunoprecipitation (ChIP) and ChIP-sequencing

ChIP-seq libraries were constructed using NEBNext ChIP-Seq Library Preparation Kits (NEB #E6240) following the manufacturer's instructions. Libraries were sequenced on an Illumina HiSeq2500 in single read mode with the read length of 50 nt following manufacturer's instructions. Base calls were performed with RTA 1.13.48.0 followed by conversion to FASTQ with bcl2fastq 1.8.4, and produced approximately 30 million reads per sample. ChIP-seq data were mapped to the mouse genome build NCBI37/mm9 using Bowtie (v1.1.1; http://bowtie-bio.sourceforge.net/index.shtml) with "-v 3 -k 11 -m 10 -t --best –strata" settings, and HOMER tagdirectories were created with *makeTagDirectory*. Bcl11b peaks were identified with *findPeaks.pl* against a matched control sample using the settings "*-P .1 -LP .1 -poisson .1 -style factor*". Bcl11b peak reproducibility was determined by a HOMER adaptation of the IDR (Irreproducibility Discovery Rate) package according to ENCODE guidelines (https://sites.google.com/site/anshulkundaje/projects/idr). Only reproducible high quality peaks, with a normalized peak score ≥ 15, were considered for further analysis.

Motif enrichment analysis was performed with the *findMotifsGenome.pl* command in the HOMER package using a 200bp window. Consistent with the varied roles of Bcl11b both in direct DNA binding and in binding via other transcription factors (13), Bcl11b ChIP-seq peaks were not dominated by a single specific motif. However, ETS-family and Runx-family motifs were by far the most highly enriched at the ~20,000 high-quality Bcl11b occupancy sites (>40% of sites, and >25% of sites, respectively; both P values <1E-1500; see Fig. S8H).

Quantitative real-time RT-PCR (qPCR)

Two to three PLBD, ERT2Cre, or C57BL/6J 6-12 wk old age-matched mice were pooled for each replicate as indicated. To isolate DN populations, mice were sacrificed, their thymuses were removed, and single-cell suspensions were made. Mature cells were depleted by staining with biotinylated Abs to CD8a (53-6.7), TCRγδ (GL3), TCRβ (H57597), Ter119 (Ter119), NK1.1 (PK136), CD11c (N418), and CD11b (M1/70), after which the cells were incubated with streptavidin-coated magnetic beads and then passed through a magnetic column (Miltenyi Biotec, Auburn, CA). Eluted DN cells were stained with CD117 (cKit)-PE (2B8) CD25-APCe780 (PC61.5). Kit-low, CD25+ DN3 cells were purified by FACS and cultured on OP9 DL1 monolayers for 4 days. 1.0×10^5 PLBD DN3 cells were frozen as a pellet in the -80 as a pre-treatment, "day 0" control. After sorting, cells were treated with 100nM 4-OHT for 44-48 h (as in "Bcl11b deletion after commitment", above). Sigma 4-OHT (H7904) was reconstituted in EtOH to 1mM and further diluted in OP9 media for treating cells. On day 4 post-4-OHT-treatment, CD25+ cells from PLBD and controls were purified by FACS, noting that while the controls further downregulated Kit and most controls also downregulated CD25 expression (transitioning through β-selection), the induced Bcl11b-knockout cells generally increased or at least maintained cell-surface Kit expression. cDNA was prepared from total RNA using RNeasy extraction kits (Qiagen) and reverse transcribed using random hexamers and SuperscriptIII (Invitrogen). Specific gene expression in

cDNA samples was measured by qPCR (ABI Prism 7900HT) using SyberGreenER mix (Invitrogen). Results were calculated (Δ Ct method) and normalized to *Actinb* levels. Primers used for qPCR were described previously [Actb exons 1-2; Kit (5)], or are as follows:

Zbtb16 FW 5′- GTG CCC AGT TCT CAA AGG AG -3′

Zbtb16 RV 5′- GAA AGC GTT TCC CAC ACA G -3′

Tnni1 FW 5′- GAA TGT GGA GGC TAT GTC TGG -3′

Tnni1 RV 5′- TGT CAT ACA GCA AGC CAA CC -3′

2. SUPPLEMENTARY TEXT:

DETAILED REVIEW OF T-CELL GENE REGULATORY NETWORK MODEL

Situating Bcl11b in the larger gene network framework

The new results defining Bcl11b targets enable us to add new linkages to the T-cell specification gene regulatory network (14) relating the T-cell fate to other pathways. This update also includes recent data for gene expression changes caused by E2A deletion (11), or acute PU.1-antagonism (7) in DN2 cells developing from fetal precursors, and for the interaction of Notch signaling with PU.1 gain of function effects in DN2 cells and a DN3-like cell line (15), as well as recent dissection of the "and" logic regulators driving the activation of the *Bcl11b* gene itself in the DN2a stage(16). The new comprehensive model is shown in Fig. 1 and elements of the circuits controlled by Bcl11b in Fig. 4. The major transitions in T-cell developmental gene regulation (4) occur between DN2a and DN2b, and then again between DN3a and post-β-selection stages (DP), which we have termed the boundaries between "phase 1" and "phase 2", and between "phase 2" and "phase 3", respectively (17). In the following, "phase 1" refers to genes that are expressed in the ETP (DN1) and/or DN2a stages, whereas "phase 2" refers to genes that are expressed in stages from DN2b through DN3a. The phase 1-phase 2 boundary also coincides with lineage commitment, distinguishing the states shown in Fig. 1A and Fig. 1B, respectively.

Commitment tightly coincides with the activation of *Bcl11b* expression (16). With its relatively late onset, this is not simply a direct response to Notch signaling. To explain the timing of commitment, at least five other transcription factors have to be included (2, 16, 18-21): Runx1, E2A, Hes1, GATA-3, and TCF-1 (see Fig. 2C). Runx1 (and other Runx family members) and E2A (encoded by *Tcf3*) are already expressed at high levels by the precursors when they enter the thymus, and both play continuing roles throughout the process. E2A levels are quite constant, but its T-cell heterodimerization

partner HEB (*Tcf12*) is upregulated during the DN2a-DN3a progression, while its stem/progenitor heterodimer partner Lyl1 is downregulated (4, 5)(Fig. 2B), and this could affect some aspects of its role. Hes1 is a repressive bHLH factor that is directly activated by Notch signaling in ETPs and expressed through all the DN stages, but is expressed in early B cells as well. GATA-3 and TCF-1 (encoded by *Tcf7*) are substantially T-lineage specific. Both are upregulated in a Notch dependent way at ETP stage, and *Tcf7* at least is a direct Notch target (22, 23). Bcl11b is turned on and its role begins while PU.1 is still present, and under the aegis of all these factors' activities (16).

Also expressed through these stages and globally important for T cell development are *Gfi1* and *Ikzf1* (Ikaros), encoding two zinc finger transcription factors (24-26), and *Myb*, which is significant even into the DP stage (27, 28) and may provide some positive input to *Gata3* (15, 29). *Gfi1* is thought to be activated in an E protein-dependent way (11, 30), and may, like Bcl11b, repress Id protein expression (31, 32)(not shown in Fig. 1). *Ikzf1* is expressed in many non-T hematopoietic cells, but detailed analysis of *Ikzf1* regulatory elements suggests that Runx and TCF-1 sustain its T-cell specific expression (33), whereas the binding of E proteins also observed at this locus (33) may actually have a damping negative effect on *Ikzf1* in pro-T cells (11, 15)(Fig. 1). Ikaros has recently been shown to be needed to enforce sharp transitions in global developmental gene regulation during T-cell specification, though the mechanism is not clear (26). It has also been implicated in damping expression of some Notch response genes in later T-cell developmental stages (34, 35), but the relevance of this mechanism to the stages around commitment is not known.

Phase 1 Network: PU.1 and Notch to PU.1 shutoff

<u>*Notch1* regulation: a key role for E proteins.</u> Notch signaling in early T cells depends strongly on bHLH E protein activity, mostly E2A, which directly regulates *Notch1* expression and signaling from prethymic stages through DN3 stage (15, 36-38). Target genes of E proteins have been inferred from E2A knockout and inducible E2A addback

experiments (30, 37, 39), from E2A binding in DN3 cells in vivo (40), and from the acute effects of forced expression of the antagonist Id2 (15). As both Id2 and Id3 are normally low in phase 1 pro-T cells (Fig. 2B)(41), the only endogenous modulators of this critical input to *Notch1* during phase 1 to early phase 2 appear to be heterodimer partners of E2A, namely Lyl1 which predominates during phase 1 (42), and HEB (*Tcf12* gene product) which predominates during Phase 2 (43). Both support high *Notch1* expression even from prethymic stages, well before Bcl11b is activated.

Phase 1 regulatory gene set. The precommitment "phase 1" period is defined by strong expression of the B-cell, dendritic-cell, and myeloid-cell associated transcription factor PU.1 (encoded by *Spi1*, alias *Sfpi1*), and a set of stem-cell associated growth factor receptor and regulatory genes including *Erg, Mycn, Hhex, Bcl11a, Gfi1b, Kit,* and *Lyl1* [reviewed by (17)]. Cells within the ETP population, presumably very early phase 1 cells, also express *Flt3, Mef2c,* and *Lmo2.* This regulatory gene combination overlaps the 10-gene stem/progenitor-cell signature (44) and includes multiple T-lineage proto-oncogenes (17). Interestingly, another one of the phase 1 genes is a highly Notch-dependent gene, *Nrarp*, which acts as a negative feedback regulator of Notch signaling itself (45, 46). Silencing of all these phase 1 regulatory genes is a major landmark of commitment.

