



E proteins orchestrate dynamic transcriptional cascades implicated in the suppression of the differentiation of group 2 innate lymphoid cells

Received for publication, April 8, 2020, and in revised form, August 11, 2020. Published, Papers in Press, August 19, 2020, DOI 10.1074/jbc.RA120.013806

Vincent Peng^{1,†}, Constantin Georgescu^{2,‡}, Anna Bakowska³, Aneta Pankow³, Liangyue Qian³, Jonathan D. Wren², and Xiao-Hong Sun^{3,*}

From the ¹Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri, and the ²Program in Genes and Human Diseases and ³Program in Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA

Edited by Peter Cresswell

Group 2 innate lymphoid cells (ILC2s) represent a subset of newly discovered immune cells that are involved in immune reactions against microbial pathogens, host allergic reactions, as well as tissue repair. The basic helix-loop-helix transcription factors collectively called E proteins powerfully suppress the differentiation of ILC2s from bone marrow and thymic progenitors while promoting the development of B and T lymphocytes. How E proteins exert the suppression is not well understood. Here we investigated the underlying molecular mechanisms using inducible gain and loss of function approaches in ILC2s and their precursors, respectively. Cross-examination of RNA-seq and ATAC sequencing data obtained at different time points reveals a set of genes that are likely direct targets of E proteins. Consequently, a widespread down-regulation of chromatin accessibility occurs at a later time point, possibly due to the activation of transcriptional repressor genes such as *Cbfa2t3* and *Jdp2*. The large number of genes repressed by gain of E protein function leads to the down-regulation of a transcriptional network important for ILC2 differentiation.

Group 2 innate lymphoid cells (ILC2s) are a subset of innate lymphoid cells (ILCs) that play important roles in initiating type 2 immunity (1). They are enriched in barrier locations such as the lung, intestine, and skin (2). They respond to signals released by tissue damages, namely IL-25, IL-33 and TSLP, and produce type 2 cytokines primarily consisting of IL-5, IL-9, and IL-13 (3). However, ILC2s exist as heterogeneous populations with diverse characteristics and functionality due to the influences of the tissues in which they reside and the cell origins from which they arise (4, 5). A complete knowledge of the ontogeny of ILC2s will help understand ILC2-mediated immunity.

ILC2s are generated during fetal, neonatal, and adult stages, among which the neonatal phase contributes to the majority of the tissue-resident ILC2s (2). ILC2s differentiate from progenitors such as common lymphoid progenitors (CLP) or lymphoid-multipotent progenitors, which are also responsible for making B and T lymphoid cells as well as NK cells (6). ILC2 dif-

ferentiation has been thought to take place in the fetal liver and bone marrow (3). However, we have established that this also occurs in the thymus, where ILC2s not only differentiate from multipotent early T cell progenitors, namely ETP or DN1 (CD4⁻CD8⁻CD25⁻c-kit⁺) cells, but also from committed T cell precursors, namely CD4⁻CD8⁻CD25⁺c-kit⁻ (DN3) cells (5, 7, 8). In fact, the thymus is a fertile niche for ILC2 differentiation when a critical checkpoint that ensures proper T cell development is disabled.

The basic helix-loop-helix family of transcription factors, collectively called E proteins, serve as a critical checkpoint to promote B and T cell development and suppress ILC2 differentiation (5, 7–10). In the thymus, the *Tcf3* and *Tcf12* genes are expressed to produce their respective E proteins, E12 and E47 from *Tcf3* and HEB from *Tcf12*. These proteins dimerize among themselves and activate transcription by binding to the DNA sequences called E boxes. When both genes are deleted using a thymus-specifically expressed Cre, *plck-Cre*, T cell development is blocked at the DN3 stage when the Cre transgene is first expressed (7, 11). Consequently, an abundance of ILC2s are generated in *Tcf3* and *Tcf12* double knock-out mice (dKO). E protein function can be abolished by their naturally occurring inhibitors, Id proteins, which dimerize with E proteins and prevent DNA binding. When Id1 is ectopically expressed in the thymus, T cell development is arrested and ILC2 development is drastically enhanced, leading to accumulation of 10–100-fold more ILC2s in different tissues (7). The promotion of ILC2 differentiation in the absence of E protein activities are not simply the result of T cell developmental blockade because numerous other animal models with T cell defects (e.g. *Rag1*^{-/-} mice) do not exhibit such robust enhancement of ILC2 production.⁴ Thus, it is likely that E proteins play a specific role in suppressing ILC2 differentiation.

To understand the underlying mechanisms, we sought to elucidate the transcriptional programs controlled by E proteins in the context of ILC2 differentiation. We measured transcriptomes and chromatin accessibilities in T cell precursors and in ILC2s upon inducible loss or gain of E protein function. We observed dynamic changes in gene regulation, which ultimately

This article contains supporting information.

[†]These authors contributed equally to this work.

* For correspondence: Xiao-Hong Sun, sunx@omrf.org.

⁴Bajana and X-H Sun, unpublished data.

leads to a suppression of ILC2 transcriptional programs by E proteins.

Results

Altered gene expression immediately following inducible E protein ablation in T cell precursors

In addition to our previous report of the substantial increase in ILC2 numbers in *Tcf3* and *Tcf12* double knockout mice, we have also shown that deletion of these genes using tamoxifen-inducible Cre enhances ILC2 differentiation from CLP, DN1, and DN3 cells by 20–40-fold when cultured on OP9-DL1 stromal cells for 11–14 days (5, 7). To detect changes in gene expression immediately following E protein ablation, we isolated sufficient DN1 and DN3 cells for RNA-seq after brief cultures on OP9-DL1 stromal cells. We seeded stromal cells with DN1 or DN3 cells isolated from ROSA26-CreERT2;ROSA26-Stop-tdTomato;*Tcf3^{fl/fl}*;*Tcf12^{fl/fl}* (dKO) mice or control mice that were without the floxed alleles of the E protein genes. 4 days later, tamoxifen was added and tdTomato⁺ (indicative of successful Cre-mediated deletion) cells were sorted 24 and 72 h later for RNA-seq (Fig. 1a). At these time points, few ILC2s were detected (5), thus allowing the evaluation of early changes in gene expression in ILC2 precursors.

The transcriptomes of dKO cells and controls at 24 and 72 h were compared separately (Fig. 1b). Although there were substantial overlaps between the differential genes at 24 and 72 h in DN1 and DN3 cells, differences were also found in terms of the magnitude of changes and genes differentially expressed. DN3 cells harbor a higher level of endogenous E proteins than DN1 cells, mostly due to the up-regulation of *Tcf12* (RRID: SCR_019014), and thus the peak time of the changes for some of the genes were delayed compared with DN1 cells. Overall, we obtained lists of 1140 and 1376 differentially regulated protein coding genes at either time point in DN1 and DN3 cells, respectively (Table S1).

DN1 and DN3 cells, being at different developmental stages, have distinct transcriptomes. Because E protein ablation promoted ILC2 differentiation from both cell types, a common set of differential genes will be extremely informative for understanding how E proteins suppress ILC2 differentiation. Remarkably, a profound number of genes up-regulated due to E protein ablation are known to be associated with ILC differentiation. For example, this list consists of *Zbtb16*, *Rora*, *Maf*, *Lmo4*, *Klf6*, *Icos*, *Itgb7*, *Ikzf2*, *Irf4*, and *Gata3* (Fig. 1c). Conversely, among the genes down-regulated by the loss of E proteins are genes previously known to be activated by E proteins and important for T cell development (Fig. 1d) (12). These include *Rag1*, *Rag2*, *Notch1*, *Notch3*, *Cd3g*, *Cd3d*, *Ptcra*, and *Spib*. These data indicate an instrumental role for E-protein activity in controlling the switch between ILC2 and T cell development.