Different alternative fates are initially restricted using different parts of the T-cell network, not a single "master" switch. Hes1 antagonizes the myeloid fates via repression of *Cebpa* (21). TCF-1 and GATA-3 are antagonistic with the B cell fate (2, 47-49), but are embedded within the ILC as well as the T-cell programs [reviewed by (50, 51)]. Whereas PU.1 is an obvious bridge to the myeloid fates (15, 52-54), and its silencing may break that bridge, the fates that Bcl11b is mainly suppressing depend on factors that are initially silent in phase 1 before Bcl11b is turned on. In this light, the phenotype of Bcl11b knockouts reveals that Bcl11b repression targets gain potential

activity during commitment, not lose it, and then remain conditionally activatable, long into T cell development.

PU.1—Notch mutual antagonism. Myeloid and dendritic potential of T-cell precursors probably depends on PU.1, but loss of function experiments show that PU.1 at normal levels is needed for optimal proliferative expansion of ETP and DN2a/2b cells alike (7, 55). Although it appears to limit the pace of T-cell differentiation, PU.1 in normal thymocytes does not promote transdifferentiation to myeloid or dendritic cell fates *in vivo*. This is at least partly because strong Notch signals like those in the thymus keep PU.1 action compatible with the T cell program (15, 52, 54). There is a specific mutual antagonism between Notch signaling and PU.1, although PU.1 and Notch1 do not repress each other's expression transcriptionally. Instead, Notch signaling selectively prevents activation of the myeloid programs by PU.1. This may be because PU.1's myeloid partner *Cebpa* is kept repressed by Notch-induced Hes1 (21). However, strong Notch signaling still allows PU.1 to activate progenitor-cell associated regulatory genes including *Bcl11a,* and in some contexts *Lmo2* (7, 15). In turn, if Notch signaling is low, PU.1 can repress key T-cell regulatory genes including *Tcf7, Gata3, Zfpm1* (GATA-3 cofactor FOG1), *Hes1, Ets1,* and *Myb*, and at sufficiently high levels, PU.1 can inhibit Notch signaling itself (15)(Fig. 1A). Thus PU.1 establishes a threshold for effective Notch signaling and Notch signaling filters PU.1 impacts.

PU.1 targets: complex modes of developmental regulation. PU.1 binds many more genomic sites than genes that it regulates (4), so perturbation is essential to define its targets. Because PU.1 protein has a very long half-life (56), and because its deletion reduces cell yield, we have used an obligate repressor form of PU.1 to perturb its activity acutely and competitively, to probe PU.1-dependent transcriptional circuitry and to generate samples for global RNA-seq analysis (7). The results have yielded four take-home lessons (7). First, endogenous PU.1 not only delays expression of T-cell genes but also negatively regulates another set of alternative-lineage genes. Notably, these

turn out to be key NK and ILC genes including *Il2rb* and *Zbtb16*, which are also repressed by Bcl11b in phase 2 cells. Second, many of PU.1's repressive effects appear to be indirect. This is implied since the obligate repressor form of PU.1 does not mimic the repressive effects of wildtype PU.1, but rather enables cells to express even higher levels of its NK-lineage and T-cell associated repression targets (7). However, the mechanism of this repressive effect is still under investigation. Third, PU.1 effects on other members of its own cohort of phase 1 regulatory genes are gene-specific. While wildtype PU.1 positively regulates some phase 1 genes, like *Bcl11a, Lmo2,* and *Mef2c,* it appears to exert moderate damping repression on others like *Gfi1b*, *Kit, Mycn* and *Erg* despite their normal expression pattern strongly overlapping with PU.1 (Fig. 1A). Through its Notch-signal inhibitory effect and/or directly, elevated levels of PU.1 can also repress the Notch-dependent phase 1 gene *Nrarp*, even though it is also normally coexpressed with PU.1 in early pro-T cells. Finally, the genes most vulnerable to the PU.1 obligate repressor are enriched for cell biology and signaling genes, encoding growth factor receptors, cytoskeletal and adhesion regulators, kinases, and heterotrimeric G-protein coupled receptor signaling components, implying that these are many of the genes that endogenous PU.1 activates in unperturbed phase 1 cells (Table S3)(7).

How PU.1 limits Notch responses is not known. One known negative regulator of Notch signaling, the feedback inhibitor Nrarp, is actually vulnerable to negative regulation by PU.1 as noted above, the opposite response from what would explain PU.1's effect on Notch. Other conventional Notch inhibitors (e.g. Numb, Spen) are not found among the PU.1-dependent targets. Conceivably some PU.1 effects, such as blunting Notch signaling in early T cells, may be established by activating cross-acting signal pathway mediators or affecting cell surface dynamics so as to interfere with Notch activation biochemistry.

Transition from phase 1 to phase 2

GATA-3, dose dependence, and incoherent circuits. The early action of Notch signaling turns on both *Gata3* and *Tcf7* (encoding TCF-1) in ETP/DN1 stage, while the phase 1 regulatory factors are still prominent. As the cells begin to turn on a DN2 surface phenotype, they also strongly upregulate *Il7r*, which, at first in combination with Kit, exerts an increasingly potent stimulus to cell proliferation and survival. GATA-3 and TCF-1 then collaborate with Notch and Runx1 to bring about the transition to phase 2. GATA-3 rapidly becomes important for survival of the ETP/DN1 cells. At physiological levels, GATA-3 helps to establish TCR expression and promotes expression of *Kit, Ets1*, the gene encoding its own cofactor *Zfpm1*, and later *Bcl11b,* while restraining PU.1 and several PU.1 target genes (2, 15, 47, 57)(Fig. 1). However, GATA-3 is highly dose-dependent in its actions and must have its expression tuned by "soft" damping-type repression. GATA-3 genomic site occupancy is acutely stage-dependent even when its levels are nearly constant (4), implying that it is titrated to occupy regulatory sites effectively where, and only where, other factors or chromatin conditions assist. Even a fewfold downregulation of GATA-3 protein causes cell death (2), but overexpression in the same stages can also abort T cell development and/or promote redirection to a mast-cell fate (11, 58). The *Gata3* enhancer has sites for inputs from TCF-1, E proteins, and the Notch-interacting transcription factor RBP-J (59), but it has a variety of restraints on it. As noted above, in low Notch signaling PU.1 can reduce GATA-3 protein levels (15). Under strong Notch signaling, once *Gata3* is induced, E2A must limit *Gata3* expression to sustain T-cell development (11, 15). Furthermore, *Gata3* expression rises slightly in many Bcl11b knockout samples in this study, suggesting that Bcl11b again parallels E2A to provide restraint.

Phase 1 to Bcl11b. The onset of expression of *Bcl11b* marks the transition to phase 2 (Fig. 1B), followed by upregulation of multiple factors that reinforce the initial specification factors (see below)(5). Using a nondisruptive fluorescent reporter allele of *Bcl11b* to dissect inputs to *Bcl11b* activation at the single cell level (16), we have shown that Bcl11b activation depends on Notch signaling, GATA-3, TCF-1, and Runx1. This appears like

classic "AND" logic, useful for stringent cell type specificity. However, shRNA knockdown of candidate transcription factors at different stages before commitment, or removal of the cells from Notch signals, reveal that the critical time window for action of at least two of these regulators is confined to ETP stage, i.e. well before actual expression of Bcl11b. GATA-3, TCF-1, and Notch signaling appear to act in a hit-and-run way, controlling the probability that a given Bcl11b allele will be activated but not regulating expression amplitude by the time the gene is expressed. Surprisingly, of these factors it is only Runx1, the one input of the four that is already expressed before T-cell specification, that provides the immediate amplitude control for *Bcl11b* both in turning its expression on and in sustaining it, with or without Notch signaling (16)(Fig. 1B). The asynchronous behavior of these four combinatorial inputs strongly implies that discrete functions are needed to prepare the Bcl11b locus for activation, distinct from those needed to control transcriptional initiation once it is active. Note that in pre-commitment cells *Bcl11b* is silenced by H3K27me3 polycomb repressed marks on the chromatin and substantial CpG methylation on the DNA (60)[and data from (61)], marks which need to be removed or modified. Similar repressive mark removal occurs on a distal enhancer for *Bcl11b*, 850 kb downstream (60). The asynchrony of action among GATA-3, TCF-1, and Runx1 could reflect a division of labor between chromatin opening functions and active transcriptional initiation rate control.

<u>PU.1 silencing.</u> PU.1 silencing is important for eliminating access to dendritic-cell, myeloid, and possibly mast-cell fates given that all these alternative programs depend on it or are promoted by it (62, 63). The termination of its expression depends on many of the same inputs as Bcl11b activation: Runx family factors, probably Runx1 in this context (19, 20); GATA-3 (2, 58); and TCF-1 (18)(Fig. 1A). Bcl11b itself does not appear to be directly required, especially since PU.1 is ultimately repressed in NK cells which lack Bcl11b as well as in T cells. Runx1 is important to activate PU.1 in early hematopoietic precursors (19, 64), but in early T cells it represses PU.1 through multiple contacts at its major enhancer and a secondary silencing element (19, 20, 65). GATA-3 dosage

dynamically affects PU.1 levels by the DN2-3 stages (2, 58). Knockdown of TCF-1 in DN2 stage cells can also greatly increase diversion to dendritic cells, a PU.1-promoted fate, instead of T cells (16).

Other changes in the phase 1 to phase 2 transition

Ets1 activation and Runx paralog shift. As pro-T cells undergo commitment, there is upregulation of a variety of T-cell genes. Some are late-activated Notch target genes, whereas others encode additional members of regulatory factor families. Early T-cell development includes expression of multiple Ets family transcription factors, but DN2a to DN2b is a stage when there is substantial upregulation of the eponymous founding member of the family, Ets1 (5, 41). Ets1 is important not only for T cell development but also for NK cell and one subtype of innate lymphoid cell development as well (66-68). Ets2 also begins to be expressed at this stage. Over the next few stages, other Ets family factors including Elk3 and Etv6 as well as PU.1 are downregulated, with Erg turned off slightly later (5, 41). The increased expression of Ets1 may depend on PU.1 downregulation, since PU.1 negatively regulates Ets1, while GATA-3, which peaks in DN2a/2b stage, somewhat upregulates it (2, 15, 54)(Fig. 1B). In parallel, expression of Runx family members shifts so that Runx3 and Runx2, which are both expressed in ETP (DN1) cells, are downregulated while Runx1 is upregulated, in part activated by E proteins (11, 15). As Runx and Ets factors frequently collaborate, these coincident paralog swaps increase the possibility that different family members have different function on target genes, even if they work through the same target DNA binding sites.

Lef1 activation and control by TCF-1. TCF-1 (encoded by *Tcf7*) is important for T-cell development from an early stage and is upregulated initially in ETP (DN1) stage (22, 69). Full-length TCF-1 can act in an autoregulatory positive feedback to promote *Tcf7* expression, as well as promoting other T-cell regulatory gene expression and blocking B cell development (23). By DN2b stage, a paralog of *Tcf7, Lef1*, is also strongly activated, partly under the control of Notch signaling and partly driven by TCF-1 (Fig. 1B).

However, the most abundant isoforms of TCF-1 expressed in the thymus are not full-length, but truncated structural variants that are biased for repression rather than for activation (70). These isoforms act to set a damper on Lef1 expression levels, a function important to prevent leukemia (70, 71), and possibly also to delay timing of *Lef1* activation (Fig. 1A). In addition to participation in binding ensembles at many active enhancers (72), roles of TCF-1 have been inferred for repression of *Spi1 (Sfpi1,* PU.1) as well as for downregulation of *Lef1*, but it is not clear yet which roles of TCF-1 are unique and which can be mediated by LEF-1 as well.

Modulation of growth factor signaling by Bcl11b. Bcl11b acts not only to repress phase 1 genes and non-T cell genes, but also to modulate growth behavior of the cells. When Bcl11b is deleted, *Il2rb* and *Kit* are consistently upregulated, but *Il7r* is also consistently well expressed, albeit not as consistently above the controls (1). The combination of higher Kit and strong IL-7R expression is functionally effective to sustain proliferation in vitro if cells have lost Bcl11b before commitment, although for unknown reasons the same growth factor gene upregulation is not effective after commitment.

TCR gene activation. Expression of T-cell receptor genes depends on a complex recombination mechanism as well as transcriptional activation, so it is mostly outside of the scope of this GRN model. *Rag1* and the mutagenic polymerase *Dntt*, which are important for the recombination process, are linked into the network as targets of E proteins (11, 30, 37, 39), with GATA-3 providing additional input to *Rag1* (47) and Bcl11b into *Dntt* as shown in the present study. However, three sets of TCR-coding gene segments are also rendered accessible transcriptionally and thus opened for recombination during the stages immediately following commitment: the *Tcrb* locus, the *Tcrd* locus which is embedded in the center of the *Tcra* locus, and the complex cluster of recombination units making up the *Tcrg* loci. Several regulators in the GRN model directly affect activity of these loci. The *Tcrb* locus appears to receive positive inputs from Ets1, Runx1, GATA-3, and E proteins (73-76), but the most significant

functionally is the stringent requirement of the *Tcrg* locus for Stat5a/Stat5b input triggered by IL-7 receptor signaling (77, 78). Whereas most *Tcr* gene complex rearrangement is positively regulated by E proteins (79), the first fetal lineage of TCRγδ cells is also distinctive for its preference for low E protein and unusual tolerance of high Id2 levels (80), which are typical of the early fetal thymus (81). Bcl11b-deficient pro-T cells are very severely blocked in TCRαβ T-cell development but not completely blocked in TCRγδ T-cell development, and this can be related to the different regulatory requirements of these different *Tcr* complexes for opening.

SUPPLEMENTARY REFERENCES

1. Li L, Leid M, & Rothenberg EV (2010) An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. *Science* 329(5987):89-93.

2. Scripture-Adams DD, *et al.* (2014) GATA-3 dose-dependent checkpoints in early T cell commitment. *J Immunol* 193(7):3470-3491.

3. Li P, *et al.* (2010) Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. *Science* 329(5987):85-89.

4. Zhang JA, Mortazavi A, Williams BA, Wold BJ, & Rothenberg EV (2012) Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* 149(2):467-482.

5. Yui MA, Feng N, & Rothenberg EV (2010) Fine-scale staging of T cell lineage commitment in adult mouse thymus. *J Immunol* 185(1):284-293.

6. Li B & Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.

7. Champhekar A, *et al.* (2015) Regulation of early T-lineage gene expression and developmental progression by the progenitor cell transcription factor PU.1. *Genes Dev* 29(8):832-848.

8. Johnson WE, Li C, & Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8(1):118-127.

9. Mortazavi A, *et al.* (2013) Integrating and mining the chromatin landscape of cell-type specificity using self-organizing maps. *Genome Res* 23(12):2136-2148.

10. Kastner P, *et al.* (2010) Bcl11b represses a mature T-cell gene expression program in immature $CD4^+CD8^+$ thymocytes. *Eur J Immunol* 40(8):2143-2154.

11. Xu W, *et al.* (2013) E2A transcription factors limit expression of Gata3 to facilitate T lymphocyte lineage commitment. *Blood* 121(9):1534-1542.

12. Zhang HM, *et al.* (2012) AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res* 40(Database issue):D144-149.

13. Avram D, Fields A, Senawong T, Topark-Ngarm A, & Leid M (2002) COUP-TF (chicken ovalbumin upstream promoter transcription factor)-interacting protein 1 (CTIP1) is a sequence-specific DNA binding protein. *Biochem.J* 368(Pt 2):555-563.

14. Kueh HY & Rothenberg EV (2012) Regulatory gene network circuits underlying T cell development from multipotent progenitors. *Wiley Interdiscip Rev Syst Biol Med* 4(1):79-102.

15. Del Real MM & Rothenberg EV (2013) Architecture of a lymphomyeloid developmental switch controlled by PU.1, Notch and Gata3. *Development* 140(6):1207-1219.

16. Kueh HY, *et al.* (2016) Asynchronous combinatorial action of four regulatory factors activates *Bcl11b* for T cell commitment. *Nat Immunol* 17(8):956-965.

17. Yui MA & Rothenberg EV (2014) Developmental gene networks: a triathlon on the course to T cell identity. *Nat Rev Immunol* 14(8):529-545.

18. Rosenbauer F, *et al.* (2006) Lymphoid cell growth and transformation are suppressed by a key regulatory element of the gene encoding PU.1. *Nat Genet.* 38(1):27-37.

19. Huang G, *et al.* (2008) PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. *Nat Genet.* 40(1):51-60.

20. Zarnegar MA, Chen J, & Rothenberg EV (2010) Cell-type-specific activation and repression of PU.1 by a complex of discrete, functionally specialized cis-regulatory elements. *Mol Cell Biol* 30(20):4922-4939.

21. De Obaldia ME, *et al.* (2013) T cell development requires constraint of the myeloid regulator C/EBP-α by the Notch target and transcriptional repressor Hes1. *Nat Immunol* 14(12):1277-1284.

22. Germar K, *et al.* (2011) T-cell factor 1 is a gatekeeper for T-cell specification in response to Notch signaling. *Proc Natl Acad Sci U S A* 108(50):20060-20065.

23. Weber BN, *et al.* (2011) A critical role for TCF-1 in T-lineage specification and differentiation. *Nature* 476(7358):63-68.

24. Möröy T & Khandanpour C (2011) Growth factor independence 1 (Gfi1) as a regulator of lymphocyte development and activation. *Semin Immunol* 23(5):368-378.

25. Ng SY, Yoshida T, Zhang J, & Georgopoulos K (2009) Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. *Immunity* 30(4):493-507.

26. Arenzana TL, Schjerven H, & Smale ST (2015) Regulation of gene expression dynamics during developmental transitions by the Ikaros transcription factor. *Genes Dev* 29(17):1801-1816.