Not all of the genes found in our analyses are directly regulated by E proteins. To identify direct targets, we compared the list of genes altered in DN3 cells to E2A-ChIP-seq data generated by Murre and colleagues (Fig. 1e) (12). Eighty-nine differentially regulated protein-coding genes in cultured DN3 cells were found to be bound by E2A proteins in *Rag2^{-/-}* thymo-

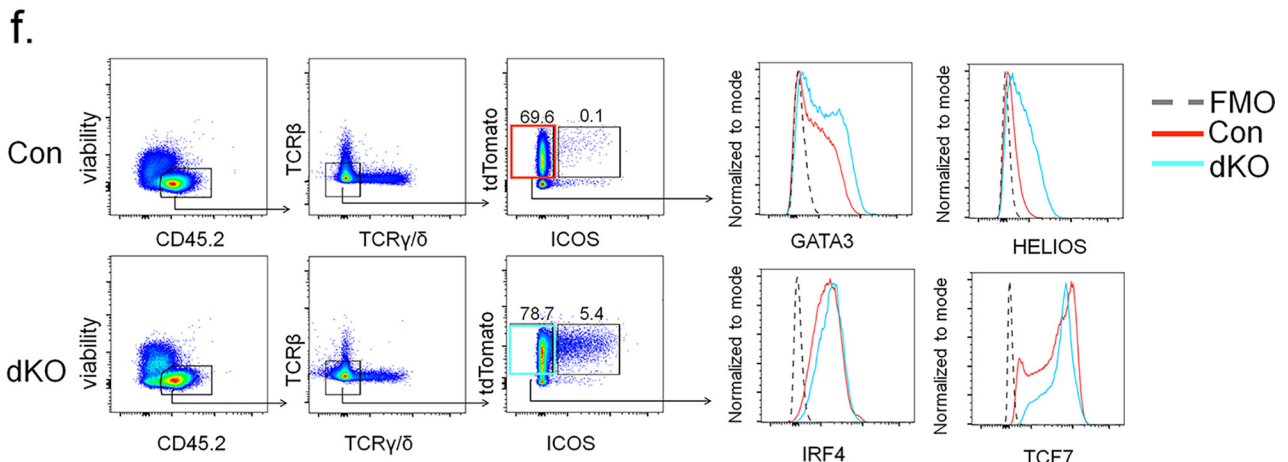
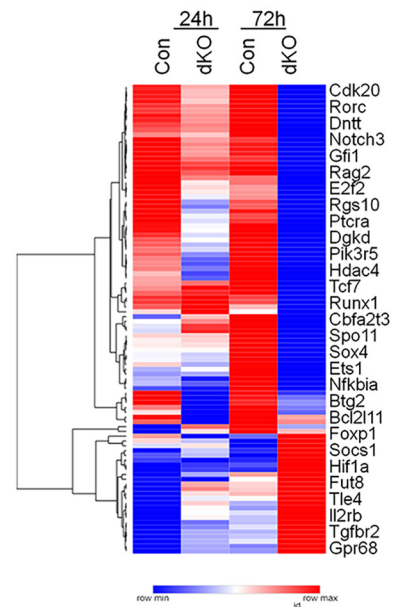
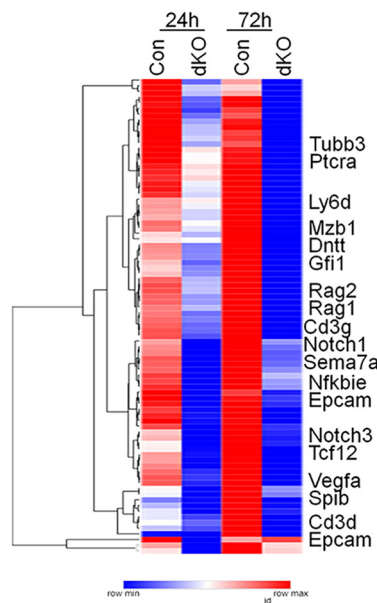
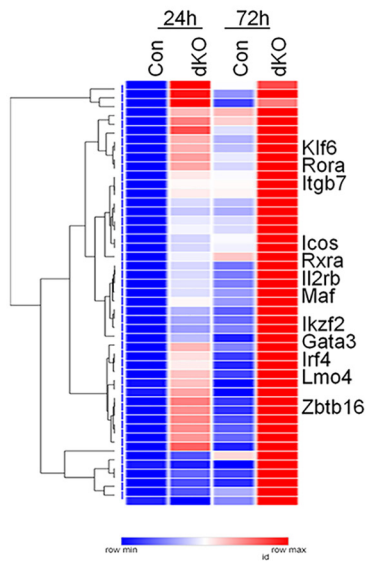
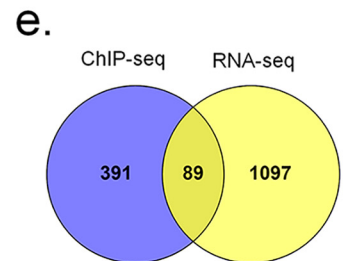
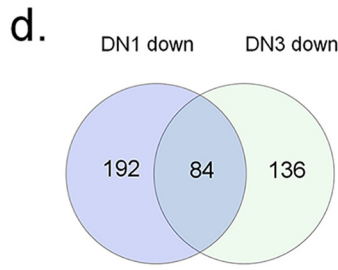
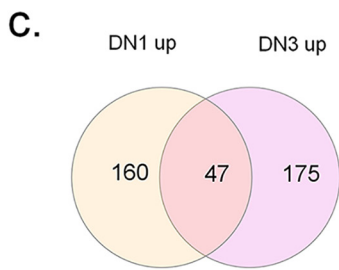
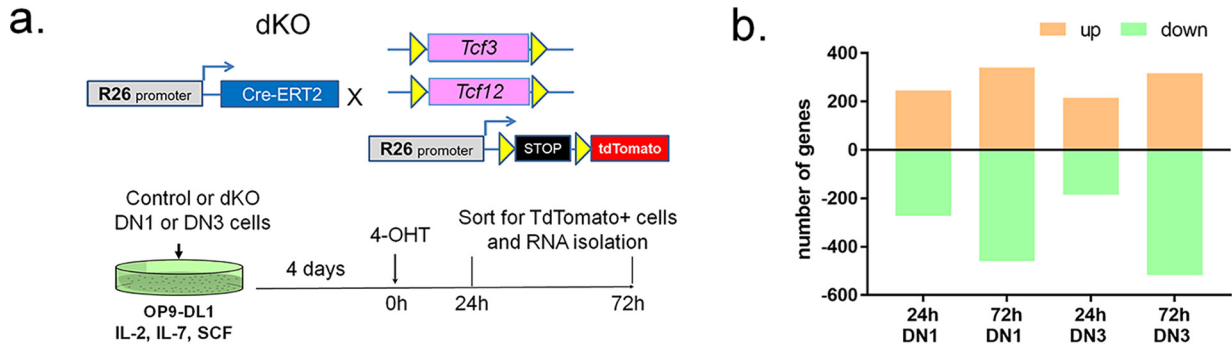
cytes, which are mostly cells arrested at the DN3 stage, suggesting that these genes are likely direct targets of E proteins. Although the overlap between our RNA-seq results and the ChIP-seq data are highly significant in a Fisher test, the small number of genes overlapped in the two datasets may be due to changes in gene expression that already occurred after a 4-day culture of DN3 cells on OP9-DL1 stroma before E protein deletion, such that their transcriptional profile is intrinsically different from that of *ex vivo* *Rag2^{-/-}* DN3 cells, which lack pre-TCR signaling.

To further verify our results, we examined the protein levels of several transcription factors whose transcripts were found to be altered by E protein ablation and to which suitable antibodies are available for flow cytometry (Fig. 1f). As we did for RNA-seq, we cultured DN3 cells from ROSA26-CreERT2;ROSA26-Stop-tdTomato;*Tcf3^{fl/fl}*;*Tcf12^{fl/fl}* (dKO) or mice expressing only tdTomato on OP9-DL1 stromal cells for 4 days and then added tamoxifen. 2 days later, cells were harvested and stained with antibodies against CD45.2, TCR β , TCR $\gamma\delta$, and ICOS along with transcription factors, GATA3, HELIOS, IRF4, and TCF1. At this time point, very few cells have differentiated into ILC2, which would express ICOS, even though E protein-deficient cells had more ICOS⁺ cells (Fig. 1f). These ICOS⁺ cells were excluded in our analyses to ensure a fair comparison. CD45.2⁺TCR β TCR $\gamma\delta$ ICOS⁺tdTomato⁺ cells were examined for levels of the transcription factors. Compared with tdTomato only controls, GATA3 and HELIOS expression is markedly increased in dKO cells, which is consistent with the elevation of their transcripts (Fig. 1f). However, the IRF4 level changed little even though its transcript was also augmented by E protein ablation. Perhaps, a different kinetics applied to the synthesis is the IRF4 protein. In contrast, the transcript for TCF1 (encoded by *Tcf7*) was found to be down-regulated upon E protein deletion, the protein level did not increase but showed a trend of reduction (Fig. 1f). We have also measured the levels of these proteins on day 3 post induction and similar results were obtained (data not shown). Together, these results largely agree with our transcriptome analyses.

Transcript levels altered by inducible E protein activity in ILC2s

We next took a complementary approach by examining gene expression in ILC2s after overexpression of an inducible E protein, E47-ER, in which E47 was fused to a modified hormone-binding domain of estrogen receptor (13). E47 activity can then be quickly induced by releasing the fusion protein from the cytoplasm to the nucleus upon binding to tamoxifen. We made use of our Id1 transgenic mice, in which a large amount of functional ILC2s are produced in the thymus (7). Thymic ILC2s were harvested from these mice by sorting for Lin⁻Thy1⁺ICOS⁺ST2⁺ cells and stimulated to proliferate with IL2, IL7, IL25, and IL33. These cells were then transduced with retroviruses expressing E47-ER and EGFP or EGFP alone at a relatively high efficiency (about 30%). Transduced cells were sorted for EGFP expression and expanded before treatments with tamoxifen for 4 and 16 h, respectively, and RNA-seq was performed (Fig. 2a). These short time points were chosen to detect

Transcriptional regulation by E proteins in ILC2s



immediate effects of E47 and to avoid any toxicity of E47 known to exist in B and T cells. Interestingly, ILC2s appeared to have tolerated E47 remarkably well within 16 h (data not shown).

At 4 and 16 h post induction, 436 and 596 genes were found to be differentially regulated, respectively (Fig. 2b, Table S2). In both cases, many more genes were found to be up-regulated by E47 than down-regulated. The differential genes can be divided into two sets: time-independent and time-dependent (Fig. 2c). The former includes genes whose expression changed statistically similarly at 24 and 72 h, which may be indicative of direct effects of E47. The latter consist of genes that exhibit slower kinetics in their activation or are indirectly affected by E47.

To narrow down the list of genes controlled by E proteins, we compared genes altered by loss of E proteins in either DN1 or DN3 cells (Fig. 1) to those affected by E47 overexpression in ILC2s, aiming to find genes altered in opposite directions. Genes that passed the threshold of combined \log_2 -fold change (LFC) > 1 and false discovery rate (FDR) < 0.05 were obtained (Fig. 2d, Table S3). Between the DN1 and DN3 sets, we found 80 protein-coding genes in common, which might be more relevant to the suppression of ILC2 differentiation (Fig. 2e). It is noted that the majority of the genes in this group are known to be direct targets of E proteins such as *Tcf12*, *Cdk20*, *Gfi1*, and *Ets1* (Fig. 1e). We reasoned that shortage of down-regulated genes found in this assay was because the 16-h window is too narrow to detect decreases in mRNA levels, which heavily depend on the repression of transcription by newly made repressors and turnover of existing transcripts.

Dynamic changes in chromatin accessibility upon inducible E47 expression

To better understand how E proteins inhibit ILC2 differentiation or function, we need to further examine the genes directly or indirectly down-regulated by overexpression of E47 in ILC2s. We thus determined the chromatin accessibility in ILC2s using the technique called Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) (14). We used retroviral transduced cells carrying E47-ER or a control vector treated as described in Fig. 2a for ATAC-seq.

Using a threshold of over 2-fold change with a FDR < 0.01, we found about 1300 peaks increased by E47 at 4 h but less than 100 peaks decreased (Fig. 3b; Table S4). However, by 16 h, 815 peaks were elevated by E47 but 4735 peaks were reduced, which suggest major changes in chromatin accessibility, resulting in transcriptional repression (Fig. 3b, Table S4). These changes correspond to less than 2% of all peaks detected, suggesting

that the assays were specific to E47 activity rather than nonspecific effects. Approximately 80% of differential peaks are associated with protein coding genes. Regarding genomic distribution, most peaks reside in the intronic and intergenic regions (Fig. 3c), consistent with the primary roles of E proteins and other transcription factors in regulating transcription through enhancers or super-enhancers.

The differential peaks obtained between E47-ER and control cells were then assigned to the nearest annotated gene. The numbers of up-regulated genes are 973 and 666 at 4 and 16 h post tamoxifen addition, respectively. Among them, 483 genes are common at both time points. The smaller number of up-regulated genes at 16 h might be attributed to negative feedback mechanisms. In contrast, the numbers of down-regulated genes were drastically different: 82 versus 2635 at 4 and 16 h, respectively. Most of the genes found to be repressed at 4 h remained down-regulated at 16 h. Because E47 is best known as a transcriptional activator, the profound increase in genes inhibited by E47 induction at 16 h suggests the induction of transcriptional repressors by E proteins.

Analyses of transcription factor-binding sites within differential peaks reveals that at 4 h, E boxes recognized by basic helix-loop-helix transcription factors such as E proteins were most abundant, representing over 50% of the binding sites found compared with a background of 18% in all open chromatin regions (Fig. 3d). This suggests the direct involvement of E47 in the increased chromatin accessibility observed (Fig. 3a). Interestingly, by 16 h, the predominant binding sites found in the differential peaks have shifted to the sites bound by bZip and GATA transcription factors. The known motif analysis revealed binding sites for different subgroups of bZip proteins including ATF3, BATE, JUN, FOS, BACH2, and MAF (Fig. S1). The large increase in the number of down-regulated regions (~4,735 peaks) may result from the secondary effects of the gene products initially elevated by E47 induction. These repressed motifs predominated over the E boxes directly associated with E47 binding. Indeed, when only elevated peaks were analyzed for their binding sites, E boxes remained the most prevalent (data not shown). Together, these analyses illustrated a dynamic change in chromatin accessibility within the 16 h after induction of E47 activity.

Genes initially activated by E47 may lead to repression of ILC2 genes later

We next correlated the ATAC-seq data with the RNA-seq results in EGFP or E47-ER/EGFP retrovirus-transduced ILC2s

Figure 1. Transcriptome changes immediately after E protein ablation. a, experimental design. DN1 and DN3 thymocytes from ROSA26-CreERT2; ROSA26-stop-tdTomato;E2A^{fl/fl};HEB^{fl/fl} (dKO) and ROSA26-CreERT2;ROSA26-stop-tdTomato (control) mice were co-cultured on OP9-DL1 stromal cells. Biological duplicates were collected 24 and 72 h later and sorted for tdTomato expression before RNA isolation and sequencing. b, numbers of differentially expressed genes comparing dKO and control at the indicated time points and cell preparations. Genes with a \log_2 -fold change > 1 and FDR < 0.05 were counted. c and d, Venn diagrams show the overlaps of genes up and down-regulated in DN1 and DN3-derived cells. Expression levels of genes commonly found in both DN1 and DN3-derived cells are shown in heat maps below each Venn diagram with representative genes listed. e, Venn diagram compares protein-coding genes differentially expressed in DN3-derived cells to genes assigned to anti-E protein (E47) ChIP-seq peaks detected in *ex vivo* Rag2^{-/-} thymocytes. Fisher test was used to determine the statistical significance of the overlap ($p < 0.00001$). Heat map shows the relative levels of expression in DN3-derived cells with representatives listed on the side. f, intracellular staining of transcription factors in DN3 cells from ROSA26-CreERT2;ROSA26-stop-tdTomato;E2A^{fl/fl};HEB^{fl/fl} (dKO) and plck-Cre;ROSA26-stop-tdTomato (control) mice were cultured as described in a. 2 days later, the cells were stained with antibodies against surface markers and the indicated transcription factors. Gating strategies are as illustrated. Data shown are representatives of cultures set up independently with two mice of each strain. Similar results were also obtained on day 3 after induction.