27. Emambokus N, *et al.* (2003) Progression through key stages of haemopoiesis is dependent on distinct threshold levels of c-Myb. *EMBO J.* 22(17):4478-4488.

28. Allen RD, III, Bender TP, & Siu G (1999) c-Myb is essential for early T cell development. *Genes Dev.* 13:1073-1078.

29. Gimferrer I, *et al.* (2011) Regulation of GATA-3 expression during CD4 lineage differentiation. *J Immunol* 186:3892-3898.

30. Schwartz R, Engel I, Fallahi-Sichani M, Petrie HT, & Murre C (2006) Gene expression patterns define novel roles for E47 in cell cycle progression, cytokine-mediated signaling, and T lineage development. *Proc Natl Acad Sci U S A* 103(26):9976-9981.

31. Li H, Ji M, Klarmann KD, & Keller JR (2010) Repression of Id2 expression by Gfi-1 is required for B-cell and myeloid development. *Blood* 116(7):1060-1069.

32. Yücel R, Karsunky H, Klein-Hitpass L, & Möröy T (2003) The transcriptional repressor Gfi1 affects development of early, uncommitted c-Kit$^+$ T cell progenitors and CD4/CD8 lineage decision in the thymus. *J. Exp. Med.* 197(7):831-844.

33. Yoshida T, *et al.* (2013) Transcriptional regulation of the *Ikzf1* locus. *Blood* 122(18):3149-3159.

34. Geimer Le Lay AS, *et al.* (2014) The tumor suppressor Ikaros shapes the repertoire of Notch target genes in T cells. *Sci Signal* 7(317):ra28.

35. Chari S & Winandy S (2008) Ikaros regulates Notch target gene expression in developing thymocytes. *J Immunol.* 181(9):6265-6274.

36. Yashiro-Ohtani Y, *et al.* (2009) Pre-TCR signaling inactivates Notch1 transcription by antagonizing E2A. *Genes Dev* 23(14):1665-1676.

37. Ikawa T, Kawamoto H, Goldrath AW, & Murre C (2006) E proteins and Notch signaling cooperate to promote T cell lineage specification and commitment. *J Exp Med* 203(5):1329-1342.

38. Pereira de Sousa A, *et al.* (2012) Inhibitors of DNA binding proteins restrict T cell potential by repressing Notch1 expression in Flt3-negative common lymphoid progenitors. *J Immunol* 189(8):3822-3830.

39. Welinder E, *et al.* (2011) The transcription factors E2A and HEB act in concert to induce the expression of FOXO1 in the common lymphoid progenitor. *Proc Natl Acad Sci U S A* 108(42):17402-17407.

40. Miyazaki M, *et al.* (2011) The opposing roles of the transcription factor E2A and its antagonist Id3 that orchestrate and enforce the naive fate of T cells. *Nat Immunol* 12(10):992-1001.

41. David-Fung ES, *et al.* (2009) Transcription factor expression dynamics of early T-lymphocyte specification and commitment. *Dev Biol* 325(2):444-467.

42. Zohren F, *et al.* (2012) The transcription factor Lyl-1 regulates lymphoid specification and the maintenance of early T lineage progenitors. *Nat Immunol* 13(8):761-769.

43. Braunstein M & Anderson MK (2012) HEB in the spotlight: Transcriptional regulation of T-cell specification, commitment, and developmental plasticity. *Clin Dev Immunol* 2012:678-705.

44. Wilson NK, *et al.* (2010) Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* 7(4):532-544.

45. Yun TJ & Bevan MJ (2003) Notch-regulated ankyrin-repeat protein inhibits Notch1 signaling: multiple Notch1 signaling pathways involved in T cell development. *J Immunol.* 170(12):5834-5841.

46. Krebs LT, Deftos ML, Bevan MJ, & Gridley T (2001) The Nrarp gene encodes an ankyrin-repeat protein that is transcriptionally regulated by the notch signaling pathway. *Dev Biol* 238(1):110-119.

47. García-Ojeda ME*, et al.* (2013) GATA-3 promotes T cell specification by repressing B cell potential in pro-T cells. *Blood* 121(10):1749-1759.

48. Nechanitzky R*, et al.* (2013) Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. *Nat Immunol* 14(8):867-875.

49. Banerjee A, Northrup D, Boukarabila H, Jacobsen SE, & Allman D (2013) Transcriptional repression of *Gata3* is essential for early B cell commitment. *Immunity* 38(5):930-942.

50. De Obaldia ME & Bhandoola A (2015) Transcriptional regulation of innate and adaptive lymphocyte lineages. *Annu Rev Immunol* 33:607-642.

51. Serafini N, Vosshenrich CA, & Di Santo JP (2015) Transcriptional regulation of innate lymphoid cell fate. *Nat Rev Immunol* 15(7):415-428.

52. Laiosa CV, Stadtfeld M, Xie H, de Andres-Aguayo L, & Graf T (2006) Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBPα and PU.1 transcription factors. *Immunity* 25(5):731-744.

53. Dionne CJ*, et al.* (2005) Subversion of T lineage commitment by PU.1 in a clonal cell line system. *Dev Biol* 280(2):448-466.

54. Franco CB, *et al.* (2006) Notch/Delta signaling constrains reengineering of pro-T cells by PU.1. *Proc Natl Acad Sci U S A* 103(32):11993-11998.

55. Dakic A, *et al.* (2005) PU.1 regulates the commitment of adult hematopoietic progenitors and restricts granulopoiesis. *J. Exp. Med.* 201(9):1487-1502.

56. Kueh HY, Champhekar A, Nutt SL, Elowitz MB, & Rothenberg EV (2013) Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. *Science* 341(6146):670-673.

57. Spicuglia S, *et al.* (2002) Promoter activation by enhancer-dependent and -independent loading of activator and coactivator complexes. *Mol.Cell* 10(6):1479-1487.

58. Taghon T, Yui MA, & Rothenberg EV (2007) Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3. *Nat Immunol* 8(8):845-855.

59. Ohmura S, *et al.* (2016) Lineage-affiliated transcription factors bind the *Gata3* Tce1 enhancer to mediate lineage-specific programs. *J Clin Invest* 126(3):865-878.

60. Li L, *et al.* (2013) A far downstream enhancer for murine Bcl11b controls its T-cell specific expression. *Blood* 122(6):902-911.

61. Ji H, *et al.* (2010) Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* 467(7313):338-342.

62. Carotta S, Wu L, & Nutt SL (2010) Surprising new roles for PU.1 in the adaptive immune response. *Immunol Rev* 238(1):63-75.

63. Sugiyama D, *et al.* (2008) Differential context-dependent effects of friend of GATA-1 (FOG-1) on mast-cell development and differentiation. *Blood* 111(4):1924-1932.

64. Hoogenkamp M, *et al.* (2009) Early chromatin unfolding by RUNX1: a molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program. *Blood* 114(2):299-309.

65. Hoogenkamp M, *et al.* (2007) The *Pu.1* locus is differentially regulated at the level of chromatin structure and noncoding transcription by alternate mechanisms at distinct developmental stages of hematopoiesis. *Mol Cell Biol* 27(21):7425-7438.

66. Cauchy P, *et al.* (2015) Dynamic recruitment of Ets1 to both nucleosome-occupied and -depleted enhancer regions mediates a transcriptional program switch during early T-cell differentiation. *Nucleic Acids Res*.

67. Ramirez K, *et al.* (2012) Gene deregulation and chronic activation in natural killer cells deficient in the transcription factor ETS1. *Immunity* 36(6):921-932.

68. Zook EC, *et al.* (2016) The ETS1 transcription factor is required for the development and cytokine-induced expansion of ILC2. *J Exp Med* 213(5):687-696.

69. Schilham MW, *et al.* (1998) Critical involvement of Tcf-1 in expansion of thymocytes. *J. Immunol.* 161:3984-3991.

70. Tiemessen MM, *et al.* (2012) The nuclear effector of Wnt-signaling, Tcf1, functions as a T-cell-specific tumor suppressor for development of lymphomas. *PLoS Biol* 10(11):e1001430.

71. Yu S, *et al.* (2012) The TCF-1 and LEF-1 transcription factors have cooperative and opposing roles in T cell development and malignancy. *Immunity* 37(5):813-826.

72. Dose M, *et al.* (2014) β-Catenin induces T-cell transformation by promoting genomic instability. *Proc Natl Acad Sci U S A* 111(1):391-396.

73. McMillan RE & Sikes ML (2009) Promoter activity 5' of Dβ2 is coordinated by E47, Runx1, and GATA-3. *Mol Immunol* 46(15):3009-3017.

74. Oestreich KJ, *et al.* (2006) Regulation of TCRβ gene assembly by a promoter/enhancer holocomplex. *Immunity* 24(4):381-391.

75. Eyquem S, Chemin K, Fasseu M, & Bories JC (2004) The Ets-1 transcription factor is required for complete pre-T cell receptor function and allelic exclusion at the T cell receptor β locus. *Proc.Natl.Acad.Sci.U.S.A* 101(44):15712-15717.

76. Wotton D, Ghysdael J, Wang S, Speck NA, & Owen MJ (1994) Cooperative binding of Ets-1 and core binding factor to DNA. *Mol. Cell. Biol.* 14:840-850.