Transcriptional regulation by E proteins in ILC2s

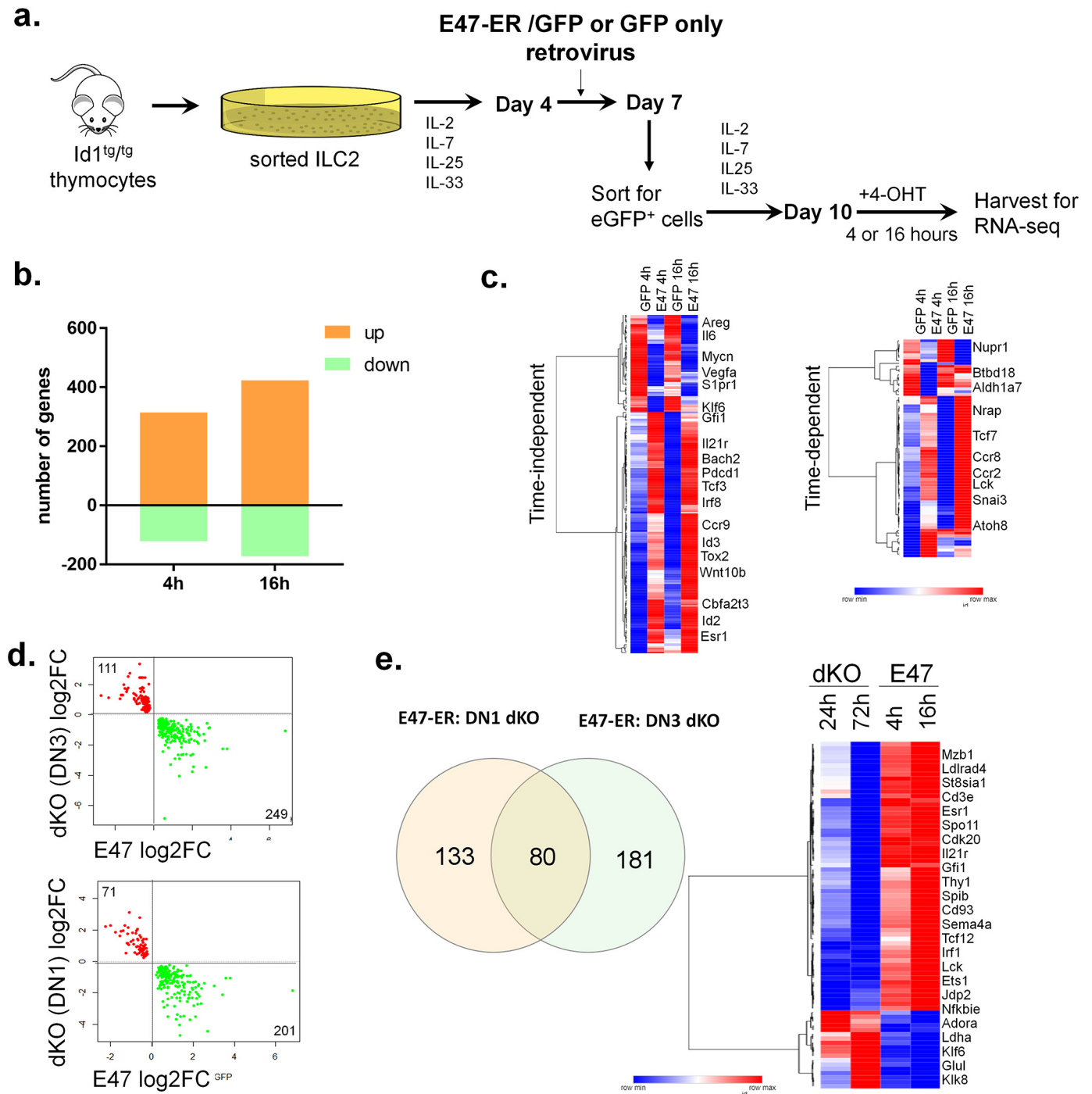


Figure 2. Transcriptome changes upon inducible expression of E47-ER. *a*, experimental scheme. *b*, numbers of differentially expressed genes comparing E47-ER/EGFP to EGFP-transduced cells at 4 and 16 h. Genes with LFC > 0.5 and FDR < 0.05 were counted. *c*, expression levels of genes altered similarly (time-independent) and differently (time-dependent) at 4 and 16 h are separately shown in heat maps with genes of interest listed. *d*, gene expression changed in opposite direction in E protein-deficient and E47-ER expressing cells are plotted (FDR < 0.05). *e*, Venn diagram shows the number of shared protein-coding genes in *d* between DN1 and DN3-derived cells. Relative levels of the 80 common genes are displayed in the heat map with representatives listed.

after the same treatments with tamoxifen. We looked for genes whose expression patterns changed in the same manner upon E47 induction. The thresholds of ATAC-seq data were set at a LFC > 1 and FDR < 0.01. Because we reasoned that changes in mRNA levels, particularly the reduction of transcripts, may be slower than alterations in chromatin accessibility, the threshold for the RNA-seq data were set at FDR < 0.05 without the LFC cutoff to maximize the detection of any trend of gene expres-

sion change (Fig. 4*a*). Four hours after E47 induction, the majority of the common differential genes are up-regulated. In contrast, by 16 h, the majority of genes were down-regulated. Additionally, we observed a significant fraction of genes that were increased as detected by RNA-seq but negatively regulated by ATAC-seq, which may illustrate the enhanced sensitivity of ATAC-seq in detection of dynamic gene repression at 16 h. The delayed appearance of down-regulated

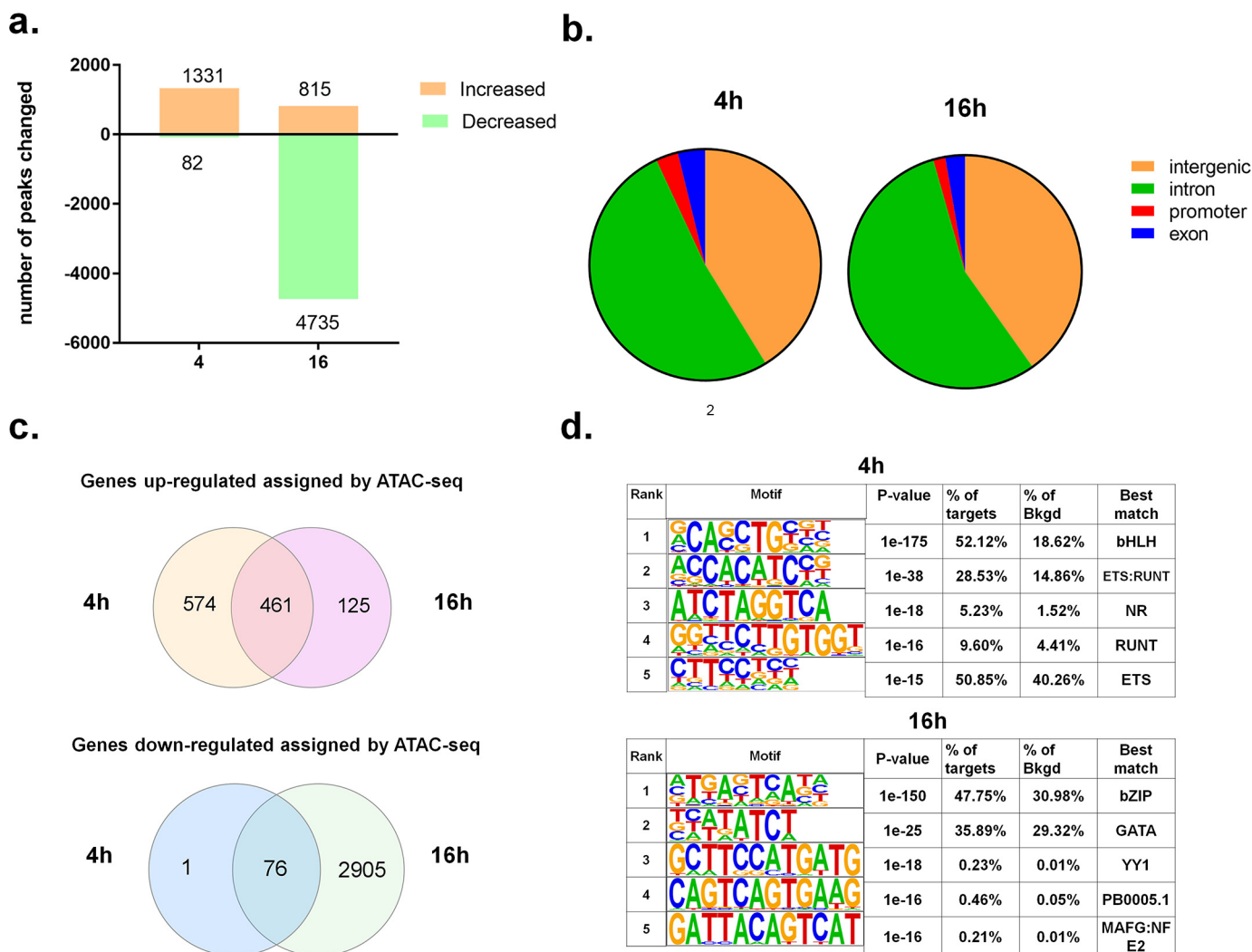


Figure 3. ATAC-seq analyses in E47-ER expressing cells. *a*, cells used in ATAC-seq analyses were treated as described in the legend to Fig. 2*a*. Numbers of peaks whose levels were increased or decreased compared with E47-ER expressing and control cells at 4 and 16 h are as indicated. The cutoffs for peak calling are LFC > 1 and FDR < 0.01. *b*, distribution of differential peaks across the genome at each time point. *c*, numbers of genes assigned to the peaks and those that overlap between the two time points. *d*, tables showing the top 5 most enriched transcription factor-binding motifs in differentially accessible peaks at 4 and 16 h. Binding sites that best match the transcription factor families are shown in the *right column*. NR stands for nuclear receptors.

genes supports our hypothesis that E47 indirectly causes gene repression.

We then set out to identify putative candidates that may mediate the profound chromatin remodeling observed at 16 h. Fig. 4*b* shows a list of more dramatically up-regulated genes at both 4 and 16 h. Among them, several transcriptional repressors stood out. The signal tracks of a few select genes are displayed in Fig. 4*c*. For example, expression of *Cbfa2t3*, which encodes MTG16 (also called ETO2), was significantly enhanced at 4 h and continued to increase by 16 h. The differential peaks were detected ~1, 3, and 13 kb upstream of the transcription start site. These peaks overlap with the region observed in ChIP against methylated K4 of histone 3 (H3K4me) in thymocytes (available at ENCODE, RRID:SCR_015482), suggesting that these regions may contain enhancers. MTG16 is a known transcriptional repressor capable of recruiting histone deacetylases (15, 16). Interestingly, *Cbfa2t3* has been shown to be expressed at very low levels in ILC2Ps compared with their precursors in the bone marrow (Fig. S2) (17), suggesting that

down-regulating this gene is necessary for ILC2 differentiation. ILC2s in small intestine also express *Cbfa2t3* at a low level. Furthermore, we found that 127 genes up-regulated in *Cbfa2t3*^{-/-} hematopoietic stem cells (LSK) (18) overlap with the genes assigned to diminished peaks at 16 h in ILC2s ($p < 2.2e-16$) (Fig. 4*d*). Considering the difference of the cell types in which the two sets of data were generated, the overlap is remarkable and supports the repressive role of *Cbfa2t3* products in ILC2s.