77. Yao Z, *et al.* (2006) Stat5a/b are essential for normal lymphoid development and differentiation. *Proc.Natl.Acad.Sci.U.S.A* 103(4):1000-1005.

78. Ye SK, *et al.* (2001) The IL-7 Receptor controls the accessibility of the TCRγ Locus by Stat5 and histone acetylation. *Immunity* 15(5):813-823.

79. Agata Y, *et al.* (2007) Regulation of T cell receptor β gene rearrangements and allelic exclusion by the helix-loop-helix protein, E47. *Immunity* 27(6):871-884.

80. Bain G, Romanow WJ, Albers K, Havran WL, & Murre C (1999) Positive and negative regulation of V(D)J recombination by the E2A proteins. *J. Exp. Med.* 189:289-300.

81. David-Fung ES, *et al.* (2006) Progression of regulatory gene expression states in fetal and adult pro-T-cell development. *Immunol Rev* 209:212-236.

82. Hirose S, *et al.* (2015) Bcl11b prevents the intrathymic development of innate CD8 T cells in a cell intrinsic manner. *Int Immunol* 27(4):205-215.

83. Heng TSP, Painter MW, & Consortium TIGP (2008) The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol.* 9(10):1091-1094.

84. Egawa T & Littman DR (2008) ThPOK acts late in specification of the helper T cell lineage and suppresses Runx-mediated commitment to the cytotoxic T cell lineage. *Nat.Immunol* 9(10):1131-1139.

SUPPLEMENTARY TABLE LEGENDS

Table S1: Identification of Gold Standard Bcl11b Knockout gene sets and comparison with previously published Bcl11b knockout results

A: Lists of Gold Standard genes upregulated and downregulated in Bcl11b knockouts. B. and C.: Genes with actual $\log_2$ fold change levels, listed by each component comparison of sample tracks that was used to identify the gold standards. Positive $\log_2$ fold change levels indicate upregulation when Bcl11b is deleted. B: Genes in each component listed in order of false discovery rate (all <0.05). C: Genes in each component list after removal of miR and Gm transcripts, listed in order of fold change between Bcl11b knockout and wildtype. D: For reference, in addition to the intersection list shown in A, a union list is provided of all the genes showing up as significant in any of these three comparisons. E. and F.: Venn diagram comparison of the Gold Standard list of Bcl11b-knockout sensitive genes with genes identified in previous studies (3, 10). E: comparison of genes upregulated in Bcl11b knockouts. F: comparison of genes downregulated in Bcl11b knockouts.

Table S2: Metacluster identities of genes in Self Organizing Map and gene expression regulator representation in Bcl11b-regulated genes

A: List of memberships of each metacluster in the SOM, with each gene in a metacluster annotated by the coordinates of the component cluster it falls within. Memberships of all metaclusters (#0-299) are given. Note that clusters and metaclusters are defined by similarity of expression patterns, and do not contain equal numbers of genes. B: Genes in metaclusters associated with different responses to Bcl11b deletion, which also encode annotated transcription factors, chromatin regulators, or transcription cofactors. The metaclusters considered to fall into different classes of Bcl11b response are listed at the top. Then the lists of genes encoding transcriptional regulators (12) in each class are presented.

Table S3: Overlap of Bcl11b gold standard genes with PU.1 obligate repressor-sensitive target genes

Lists of genes shown to be acutely down or up regulated by the PU.1-Engrailed obligate repressor are presented, with the gold standard lists of Bcl11b-knockout affected genes. Genes upregulated and genes downregulated in the Bcl11b knockout samples are each presented with the lists of genes that are also affected (up or down) by the PU.1 obligate repressor. Note that although there are overlaps in all four combinations, there is little bias for a Bcl11b effect to be in the same or the opposite direction from the PU.1 antagonist effect.

Table S4: Overlap of Bcl11b gold standard genes with E2A- sensitive targets

A: Bcl11b knockout gold standard gene lists. B: Lists of genes affected by ≥1.5 x in E2A knockout DN2 cells as compared with control DN2 cells, from ref. (11). C: Venn diagram overlap of representation of Bcl11b knockout upregulated genes, among E2A up- or down-regulated genes. Note the highly biased pattern of overlap. D: Venn diagram overlap of representation of Bcl11b knockout downregulated genes, among E2A up- or down-regulated genes. Note the low but also biased pattern of overlap. A χ^2 calculation (2x2) yields a chi-square statistic of 18.2609, *p*-value 0.000019.

Table S5: Characteristics of RNA-seq samples

The Table presents the characteristics of the RNA-seq samples used in this study. Experimental conditions are included in the descriptive names, column A. Column A also indicates whether the cells were Bcl11b knockout ("KO") or controls. Columns B & C report the short name and sequence track number. Columns D-K present trait scoring. Column O describes the source of the input material: most of the samples are in vitro-

differentiated DN cells derived from fetal liver hematopoietic precursors, but whether or not the precursors were frozen before use is reported in this column. Column P reports the IL-7 concentrations used in the cultures, if any. Column Q indicates the genetic background. YFP= mouse strain includes ROSA26R-eYFP as a Cre recombination indicator gene. These samples were sorted on YFP as well as on Kit/CD44/CD25 phenotype. Note that both Bcl11b$^{fl/fl}$ transgenic lines in our study were repeatedly bred to B6 stock but could not be presumed to be fully congenic at the time these samples were harvested. They are therefore indicated as (129,B6) genetic background. Color annotations indicate whether the samples were published and the data deposited before this study (yellow), and if the samples were part of the gold standard set, the group they were included in is indicated. Green: group A, protocol I. Blue: group B, protocol I. Rose-gray: group C, protocol II.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1.Phenotypes of pro-T cell developmental progression with and without the *Bcl11b* gene. A) Characteristic differentiation profiles of control and Bcl11b-knockout cells developing from fetal liver precursors in a protocol I comparison. Fetal liver precursor cells from the indicated genotypes with a ROSA26R-eYFP Cre recombinase reporter were Cre-transduced prior to initiation of T-cell differentiation culture. Panels show phenotypes of the YFP+ cells after 9 or 12 days of culture. Cartoon below shows pro-T cell developmental stages corresponding to different phenotypes. Note pronounced delay in downregulation of Kit in Bcl11b-deficient cells and persistence of some CD44 expression as control cells generate Kit^{low} $CD44^{low}$ true DN3 cells. B) "Retrograde" differentiation in DN3 cells acutely deprived of Bcl11b after commitment. PLBD mice and control mice with the ROSA26-CreERT2 gene crossed out to a B6 background were used as sources of fresh DN3 thymocytes, sorted from Lin- DN thymocytes as indicated in the top row (day 0). Sort gates to purify DN3 cells are shown on the figure. The cells were then incubated for 4 days in OP9-DL4 co-culture with 100 nM 4-OHT to activate Cre. Lower row (day 4) shows that while the control cells progress rapidly through DN3 to DN4 stage, the Bcl11b-deleted PLBD cells acquire a DN2-like Kit-high phenotype. (Flow cytometry plots came from FACSAria IIu sorter for top row and from MACSQuant analyzer in the bottom row.)

Figure S2: RNA-seq profiles of key genes in Bcl11b KO and control cells. Figure shows UCSC genome browser displays of RNA transcript levels for 12 genes in controls at different stages and in Bcl11b knockout cells. Within each panel, all the vertical scales are identical, ranging from backgrounds of 0.02 fpm (fragments per million reads) to maxima scaled to the peak expression of the genes shown (1 fpm to 50 fpm in different panels). Gene names and scale bars are shown at the top of each panel. Panels

show *Bcl11b* itself, to document the efficiency of deletion; Bcl11b-dependent genes *Gbp4* and the *Cd3g, Cd3d, Cd3g* cluster; and Bcl11b-repressed genes *Tnni1, Itga2b, Tyrobp, Fcer1g, Cxcr5, Zbtb16, Nfil3, Id2,* and *Il2rb*. All of these targets qualified as members of the "gold standard" gene sets except *Nfil3*, which is consistently upregulated in Bcl11b knockout cells made by Protocol II (see Fig. S1) but more weakly, though again consistently, upregulated in cells made by Protocol I. Of these genes, *Itga2b, Tyrobp,* and *Fcer1g* illustrate examples of genes normally expressed in a phase 1 pattern (specific for DN1 or ETP & DN2a) but further upregulated or delayed in downregulation when Bcl11b is absent. Black traces: control cells of indicated developmental types. Red traces: Bcl11b knockout DN2-like cells. Top four lanes: two independent pairs of Bcl11b knockout and control samples. Bottom three lanes: normal developmental series from DN1 (ETP) to DN2b (newly committed). Green box in Bcl11b gene plot shows exon 4, the part of the gene that is deleted in the knockout. This exon encodes 5 of the 6 zinc fingers of the Bcl11b protein that include the DNA binding and protein interaction domains (82), so loss of signal over this exon means *Bcl11b* is inactivated functionally. A hallmark of this deletion, a property of both independently derived *Bcl11b*$^{fl/fl}$ mouse lines used in this study, is that transcripts from the rest of the gene do not disappear but splice to a cryptic downstream exon before terminating. Rectangles in *Cd3* cluster, *Itga2b*, *Fcer1g*, and *Cxcr5* plots indicate region containing the relevant transcripts in these more crowded genomic regions.