However, there are likely to be other repressors that mediate the suppressive function of E proteins in addition to MTG16. Jdp2 is a transcriptional repressor dimerizing with the Jun family of leucine-zipper transcription factors (19), many of which are expressed in ILC2s. Jdp2 can recruit histone deacetylase 3, thus regulating histone modification and chromatin assembly (20, 21). This is particularly relevant in our study because we observed widespread decreases in ATAC peaks at 16 h, and these peaks are predominantly populated with bZip transcription factor binding sites (Fig. 3*d*, Fig. S1). Thus, Jdp2 is a probable candidate for mediating gene repression at 16 h. Another

Transcriptional regulation by E proteins in ILC2s

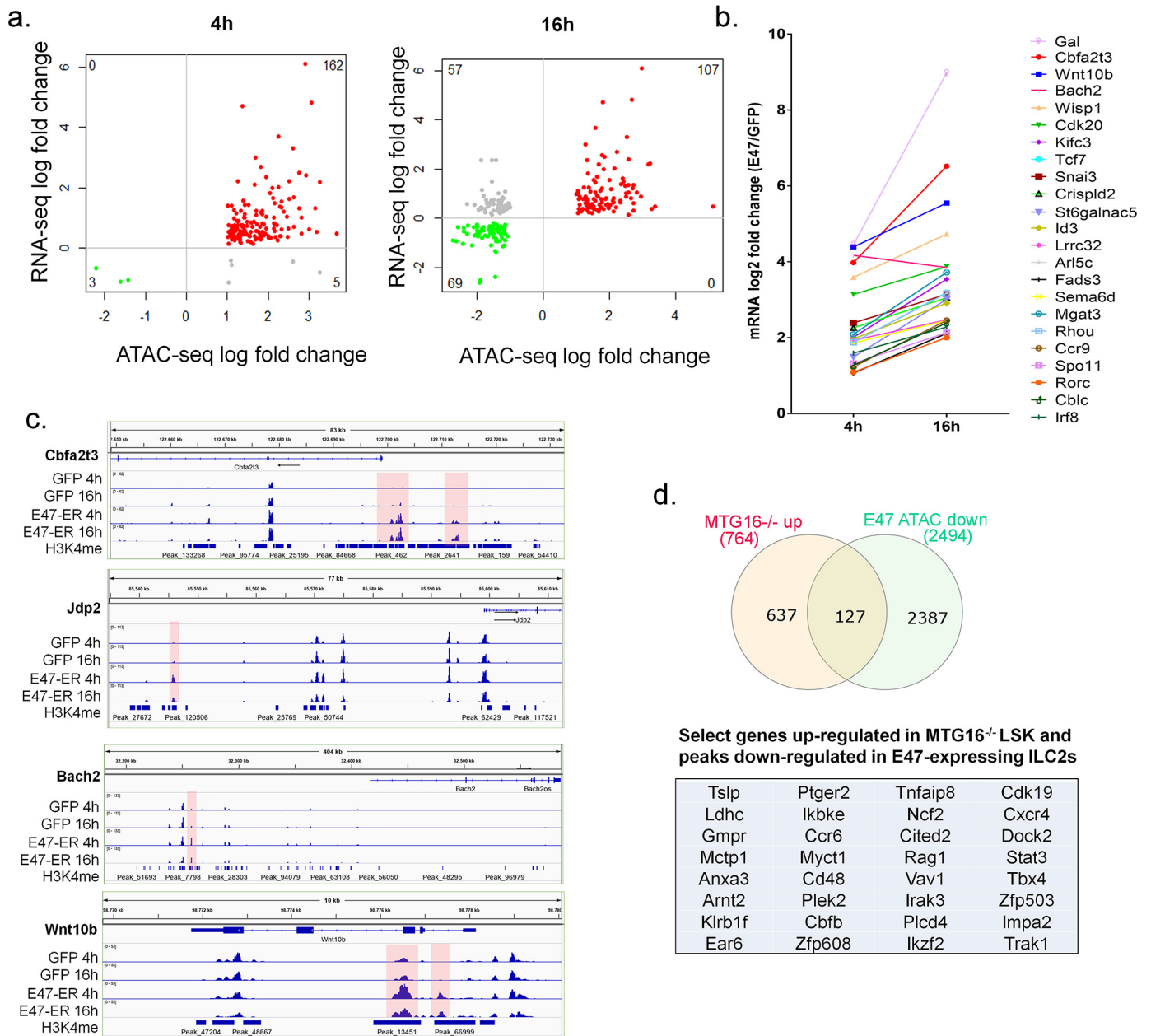


Figure 4. Initial transcriptional activation followed up widespread gene repression. *a*, correlation of mRNA levels and chromatin accessibility. LFC of ATAC peaks and mRNA levels in E47-ER expressing cells at each time point are plotted on x and y axis, respectively. *b*, top up-regulated genes at 4 and 16 h are plotted. *c*, ATAC peak tracks at the indicated time points and cells are shown for representative genes as indicated. H3K4me ChIP-seq data were generated using thymocytes by the ENCODE consortium. *d*, comparison of genes repressed by MTG16 (encoded by *Cbfa2t3*) in hematopoietic stem cells (LSK) and genes repressed in E47-ER expressing ILC2s at 16 h. Fisher test shows the statistical significance of the overlap ($p < 0.00001$). A select list of the genes are shown below the Venn diagram.

up-regulated bZip transcription repressor is Bach2 (22). Its binding sites are readily detectable in ATAC peaks at 16 h (Fig. S1). Considering the inhibition of Th2 differentiation by Bach2 (23, 24), it could also contribute to the suppression of the ILC2 fate.

Finally, several genes encoding proteins related to the Wnt signaling pathway, *Wnt10b*, *Wisp*, and *Snai*, are on top of the list of up-regulated genes shown in Fig. 4b. Given the crucial roles of Wnt signaling in regulating various developmental programs (25, 26), it may play an important role in ILC2 suppression.

Down-regulation of a network of transcriptional programs for ILC2 differentiation by E47

The large amount of ATAC peaks found to be reduced by E47 at 16 h may provide plausible hints as to how E proteins suppress ILC2 differentiation. We thus performed pathway analysis on the genes regulated in similar manners at 16 h as detected using RNA-seq and ATAC-seq (Fig. 4a). Consistent with our expectations, the top canonical pathway found to be down-regulated by E47 is related to T helper 2 cells. The top-ranking network of transcriptional programs includes key transcription factors and signaling molecules known to be ILC