Figure S3. Characterization of gold standard sample panel in terms of intrinsic developmental state clock. A) Heat map of differential expression of 90 pro-T cell regulatory genes (rows) in control and Bcl11b knockout samples of the gold standard panel set (columns). These index genes are useful for defining early T-lineage developmental progression in principal component analyses (7). Despite all being sorted in the DN2-DN3 stage interval, note that there is a range of states among the samples such that controls especially differ in terms of global "maturity"as defined by

this panel of regulatory genes. B) The hierarchy of expression of these genes in normal development is shown by listing the genes in order of their weightings along Principal Component 1, which roughly captures advancement through development from DN1 (ETP) all the way to DP stage. High values for weightings along Principal Component 2, also provided, strongly separate DN3 cells poised for β-selection from earlier and later T-cell precursors (7). C) Pairwise comparisons of RNA-seq read counts between ten pairs of Bcl11b-knockout and matched control samples (generated, Cre-treated, and processed in parallel in the same experiment). *Bcl11b* itself and 12 target genes are shown with p-values for each difference calculated by paired Student's t test. The total signal for *Bcl11b* itself is reduced, despite a lack of Bcl11b autoregulation, because the deleted exon includes most of the coding sequence of the *Bcl11b* gene.

Figure S4. Examples of RNA-seq track sets across multiple Bcl11b knockout, control, and reference samples. The figure shows the specific, repeatedly observed upregulation or deferral of downregulation of three Phase 1 genes in Bcl11b knockout cells: (A) *Cd7,* (B) *Kit,* and (C) *Bcl11a*. Samples from Bcl11b knockout cells are indicated as red tracks, samples from controls (Cre treated or reference control) are shown in black, and a set of reference samples treated with γ-secretase inhibitor to test for effects of Notch pathway inhibition are shown in dark blue. Arrows at the top of each panel show the direction of transcription of the gene indicated, and brackets connect the pairs of samples that were generated as knockout & matched control within the same experiment. The five samples at the bottom of each panel are unperturbed reference samples from (4)(FLDN = in vitro differentiated DN cells from fetal liver precursors; Thy = purified from adult thymus), included to relate the knockout phenotype to normal developmental changes of expression. The group of six samples at the top represent adult thymus-derived cells in which Bcl11b was deleted after commitment (Protocol II); with the exception of one other adult thymic DN3 control sample (ThyDN3ctrl+Cre); all the remaining samples were generated under Protocol I

conditions. All the vertical scales for all samples are the same in a given panel, from backgrounds of 0.02 fpm to maximum values of 5 fpm (*Cd7, Kit*) or 1 fpm (*Bcl11a*). As all these genes are Phase 1 genes, the normal patterns of expression decrease markedly from DN2a (FLDN2a) to post-commitment stages, ThyDN3 or ThyDP. The figure shows that despite quantitative differences in the absolute levels and the degrees to which different controls had downregulated these genes, the Bcl11b-knockout cells in each case express higher levels than the controls. Although *Bcl11a* levels in these samples are quite low, the Bcl11b-knockout samples derived from protocol I (FLDN) consistently delay silencing of *Bcl11a* relative to controls.

Figure S5. Expression patterns of gold standard Bcl11b-regulated genes. A) Trends in expression of gold standard genes: developmental change vs. impacts of loss of Bcl11b. Plots show fold change in expression of the 99 gold standard genes when Bcl11b is deleted (y axis) plotted against the fold change in the expression of the same genes as cells progress from DN1 (ETP) through commitment, to DN2b or DN3 (x axes in left, right panels respectively). Data for fold change in response to Bcl11b deletion are $\log_2$ of the geometric means of effects seen in groups in Table S1B,C. Data for fold changes in response to developmental progression are calculated from the difference of the geometric means reported in ref. (4)(**Table S2** of ref. 4). DN1-to-DN2b values show trends in response to *in vitro* commitment, DN1-to-DN3 values show trends in response to *in vivo* commitment. Note negative correlation between developmental progression and Bcl11b knockout effect, suggesting that high upregulation in Bcl11b knockout is often linked to DN1-biased (negative x axis value) expression pattern. B) Quantitative real-time PCR (qPCR) measurements of expression of three Bcl11b-repressed genes, *Tnni1, Zbtb16*, and *Kit*, in acute response to *Bcl11b* deletion at the DN3 stage. Samples of sorted DN3 thymocytes from PLBD (*Bcl11b*$^{fl/fl}$; *ROSA26-Cre-ERT2*) and control (*Bcl11b*$^{+/+}$; *ROSA26-Cre-ERT2* = "ERT2Cre") mice were processed for RNA measurement immediately (PLBD DN3, blue) or treated in parallel with 4-OHT and

cultured for four more days before re-sorting (ERT2Cre+cult – red bars; PLBD+cult – orange bars). In agreement with the surface phenotype shown in Fig. S1B, Bcl11b deletion in the Kit-low DN3 starting population led to upregulation of *Kit* RNA while normal developing controls downregulated *Kit* RNA further under these conditions. *Tnni1* and *Zbtb16* RNAs were low or undetectable (values below baseline) in the initially purified DN3 samples but were upregulated dramatically after *Bcl11b* deletion. C) Expression histograms of indicated Bcl11b-dependent gold standard genes, based on data from **Table S2** of ref. 4. Note rising patterns toward DN3. D) Expression histograms of indicated Bcl11b-repressed gold standard genes, based on data from **Table S2** of ref. 4. Note pattern diversity but frequent phase 1 (highest in DN1 and DN2a) bias.

Figure S6. Properties of the gold standard gene set: KEGG pathway enrichment and relation to Bcl11b target genes found at later developmental stages. A) Top hit pathway among genes upregulated in Bcl11b knockout cells is the Natural Killer Cell Cytotoxicity pathway. Graphic is from DAVID <https://david.ncifcrf.gov/>, a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway diagram with red stars showing the pathway components that are represented among these Bcl11b-repressed gold standard genes. Note that sometimes the names under which the genes are listed in the KEGG diagram are different from the gene names from GENCODE 1, e.g. TRAIL for Tnfsf10. B) Gold standard genes upregulated in Bcl11b knockout cells were compared with genes upregulated in Bcl11b knockout DP cells in the studies of Li et al. (3) and Kastner et al. (10), respectively. We used **Table S3** of ref. (3) and **Table S2** of ref. (10) as sources, clipping off genes that the authors reported to show less than a 2x fold change (i.e. for all genes considered, $|\log_2 FC| > 1$). Venn diagrams show the overlaps between the gene sets, indicating the names of the genes in the shared sets and the names of genes included in our gold standard set that were not reported as previous Bcl11b-affected targets. This could be a slight underestimate, since nomenclature has changed

for some genes since publication of refs. (3) and (10). C) Gold standard genes downregulated in Bcl11b knockout cells were compared with genes downregulated in Bcl11b knockout DP cells in Li et al., and Kastner et al., selected as in panel (B). Far fewer genes appear to be Bcl11b-dependent in both the early context and the later one, as compared to the number of genes that appear to be shared repression targets of Bcl11b in both contexts.

Figure S7. Broader patterns of developmental expression of gold standard genes in the Immunological Genome microarray data resource. The figure shows heat maps taken from the Immunological Genome (83) web site <www.immgen.org>, using the "My Gene Set" interactive heat map tool for seach and display (searched April 27, 2016). A) Gold standard genes dependent on Bcl11b were used to interrogate the data resource. Genes are shown in rows. Samples in database define columns. Orange to yellow color means relatively high expression. Dark blue to blue color means low expression. Columns under the purple bar represent different stages of T-cell development, with an orange star at the bottom to indicate DN3a stage. Note that key pre-commitment pro-T cell stages here are found in the narrow purple zone to the left of the orange star. Many Bcl11b dependent genes can be seen to peak at DN3a, like Bcl11b itself (red arrow on left). B) Bcl11b repressed genes, shown using the same convention as in (A). For internal comparison, the expression pattern of Bcl11b itself (red arrow on left) is also loaded with this gene set. Note that non-T expression of these genes (outside the zone marked by purple, top) is not confined to NK cells (light purple zone at right), but can also include strikingly high expression in myeloid cells (green, pink, red zones) and stem or even B cells (blue zones). In this panel, the magnification does not allow the gene names all to be shown in legible form on the figure itself. A full list of these genes in order from top to bottom is as follows, with the genes marked on the figure in bold caps as landmarks: *Gimap4, Il2rb, Txk, Fcer1g, Tyrobp, Gstml, Trf, Cd63, Cd9, Aoah, Fcgr3, Dok3, Fes, Cd7, Klrd1, Id2, Ccr2, S100a6, Cpa3,* ***KIT****, Cd244, Clnk, Gpr97,*

Iqgap2, Chn2, ***ZBTB16****, Appl2, Klhdc2, Osbpl5, Pear1, Ivns1abp, Sccpdh, 2900026A02Rik, Bace1, Ccdc135,* ***ITGA2B****, 1700112E06Rik, P2rx1, Kcne3, Col15a1, Arsi, Dlk1, Colq, Cd163l1, Chpf, Rgs3, Fdx1, Mcart6, Pgam2, Tnni1, 1700026L01Rik, Fgf3, Tas1r1, Ivns1abp (repeat, another probe set), Col9a3, Espn, Fgf15, Pdzk1, Trpm1, Arap3, Lpcat4, Rgs3, Lag3, Mcart6 (repeat, another probe set), Gramd1a, Polm, Rab19, Ecm1, Timp2,* ***BCL11B****, Tnfsf10, Cxcr5, Pou2af1, Cd72, Myo1e, Gpr183, Itgb7, Pde2a, Gimap8, Tspan32.*

Figure S8. ChIP-sequencing analysis of Bcl11b binding in DN3 cells shows sites of engagement at positively and negatively regulated "gold standard" Bcl11b targets

(A)-(G): Bcl11b occupancy patterns at the indicated genes are shown from two independent samples of Bcl11b ChIP-seq and background control samples (1% of ChIP input). Thick blue arrows at the top of each panel show the direction of transcription and the extent of the transcribed region of the indicated differentially regulated target gene. In (A)-(C), Bcl11b ChIP patterns are compared with patterns of histone H3K4me2 modification at the DN1 (pre-commitment) and DN2b (immediately post-commitment) stages, as measured in ref. (4). A) Bcl11b binding to *Tyrobp*, a gene that undergoes silencing and chromatin closing, as shown by disappearance of a peak in the H3K4me2 tracks, during commitment. B) Bcl11b binding to the *Cd3* gene cluster, three genes that undergo upregulation and chromatin opening, as shown by H3K4me2 signals, during commitment. C) Bcl11b binding to *Tnni1*, a gene that is normally silent and closed before and after commitment for most αβ-lineage T cells. D) Bcl11b binding to *Il2rb*. E) Bcl11b binding to *Zbtb16*. F) Bcl11b binding to *Cd7*. G) Bcl11b binding to *Nfil3*. Although *Nfil3* did not fulfill the criteria to be a "gold standard" gene, it was upregulated in many DN2-3 Bcl11b knockout samples as compared to controls, as illustrated in Fig. S2. H) Top three motifs enriched in Bcl11b binding peaks in DN3 cells are shown. Binding detected by ChIP does not prove function, but these results and those in Fig. 3A,B indicate specific candidate sites through which the Bcl11b functions demonstrated here may be mediated.

Figure S9. Full representation of the metaclusters in the early T specification SOM. A) The positions of all the numbered metaclusters in the total map are shown. Note that the map is actually a toroidal surface, which is opened out to a flat presentation simply for display. This is the underlying structure of the gene clusters that are differentially activated and presented as enriched or depleted signals in Fig. 3C-G. B) Labeling of major SOM regions associated with developmental gene expression change and with effects of Bcl11b deletion. These regions are defined by the differential expression patterns shown in Fig. 3C and D. Region A=genes upregulated during normal pro-T cell development. Region B=genes downregulated during normal pro-T cell development, including many phase 1 regulatory genes. Region C1: genes repressed by Bcl11b (upregulated upon Bcl11b deletion) that have expression similarities with phase 1 genes. Region C2: genes repressed by Bcl11b (upregulated upon Bcl11b deletion) that have no normal expression enrichment in normal pro-T cell development. In contrast to region C1, which overlaps region B, region C2 is apparently disjunct from region A despite its proximity.

Figure S10. Eigengene expression pattern heatmaps in representative SOM metaclusters. Each metacluster is comprised of multiple SOM unit clusters of genes that share related expression patterns across our whole set of RNA-seq samples. These collections of gene expression patterns that characterize each metacluster are shown here for four metaclusters of interest, in terms of their eigengenes. These are the metaclusters in the SOM overall the positions of which are indicated in Fig. 3C. Here, in panels (A)-(D), each row of the heatmap depicts the expression levels for the eigengene of that SOM unit in each sample. The number at the left side identifying each row is the coordinate of each SOM unit that is included within the indicated metacluster. The developmental stage and treatment used to generate each sample is depicted with

different color codes underneath the heatmap, identifying each column. A) Heatmap of metacluster 112. B) Heatmap of metacluster 212. C) Heatmap of metacluster 168. D) Heamap of metacluster 26.

Figure S11. Runx3 distal promoter upregulation in *Bcl11b* knockout cells
Browser tracks are shown as in Fig. S4, focusing on the distinctive effect of Bcl11b deletion on *Runx3* expression from the distal promoter. *Runx3* transcription units are approximately indicated by dark blue arrows, top. The cyan rectangle over the tracks indicates the position of the distal promoter of *Runx3*. Although transcripts are commonly expressed from the proximal promoter of *Runx3* in multiple T-cell subsets, expression from the distal promoter is associated with T-cell commitment to a cytolytic lineage (84). In the track key on the left, matched pairs of control and Bcl11b-knockout samples generated and analyzed within a single experiment are connected by brackets, as in Fig. S4. For details of the samples, see Table S5.

Figure S12. Trait enrichment for hypothesis testing map of SOM metaclusters, and Venn diagrams of Bcl11b regulome overlaps with regulomes of PU.1 and E2A. A) Full comprehensive heatmap of the correlation between SOM metaclusters of the indicated numbers and different developmental or treatment traits. The heatmap is laid out with a break at the point shown in order to allow cluster numbers to be legible. Metaclusters with significant positive correlation against the traits indicated at the bottom are highlighted in red and those with negative correlation are in blue. To retrieve gene lists within each cluster with a desired trait enrichment property, see Table S2.

For interpretation, note that color scale is relative to averages over the whole sample set, but not over all of T-cell biology. Thus, given the preponderance of DN1 and DN2a/2b samples in this analysis, both phase 1 and phase 1-like, the DN1 stage as

such is mainly distinguished here by a relatively small number of more highly represented metaclusters, i.e., metaclusters 244, 63, 265, 248, and 264, and underrepresentation of only a few later-activated metaclusters, i.e. metaclusters 232, 193, 64, 200, 184, 26, 25, and 59, despite its large difference from most mature T cells, which are not studied here. The nature of the comparison yields more differences between Bcl11b KO and control DN2b cells.

Also note that "DN2" here primarily refers to experimentally perturbed samples which cannot be perfectly equated with either normal DN2a or normal DN2b cells. Given the nature of this study, most of them are Bcl11b knockout in this sample set, accounting for the large overlap between clusters correlating with "Bcl11b KO" and with "DN2".

B) Venn diagram of effects shared between genes up- and downregulated by Bcl11b deletion (this study) and by acute introduction of a PU.1 obligate repressor (PU.1-ENG)(7). Note lack of bias in directions of shared effects.

C) Venn diagram of effects shared between genes up- and downregulated by Bcl11b deletion (this study) and by deletion of the E2A isoform E47 in similar fetal liver-derived DN2 cells (11). The overlaps are highly biased toward concordance, as shown, and a simple χ^2 test of dependence between these variables from the 2x2 matrix of upregulated and downregulated genes for each perturbation yields a chi-square statistic of 18.3, p-value=0.000019.

Fig. S1

A

Bcl11b+/+
Bcl11b-/-
Day 9
Day12
c-Kit
CD44
CD25
4.28
17.8
8.35
69.6
4.32
34.2
9.04
52.5
9.92
52.1
3.12
34.9
9.24
67.7
4.21
18.9
1.96
7.12
8
82.9
1.52
2.92
8.82
86.7
10.1
66.5
2
21.4
2.01
6.37
10.5
81.1
ETP
DN2a
DN2b
DN3a

B

ERT2-Cre
PLBD (Bcl11b^{ff} ERT2-Cre)
Day 0
Day 4 after 4OHT
sort gate
46.6
42.1
1.13
3.31
46.1
49.4
5.93
81.4
0.91
11.8
c-Kit
CD25

Bcl11b
Itga2b
Zbtb16
Gbp9-Gbp4
Tyrobp
Nfil3
Cd3g-Cd3d-Cd3e
Fcer1g
Id2
Tnni1
Cxcr5
Il2rb
B6-Cre DN2b/3-rep1
Bcl11bko DN2a*-rep1
B6-Cre DN2b/3-rep2
Bcl11bko DN2a*-rep2
B6 DN1
B6 DN2a
B6 DN2b