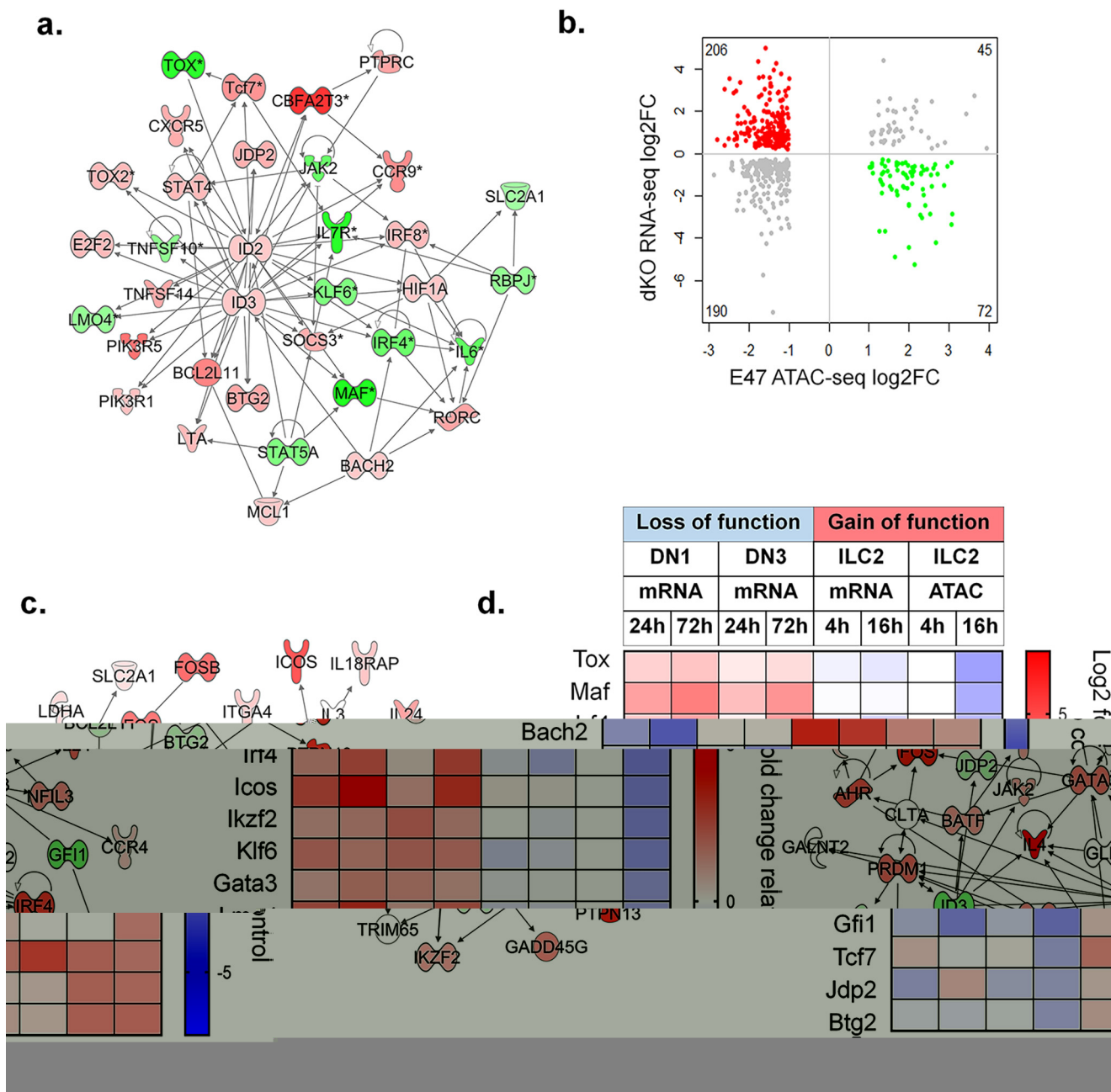


Figure 5. Loss and gain of E protein function impact a similar transcription program regulating ILC2 differentiation. *a*, ingenuity pathway analysis of differential peaks detected at 16 h after E47-ER induction. The top transcription network (score = 45) is shown. Up- and down-regulated genes are depicted in red and green. Multiple peaks assigned to the same gene are indicated with an asterisk. Arrows show connections between two genes, mostly by positive or negative influence in expression. *b*, inverse correlation between the effects of loss and gain of E protein function. LFC of differentially expressed genes in both DN1- and DN3-derived cells is plotted against the changes in E47 expressing cells at 16 h. Genes increased and decreased in E protein-deficient cells are highlighted in red and green, respectively. *c*, top network identified using the ingenuity pathway analysis of inversely correlated genes as shown in *b* (score = 38). *d*, summary of the expression profiles of key genes regulated by E proteins in relationship to ILC2 differentiation.

signatures such as IL-7R, TOX, IRF4, MAF, KLF6, and RBPJ (Fig. 5a) (4, 5, 17, 27). The reduction in chromatin accessibility of the key ILC2 signature genes is striking. For example, 15 peaks assigned to *Maf* were found to be diminished by at least 2-fold (Fig. S3). Likewise, 3 peaks each were reduced for *Tox* and *Irf4*, respectively.

Id2 and *Id3* were up-regulated by E47 in the treated cells because these genes are known to be activated by E proteins in a feedback loop that helps maintain the net activity of E pro-

teins (28). Because E47-ER overexpressed in the cells would likely outcompete endogenous Id proteins, the small increases in *Id2* and *Id3* mRNA levels may not be consequential. However, the major nodes formed around *Id2* and *Id3* highlight the importance of the helix-loop-helix family of transcription factors such as E47 in regulating ILC2 differentiation and their interplay with other transcription factors (Fig. 5a).

To correlate the ATAC-seq data with gene expression changes caused by E protein ablation in both DN1- and DN3-

Transcriptional regulation by E proteins in ILC2s

initiated ILC2 cultures (Fig. 1), we plotted fold-changes in mRNA levels and ATAC-seq peaks (Fig. 5b). We focused on the inverse relationship between the two datasets. There are 206 and 72 genes up- and down-regulated upon E protein gene deletion, which are assigned to peaks decreased and increased in ATAC-seq, respectively. Ingenuity pathway analysis of these commonly regulated genes revealed the top regulatory network consisting of key regulatory proteins in ILC2 differentiation. The same molecules found to be down-regulated by E47 expression in ILC2s, IRF4, MAF, and KLF6, are shown to be up-regulated during ILC2 differentiation of E protein-deficient DN1 and DN3 thymocytes (Fig. 5, a and c). Conversely, genes activated by E47 are diminished with E protein ablation, which include *Id3*, *Jdp2*, *Jak2*, *Bcl2l11*, and *Btg2*. In addition, several more genes encoding transcription factors potentially important for ILC2 differentiation, such as *Gata3*, *Batf*, *Nfil3*, *Ahr*, *Irf2*, *Prdm1*, *Fos*, and *Fosb* were also deemed to be up-regulated in E protein-deficient DN1 and DN3 cells and down-regulated by E47 as detected by ATAC-seq.

Discussion

This study uses inducible loss and gain of E protein function models to cross-examine transcription programs in the context of ILC2 differentiating cells or ILC2s and the consensus is summarized in Fig. 5d. By following changes in mRNA levels or chromatin accessibility shortly after alterations in E protein activity, we have a unique opportunity to postulate a cascade of regulatory events that lead to E protein-mediated suppression of ILC2 differentiation. Induction of E47 in ILC2s quickly led to the up-regulation of repressor genes, which in turn cause coordinated transcriptional repression of a network of transcriptional programs important for the ILC2 cell fate. In particular, *Cbfa2t3*, encoding MTG16 (or ETO2), is dramatically up-regulated by E47 and it is also highly expressed in T cell precursors. MTG16 not only interacts with E proteins but also forms super complexes with GATA1, E proteins, Tal1, Lmo2, and Ldb in erythroid cells (29, 30). MTG16 is also important for T cell development, as well as the differentiation of plasmacytoid dendritic cells (31, 32). It is thus conceivable that E proteins recruit MTG16 and cooperate with GATA3 and Lmo2/4 to suppress the ILC2 fate in bone marrow and T cell precursors where these proteins all exist. Equally plausible is the action of negative regulators of a large class of basic leucine zipper transcription factors including products of *Jdp2* and *Bach2*. The enrichment of bZip transcription factor-binding sites in the ATAC peaks decreased at the 16-h time point would argue for the importance of this class of transcription factors in controlling the genes necessary for ILC2 identity and function. Therefore, it is likely that the concerted efforts of multiple axes of positive and negative regulation of transcription dictate the developmental trajectory of progenitor cells of either innate or adaptive lymphoid cells.

E proteins are known to promote B and T lymphoid cell commitment, partly through their ability to facilitate the production of proteins involved in antigen receptor signaling (33). However, simply pushing the B and T cell-specific function may not be sufficient and suppression of other cell fates, such

as the myeloid and innate lymphoid paths could also be necessary (34, 35). Conversely, the differentiation into ILC2 or other ILCs requires the inhibition of E protein function by their inhibitors, the Id proteins. Id2 is well known to be essential of ILC differentiation (36). Id3 is known to be transiently up-regulated by pre-T cell receptor signaling (37). Therefore, Id2 and Id3 may act by antagonizing the E protein activities we have elucidated in this study to promote ILC differentiation from common lymphoid progenitors in the bone marrow or T cell precursors in the thymus.

Experimental procedures

Mice and OP9 stromal culture

ROSA26-CreERT2;ROSA26-Stop-tdTomato;*Tcf3*^{fl/fl};*Tcf12*^{fl/fl} (dKO) mice and controls lacking to the floxed *Tcf3* and *Tcf12* are as described (5, 7). CLP from the bone marrow and DN1 or DN3 cells from the thymus were sorted and cultured as described (5). Briefly, cells were seeded at 100 or 1000 cells/well in 48-well-plates containing a layer of OP9-DL1 stromal cells plated 24 h before. The culture medium contains 20% bovine fetal calf serum, 30 ng/ml of each IL-2, IL-7, and stem cell factor (all from R&D systems) in α -minimal essential medium (Gibco, ThermoFisher). On day 4, 4-hydroxy-tamoxifen was added to a concentration of 1 μ M. Cells were harvested 24 and 72 h later and CD45⁺tdTomato⁺ were sorted and used for RNA-seq. For intracellular staining of transcription factors, cells were analyzed 48 h after addition of tamoxifen.

Retroviral transduction in ILC2s

To purify ILC2s from the thymus of plck-Id1^{tg/tg} mice, thymocytes were stained with anti-CD90.2 (53-2.1), anti-ST2 (DIH9), anti-ICOS (C398.4A), and a mixture of lineage (Lin)-specific antibodies: anti-Fc ϵ R (MAR-1), anti-B220 (RA3-6B2), anti-CD19 (ID3), anti-Mac-1 (M1/70), anti-Gr-1 (R86-8C5), anti-CD11c (N418), anti-NK1.1 (PK136), anti-Ter-119 (Ter-119), anti-CD3 ϵ (145-2C11), anti-CD5 (53-7.3), anti-CD8 α (53-6.7), anti-TCR β (H57-597), and anti- γ TCR (GL-3; eBioscience). All antibodies are from Biologend unless specified otherwise. Lin-CD90.2⁺ICOS⁺ST2⁺ cells were sorted on a FACSAria II (BD Biosciences). The cells were plated on 96-well U-bottom plates at 80,000 cells/well in α -minimal essential medium containing 10% fetal calf serum, 10 ng/ml each IL-2, IL-7, IL-25, and IL-33. Cells were split 1:3 on day 3 and placed onto 96-well flat-bottom plates and spin-infected with retrovirus the next day. Spin-infection was performed by mixing retroviral stocks and cultured cells at a 1:1 ratio plus the addition of 8 μ g/ml of Polybrene. The retroviral stocks are as described (13). The cells were centrifuged for 1.5 h at 1,300 rpm before the plates were returned to the 37 °C incubator with 5% CO₂. After 6 h, the culture was replaced with fresh medium and incubated for 2 days. Transduced cells were sorted for the expression of EGFP and cultured for 3 additional days in U-bottom plates with the same medium. These cells with or without E47-ER were induced with 1 μ M 4-hydroxy-tamoxifen for 4 or 16 h and harvested.