Fig. S2

A

Gold standard samples compared for index gene expression levels

B

Index gene expression in normal development, PC1 and PC2 loading

	PC1	PC2
gene_name	*ETP-->DP*	*ETP,DP-->DN3a*
Spi1 (Sfpi1,PU.1)	-0.259	0.027
Hhex	-0.220	-0.041
Lyl1	-0.206	-0.113
Bcl11a	-0.192	0.117
Erg	-0.171	0.164
Gfi1b	-0.166	-0.112
Nfe2	-0.157	0.023
Ctbp2	-0.153	-0.066
Rbpms	-0.146	0.066
Irf5	-0.134	-0.135
Mdfi	-0.125	0.070
Dnmt3b	-0.124	0.057
Hes1	-0.119	0.250
Mycn	-0.112	-0.039
Hmga2	-0.109	-0.114
Fosb	-0.090	0.010
Prr13	-0.083	0.003
Runx3	-0.078	0.017
Fos	-0.074	-0.013
Arid3a	-0.071	-0.073
Nfix	-0.070	0.069
Sirt3	-0.067	-0.046
Sox13	-0.066	0.137
Creg1	-0.062	0.018
Dusp5	-0.058	-0.069
Etv6	-0.054	0.039
Ybx3 (Csda)	-0.053	0.027
Id2	-0.037	0.059
Myc	-0.032	0.077
Tsc22d1	-0.028	0.065
Cited2	-0.025	-0.095
Ern1	-0.025	0.092
Ahr	-0.022	0.148
Jun	-0.019	0.058
Mta3	-0.011	0.067
Pbx4	-0.009	0.080
Klf6	-0.004	-0.061
Hdac5	0.000	-0.096
Notch1	0.010	0.122
Nfkb2	0.019	0.090
Hdac4	0.021	0.206
Stat1	0.021	0.059
Card9	0.026	-0.093
Spib	0.026	0.280
Hdac10	0.028	-0.068
Notch3	0.031	0.148
Gtf2ird1	0.032	-0.061
Rbl2	0.033	-0.051
Tle2	0.039	-0.308
Tcf12	0.048	0.050
Arntl	0.048	-0.057
Nfkbia	0.049	-0.065
Nfatc3	0.049	-0.062
Brwd1	0.050	0.033
Tsc22d3	0.051	-0.029
Setd1b	0.051	-0.007
Tcf7	0.051	0.006
Zfp329	0.051	0.042
Zfp579	0.053	-0.011
Tcof1	0.053	-0.313
Ssbp2	0.054	-0.049
Rb1	0.054	-0.009
Zfp39	0.056	-0.042
Hivep3	0.057	0.124
Klf13	0.057	-0.059
Scml4	0.058	0.076
Phf1	0.059	-0.018
Txnip	0.059	-0.072
Gtf2h4	0.060	-0.111
Ablim1	0.068	0.025
Klf7	0.069	-0.089
Snai3	0.070	0.092
Hivep2	0.071	0.023
Mier1	0.073	0.006
Mef2a	0.073	-0.073
Nr3c1	0.078	0.084
Rag2	0.080	0.028
Id3	0.086	0.051
Satb1	0.097	-0.058
Btg2	0.105	0.050
Ets1	0.106	0.079
E2f2	0.106	0.009
Ets2	0.114	-0.053
Bcl6	0.115	-0.125
Bcl11b	0.134	0.175
Lef1	0.158	0.218
Rag1	0.161	0.100
Pou6f1	0.174	0.043
Rorc	0.249	-0.089
Ikzf3	0.263	0.074

C

Pairwise gene expression changes Bcl11b KO vs. controls

Fig. S3

Fig. S4

A Cd7
B Kit
C Bcl11a
ThyDN3 control d0
ThyDN3 ctrl+Cre d3
ThyDN3 ctrl+Cre d4
ThyDN3 Bcl11b-fl d0
ThyDN3 Bcl11bko d3
ThyDN3 Bcl11bko d4
FLDN2b control
ThyDN3 ctrl+Cre
FLDN2b ctrl+Cre
FLDN2 Bcl11bko
FLDN2 Bcl11bKO
FLDN2a +GSI
FLDN2a control
FLDN2b +GSI
FLDN1 control
ThyDP control

Fig. S5

A Bcl11b ko effect on “gold standard” genes relative to their normal expression

B Activation of Bcl11b-repressed “gold standard” genes in response to acute Bcl11b deletion

C Bcl11b dependent “gold standard” genes (downregulated in knockout)

D Bcl11b repressed “gold standard” genes (upregulated in knockout)

A

B Bcl11b repression targets

2900026A02Rik
Appl2
Cxcr5
Gpr183
Itgb7
Klhdc2
Mcart6
Osbpl5
Pde2a
Pou2af1
Rab19
Rgs3
Trf

Cd244
Cd63
Fcer1g
Fes
Lag3
Myo1e
S100a6
Tyrobp

COMMON
Cd9
Id2
Klrd1

1700006J14Rik	Clnk	Gimap4	Pdzk1
1700026L06Rik	Col15a1	Gimap8	Pear1
1700112E06Rik	Col9a3	Gpr97	Pgam2
Aoah	Colq	Gramd1a	Polm
Arap3	Cpa3	Gstm1	Sccpdh
Arsi	Dlk1	Il2rb	Tas1r1
Bace1	Dok3	Iqgap2	Timp2
Ccdc135	Ecm1	Itga2b	Tnfsf10
Ccr2	Espn	Ivns1abp	Tnni1
Cd163l1	Fcgr3	Kcne3	Trpm1
Cd7	Fdx1	Kit	Tspan32
Cd72	Fgf15	Lpcat4	Txk
Chn2	Fgf3	P2rx1	Zbtb16
Chpf			

C Bcl11b dependent targets

1300002E11Rik	Gbp6
Arl5c	Khdc1a
Bst1	Lrrc42
Ccdc109b	Ly6c1
Ccdc153	Mpzl2
Cd3d	Pcmtd1
Cd3e	Slc5a9
Cd3g	Tctex1d1
Gata1	Ttll9
Gbp4	Uaca

Figure S6

Figure S7

A
Bcl11b-dependent genes
Stem or progenitor
B cells
Dendritic cells
Monocyte
Mac
Gr
Pro-T--> pre-T
AlphaBeta T cells
iNKT
GammaDelta T cells
Stroma
NK cells
Cd3g
Ly6c1
Bst1
Gbp6
Khdc1a
Arl5c
Pcmtd1
Khdc1a
Tctex1d1
Slc5a9
Ccdc153
Ttll9
Utaca
Gata1
Lrrc42
Ccdc109b
Mpzl2
Dntt
Gbp4
Bcl11b
Cd3d
Cd3e
B
Bcl11b-repressed genes (+ Bcl11b)
Kit
Zbtb16
Itga2
Bcl11b

H

Rank	Motif	TFs	% of Targets	% of Background	P-value
1		Etv2 (ETS)	44.5%	12.4%	1E-2694
2		RUNX2	26.4%	5.78%	1E-1874
3		Ptf1a	24.2%	15.8%	1E-200

Figure S8

Figure S9

A
Meta-Clusters
metacluster number
300
200
100
0
B
Meta-Clusters
region A
region C2
reg. C1
region B

A
Metacluster 112
29,57
28,58
30,57
28,57
27,57
29,56
Stage
Treatment
B
Metacluster 212
32,56
31,56
30,56
Stage
Treatment
C
Metacluster 168
11,57
12,58
12,57
11,58
14,58
13,57
14,57
13,58
Stage
Treatment
D
Metacluster 26
2,58
1,57
4,59
3,58
5,59
6,0
7,0
Stage
Treatment
Stage
DN1
DN2
DN2a
DN2b
DN3
DP
Treatment
Bcl11b KO
GSI
Pu.1 DN
untreated

Fig. S10

Fig. S11

Runx3
Scale
chr4:
20 kb
134660000
134670000
134680000
134690000
134700000
134710000
134720000
134730000
Runx3
ThyDN3 control d0
ThyDN3 ctrl+Cre d3
ThyDN3 ctrl+Cre d4
ThyDN3 Bcl11b-fl d0
ThyDN3 Bcl11bko d3
ThyDN3 Bcl11bko d4
FLDN2b control
ThyDN3 ctrl+Cre
FLDN2b ctrl+Cre
FLDN2 Bcl11bko
FLDN2b ctrl+Cre
FLDN2 Bcl11bKO
FLDN2b ctrl+Cre
FLDN2 Bcl11bko
FLDN2b ctrl+Cre
FLDN2 Bcl11bko
FLDN2b ctrl+Cre
FLDN2 Bcl11bko
FLDN2b ctrl+Cre
FLDN2b ctrl+Cre
FLDN2 Bcl11bko
FLDN2 Bcl11bko
FLDN2b ctrl+Cre
FLDN2 Bcl11bko
FLDN2b ctrl+Cre
FLDN2a +GSI
FLDN2a control
FLDN2b +GSI
FLDN2b control
FLDN2b +GSI
FLDN2b control
Runx3
Runx3
FLDN1 control
FLDN2a control
FLDN2b control
ThyDN3 control
ThyDP control
ST00000079041_Gm16225
TTMUST00000079034_Runx3
OTTMUST00000079040_Gm16224
ENSMUST00000146152
ENSMUST00000127711
ENSMUST00000056977
ENSMUST00000119564
ENSMUST00000156478
ENSMUST00000137027
ENSMUST00000140642
ENSMUST00000127109

Fig. S12

A
p-value
10^{-10}
10^{-7}
10^{-4}
Cluster #
Bcl11b KO
DN1
DN2
DN2a
DN2b
DN3
DP
GSI
PU.1 inhib
TRAIT

B
116 genes
Bcl11b KO activated (gold std)
73 genes
4 genes
69 genes
4 genes
1
20 genes
1
Bcl11b KO repressed (gold std)
PU.1-ENG repressed
PU.1-ENG activated

C
158 genes
Bcl11b KO activated (gold std)
273 genes
2
54 genes
21 genes
5 genes
17 genes
Bcl11b KO repressed (gold std)
E2A KO repressed
E2A KO activated