RNA and ATAC sequencing

RNA was isolated using the TRIzol reagent per the vendor's instructions (Life Technology). RNA-seq was performed in duplicate by the Clinical Genomics Center at the Oklahoma Medical Research Foundation using the Ovation RNA-Seq System (NuGEN Technologies Inc., San Carlos, CA) followed by KAPA Hyper library preparation kits (KAPA Biosystems, Wilmington, MA) on an Illumina NextSeq 500 sequencing platform with 76-bp paired-end reads (Illumina, San Diego, CA).

ATAC-seq was also performed in duplicate for each time point and cell type. ATAC transpose reactions were carried out using the NX#-TDE1, TAGMENT DNA enzyme, and buffer from Illumina. The reaction was purified using Qiagen miniElute columns, which is followed by PCR amplification with matching primers. The PCR products were purified using magnetic beads and eluted in EB buffer (Qiagen). The products were sequenced using an Illumina NextSeq 500 sequencing platform with 76-bp paired-end reads.

Analyses of RNA-seq data

Raw sequencing reads (in a FASTQ format) were trimmed using Trimmomatic to remove any low-quality bases at the beginning or the end of sequencing reads as well as any adapter sequences. Trimmed sequencing reads were aligned to the *Mus musculus* genome reference (GRCm38/mm10) using STAR version 2.4.0h (38). Gene-level read counts were determined using HTSeq version 0.5.3p9 (39) with the GENCODE Release M10 (GRCm38) annotation. Only autosomal genes coding for lncRNAs, miRNAs, and protein-coding mRNAs were selected for analyses.

Read-count normalization and differentially expressed analyses was performed using the edgeR package from Bioconductor. Expression values normalized with the voom function were analyzed for differential expression using the standard functions of the limma program. Each transcript expression variation was tested with the moderated *t*-statistics and the corresponding *p* values were adjusted for multiple testing using FDR. Unless specified, sets of differentially expressed transcripts were filtered requiring at least 2-fold change in expression and a FDR below 0.05. Overrepresented functional gene sets (GO, KEGG pathways) were identified using specialized R Bioconductor packages. Ingenuity Pathway Analysis (Qiagen) was used to explore significant gene networks and pathways interactively.

ATAC data processing and analysis

Alignment and peak-calling of ATAC sequencing data were performed using the ENCODE-DCC ATAC-Seq pipeline (<https://github.com/ENCODE-DCC/atac-seq-pipeline>) (40). Briefly, data were aligned to mm10 genome using Bowtie2 (41). Aligned data were filtered for PCR duplicates, blacklisted regions, and shifted +4 bp for the plus (+) strand and -5 bp for the minus (-) strand. Peaks were called using MACS2 (42) and filtered by irreproducible discovery rate (IDR < 0.1) (<https://github.com/nboley/idr>). Differential peak calling was performed using differential binding analysis of ChIP-seq peak data (DiffBind, <http://bioconductor.org/>

[packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf](https://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf)) (43), where a consensus peak set of 60,543 peaks was derived from all samples, and a window of 150 bp was centered on peak summits for further downstream analysis. DESeq2 (44) was used to perform differential peak analysis with DE peaks identified by $q < 0.05$ and absolute log fold-change > 1. Motif enrichment was performed using Hypergeometric Optimization of Motif Enrichment program (HOMER) version 4.1 (45), where the background was set as the union of all ATAC peaks.

Data availability

Raw data for RNA-seq and ATAC-seq can be found in the GEO database GSE111518 for E protein-deficient cells on OP9-DL1 culture; GSE151739 for E47-ER transduced ILC2s. Other data are presented in the manuscript including [supporting material](#).

Acknowledgments—We are grateful to Dr. Graham Wiley for advice in the ATAC-seq experiments and the Clinical Genomics Center and Flow Cytometry Facility at the Oklahoma Medical Research Foundation for outstanding technical assistance. We thank Dr. Marco Colonna for advice and support.

Author contributions—V. P. and C. G. carried out bioinformatics analyses; A. B., A. P., L. Q., and X.-H. S. generated the data; V. P., C. G., J. D. W. and X.-H. S. analyzed the data and wrote the manuscript.

Funding and additional information—This work was supported by National Institutes of Health Grant 1R01AI126851 and the Presbyterian Health Foundation (to X.-H.S.), and National Institutes of Health Grants 1P20 GM103636-07 (to J. D. W. and X.-H. S.) and T32DK077653-28 (to V. P.). X.-H.S. holds the Lew and Mira Ward Chair in Biomedical Research at the Oklahoma Medical Research Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: ILC2s, group 2 innate lymphoid cells; CLP, common lymphoid progenitor; dKO, double knockout; LFC, log₂-fold change; FDR, false discovery rate; ATAC, assay for transposase-accessible chromatin using sequencing; IL, interleukin; EGFP, epidermal growth factor protein.

References

1. Vivier, E., Artis, D., Colonna, M., Diefenbach, A., Di Santo, J. P., Eberl, G., Koyasu, S., Locksley, R. M., McKenzie, A. N. J., Mebius, R. E., Powrie, F., and Spits, H. (2018) Innate lymphoid cells: 10 years on. *Cell* **174**, 1054–1066 [CrossRef Medline](#)
2. Schneider, C., Lee, J., Koga, S., Ricardo-Gonzalez, R. R., Nussbaum, J. C., Smith, L. K., Villeda, S. A., Liang, H. E., and Locksley, R. M. (2019) Tissue-resident group 2 innate lymphoid cells differentiate by layered ontogeny and *in situ* perinatal priming. *Immunity* **50**, 1425–1438 [CrossRef Medline](#)
3. Artis, D., and Spits, H. (2015) The biology of innate lymphoid cells. *Nature* **517**, 293–301 [CrossRef Medline](#)

Transcriptional regulation by E proteins in ILC2s

- Ricardo-Gonzalez, R. R., Van Dyken, S. J., Schneider, C., Lee, J., Nussbaum, J. C., Liang, H. E., Vaka, D., Eckalbar, W. L., Molofsky, A. B., Erle, D. J., and Locksley, R. M. (2018) Tissue signals imprint ILC2 identity with anticipatory function. *Nat. Immunol.* **19**, 1093–1099 [CrossRef Medline](#)
- Qian, L., Bajana, S., Georgescu, C., Peng, V., Wang, H. C., Adrianto, I., Colonna, M., Alberola-Ila, J., Wren, J. D., and Sun, X. H. (2019) Suppression of ILC2 differentiation from committed T cell precursors by E protein transcription factors. *J. Exp. Med.* **216**, 884–899 [CrossRef Medline](#)
- Yang, Q., and Bhandoola, A. (2016) The development of adult innate lymphoid cells. *Curr. Opin. Immunol.* **39**, 114–120 [CrossRef Medline](#)
- Wang, H. C., Qian, L., Zhao, Y., Mengarelli, J., Adrianto, I., Montgomery, C. G., Urban, J. F., Jr., Fung, K. M., and Sun, X. H. (2017) Downregulation of E protein activity augments an ILC2 differentiation program in the thymus. *J. Immunol.* **198**, 3149–3156 [CrossRef Medline](#)
- Miyazaki, M., Miyazaki, K., Chen, K., Jin, Y., Turner, J., Moore, A. J., Saito, R., Yoshida, K., Ogawa, S., Rodewald, H. R., Lin, Y. C., Kawamoto, H., and Murre, C. (2017) The E-Id protein axis specifies adaptive lymphoid cell identity and suppresses thymic innate Lymphoid cell development. *Immunity* **46**, 818–834 [CrossRef Medline](#)
- Murre, C. (2005) Helix-loop-helix proteins and lymphocyte development. *Nat. Immunol.* **6**, 1079–1086 [CrossRef Medline](#)
- Ling, F., Kang, B., and Sun, X. H. (2014) Id proteins: small molecules, mighty regulators. *Curr. Top. Dev. Biol.* **110**, 189–216 [CrossRef Medline](#)
- Wojciechowski, J., Lai, A., Kondo, M., and Zhuang, Y. (2007) E2A and HEB are required to block thymocyte proliferation prior to pre-TCR expression. *J. Immunol.* **178**, 5717–5726 [CrossRef Medline](#)
- Miyazaki, M., Rivera, R. R., Miyazaki, K., Lin, Y. C., Agata, Y., and Murre, C. (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**, 992–1001 [CrossRef Medline](#)
- Yang, Y., Wang, H. C., and Sun, X. H. (2008) Id1 induces apoptosis through inhibition of ROR γ t expression. *BMC. Immunol.* **9**, 20 [CrossRef Medline](#)
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 [CrossRef Medline](#)
- Hoogeveen, A. T., Rossetti, S., Stoyanova, V., Schonkeren, J., Fenaroli, A., Schiaffonati, L., van, U. L., and Sacchi, N. (2002) The transcriptional corepressor MTG16a contains a novel nucleolar targeting sequence deranged in t(16;21)-positive myeloid malignancies. *Oncogene* **21**, 6703–6712 [CrossRef Medline](#)
- Amann, J. M., Nip, J., Strom, D. K., Lutterbach, B., Harada, H., Lenny, N., Downing, J. R., Meyers, S., and Hiebert, S. W. (2001) ETO, a target of t(8;21) in acute leukemia, makes distinct contacts with multiple histone deacetylases and binds mSin3A through its oligomerization domain. *Mol. Cell Biol.* **21**, 6470–6483 [CrossRef Medline](#)
- Harly, C., Kenney, D., Ren, G., Lai, B., Raabe, T., Yang, Q., Cam, M. C., Xue, H. H., Zhao, K., and Bhandoola, A. (2019) The transcription factor TCF-1 enforces commitment to the innate lymphoid cell lineage. *Nat. Immunol.* **20**, 1150–1160 [CrossRef Medline](#)
- Chyla, B. J., Moreno-Miralles, I., Steapleton, M. A., Thompson, M. A., Bhaskara, S., Engel, M., and Hiebert, S. W. (2008) Deletion of Mtg16, a target of t(16;21), alters hematopoietic progenitor cell proliferation and lineage allocation. *Mol. Cell Biol.* **28**, 6234–6247 [CrossRef Medline](#)
- Aronheim, A., Zandi, E., Hennemann, H., Elledge, S. J., and Karin, M. (1997) Isolation of an AP-1 repressor by a novel method for detecting protein-protein interactions. *Mol. Cell Biol.* **17**, 3094–3102 [CrossRef Medline](#)
- Jin, C., Li, H., Murata, T., Sun, K., Horikoshi, M., Chiu, R., and Yokoyama, K. K. (2002) JDP2, a repressor of AP-1, recruits a histone deacetylase 3 complex to inhibit the retinoic acid-induced differentiation of F9 cells. *Mol. Cell Biol.* **22**, 4815–4826 [CrossRef Medline](#)
- Jonak, C. R., Lainez, N. M., Roybal, L. L., Williamson, A. D., and Coss, D. (2017) c-JUN dimerization protein 2 (JDP2) is a transcriptional repressor of follicle-stimulating hormone β (FSH β) and is required for preventing premature reproductive senescence in female mice. *J. Biol. Chem.* **292**, 2646–2659 [CrossRef Medline](#)
- Muto, A., Hoshino, H., Madisen, L., Yanai, N., Obinata, M., Karasuyama, H., Hayashi, N., Nakauchi, H., Yamamoto, M., Groudine, M., and Igarashi, K. (1998) Identification of Bach2 as a B-cell-specific partner for small maf proteins that negatively regulate the immunoglobulin heavy chain gene 3' enhancer. *EMBO J.* **17**, 5734–5743 [CrossRef Medline](#)
- Kuwahara, M., Ise, W., Ochi, M., Suzuki, J., Kometani, K., Maruyama, S., Izumoto, M., Matsumoto, A., Takemori, N., Takemori, A., Shinoda, K., Nakayama, T., Ohara, O., Yasukawa, M., Sawasaki, T., et al. (2016) Bach2-Batf interactions control Th2-type immune response by regulating the IL-4 amplification loop. *Nat. Commun.* **7**, 12596 [CrossRef Medline](#)
- Yamashita, M., and Kuwahara, M. (2018) The critical role of Bach2 in regulating type 2 chronic airway inflammation. *Int. Immunol.* **30**, 397–402 [CrossRef Medline](#)
- Grigoryan, T., Wend, P., Klaus, A., and Birchmeier, W. (2008) Deciphering the function of canonical Wnt signals in development and disease: conditional loss- and gain-of-function mutations of β -catenin in mice. *Genes Dev.* **22**, 2308–2341 [CrossRef Medline](#)
- Lento, W., Congdon, K., Voermans, C., Kritzik, M., and Reya, T. (2013) Wnt signaling in normal and malignant hematopoiesis. *Cold Spring Harb. Perspect. Biol.* **5**, a100801 [CrossRef](#)
- Schwartz, C., Khan, A. R., Floudas, A., Saunders, S. P., Hams, E., Rodewald, H. R., McKenzie, A. N. J., and Fallon, P. G. (2017) ILC2s regulate adaptive Th2 cell functions via PD-L1 checkpoint control. *J. Exp. Med.* **214**, 2507–2521 [CrossRef Medline](#)
- Neuman, K., Nornes, H. O., and Neuman, T. (1995) Helix-loop-helix transcription factors regulate Id2 gene promoter activity. *FEBS Lett.* **374**, 279–283 [CrossRef](#)
- Meier, N., Krpic, S., Rodriguez, P., Strouboulis, J., Monti, M., Krijgsveld, J., Gering, M., Patient, R., Hostert, A., and Grosveld, F. (2006) Novel binding partners of Ldb1 are required for hematopoietic development. *Development* **133**, 4913–4923 [CrossRef Medline](#)
- Tropic, T., Deng, W., Cheng, Y., Zhang, Y., Vakoc, C. R., Gregory, G. D., Hardison, R. C., and Blobel, G. A. (2009) SCL and associated proteins distinguish active from repressive GATA transcription factor complexes. *Blood* **113**, 2191–2201 [CrossRef Medline](#)
- Hunt, A., Fischer, M., Engel, M. E., and Hiebert, S. W. (2011) Mtg16/Eto2 contributes to murine T-cell development. *Mol. Cell Biol.* **31**, 2544–2551 [CrossRef Medline](#)
- Ghosh, H. S., Ceribelli, M., Matos, I., Lazarovici, A., Bussemaker, H. J., Lasorella, A., Hiebert, S. W., Liu, K., Staudt, L. M., and Reizis, B. (2014) ETO family protein Mtg16 regulates the balance of dendritic cell subsets by repressing Id2. *J. Exp. Med.* **211**, 1623–1635 [CrossRef Medline](#)
- Engel, I., and Murre, C. (2001) The function of E- and Id proteins in lymphocyte development. *Nat. Rev. Immunol.* **1**, 193–199 [CrossRef Medline](#)
- Cochrane, S. W., Zhao, Y., Welner, R. S., and Sun, X. H. (2009) Balance between Id and E proteins regulates myeloid-versus-lymphoid lineage decisions. *Blood* **113**, 1016–1026 [CrossRef Medline](#)
- Dias, S., Mansson, R., Gurbuxani, S., Sigvardsson, M., and Kee, B. L. (2008) E2A proteins promote development of lymphoid-primed multipotent progenitors. *Immunity* **29**, 217–227 [CrossRef Medline](#)
- Klose, C. S., Flach, M., Mohle, L., Rogell, L., Hoyler, T., Ebert, K., Fabianke, C., Pfeifer, D., Sexl, V., Fonseca-Pereira, D., Domingues, R. G., Veiga-Fernandes, H., Arnold, S. J., Busslinger, M., Dunay, I. R., et al. (2014) Differentiation of type 1 ILCs from a common progenitor to all helper-like innate lymphoid cell lineages. *Cell* **157**, 340–356 [CrossRef Medline](#)
- Bain, G., Cravatt, C. B., Loomans, C., Alberola-Ila, J., Hedrick, S. M., and Murre, C. (2001) Regulation of the helix-loop-helix proteins, E2A and Id3, by the Ras-ERK MAPK cascade. *Nat. Immunol.* **2**, 165–171 [CrossRef Medline](#)
- 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 [CrossRef Medline](#)
- Anders, S., Pyl, P. T., and Huber, W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 [CrossRef Medline](#)
- Encode Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 [CrossRef Medline](#)

41. Langmead, B., and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 [CrossRef Medline](#)
42. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 [CrossRef Medline](#)
43. Ross-Innes, C. S., Stark, R., Teschendorff, A. E., Holmes, K. A., Ali, H. R., Dunning, M. J., Brown, G. D., Gojis, O., Ellis, I. O., Green, A. R., Ali, S., Chin, S. F., Palmieri, C., Caldas, C., and Carroll, J. S. (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 [CrossRef Medline](#)
44. Love, M. I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 [CrossRef Medline](#)
45. 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. *Mol. Cell* **38**, 576–589 [CrossRef Medline](#)