Title: Id proteins suppress E2A-driven innate-like T cell development prior to TCR selection

Running Title: E2A drives innate-like transcriptional network in developing T cells

Authors: Sumedha Roy*, Amanda J. Moore‡, Cassandra Love§, Anupama Reddy§, Deepthi Rajagopalan§, Sandeep S. Dave§, Leping Li†, Cornelis Murre‡, Yuan Zhuang*,#

Affiliations:
*Department of Immunology, Duke University Medical Center, Durham, NC 27710
‡Department of Molecular Biology, University of California, San Diego, La Jolla, California, 92093
§Duke Institute for Genome Sciences and Policy, Duke University, Durham, NC 27710
†Biostatistics and Computational Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC 27709

#Corresponding author: Department of Immunology, Duke University Medical Center, Durham, NC 27710; yzhuang@duke.edu; Phone (919) 613-7824, Fax (919) 613-7853

Keywords
Thymus, iNKT cells, innate-like T cells, ID proteins, mouse
Id proteins have been shown to promote the differentiation of conventional αβ and γδT cells, and to suppress the expansion of invariant Natural Killer T (iNKT) cells and innate-like γδNKT within their respective cell lineages. However, it remains to be determined whether Id proteins regulate lineage specification in developing T cells that give rise to these distinct cell fates. Here we report that in the absence of Id2 and Id3 proteins, E2A prematurely activates genes critical for the iNKT cell lineage prior to TCR expression. Enhanced iNKT development in Id3-deficient mice lacking γδNKT cells suggests that Id3 regulates the lineage competition between these populations. RNA-Seq analysis establishes E2A as the transcriptional regulator of both iNKT and γδNKT development. In the absence of pre-TCR signaling, Id2/Id3 deletion gives rise to a large population of iNKT cells and a unique innate-like DP population, despite the block in conventional αβ T cell development. The transcriptional profile of these unique DP cells reflects enrichment of innate-like signature genes, including PLZF (Zbtb16) and Granzyme A (Gzma). Results from these genetic models and genome-wide analyses suggest that Id proteins suppress E2A-driven innate-like T cell programs prior to TCR selection to enforce predominance of conventional T cells.
Introduction

The thymic output of a diverse and abundant population of conventional CD4$^+$ and CD8$^+$ $\alpha\beta$ T cells constitutes the adaptive immune system that is necessary for a specific and effective immune response to antigens. A smaller but significant population of unconventional T cells also concomitantly develops in the thymus, with innate-like capabilities of mounting a non-specific but rapid immune response to antigens. These innate-like T cells have garnered increasing interest in lieu of their memory phenotype that can be harnessed in the context of allergies, infections and tumors. They often possess limited TCR diversity that allows recognition of self-ligands presented by non-classical MHC molecules such as CD1 and MR1.\cite{1} They are also characterized by high levels of expression of the innate-like transcription factor, Promyelocytic Zinc Finger (PLZF).\cite{2-6} While the transcriptional programs that drive conventional CD4$^+$ and CD8$^+$ T cell specification and development have been well characterized, little is known about the innate-specific transcriptional programs upstream of PLZF that are responsible for the divergence of innate-like T cells from conventional T cells.

Innate-like T cell populations include TCR$\alpha\beta^+$ Natural Killer T (NKT) cells, TCR$\gamma\delta^+$ NKT cells, innate-like CD8$^+$ T cells, CD8$\alpha\alpha$ intraepithelial lymphocytes (IELs), and Mucosal-associated invariant T (MAIT) cells. Invariant NKT (iNKT) cells are among the best characterized innate-like T cells, which arise in parallel with conventional $\alpha\beta$ T cells. These cells are known to stochastically express a canonical $\upsilon\alpha$14-$\upsilon\alpha$18 TCR$\alpha$ chain at the CD4$^+$CD8$^+$ double positive (DP) stage, which allows them to undergo TCR selection mediated by a CD1d molecule expressed on other conventional DP thymocytes.\cite{7,8} $\gamma\delta$NKT cells are yet another population of innate-like T cells that express a restricted $\upsilon\gamma$1.1$\upsilon\delta$6.3 TCR. Both iNKT and $\gamma\delta$NKT cells express high levels of PLZF and readily produce effector cytokines like IL-4.\cite{2,9}
Id proteins, primarily produced by *Id2* and *Id3* during T cell development, are inhibitors of the E protein transcription factors E2A and HEB. Interestingly, Id proteins play opposite roles in the development of conventional and innate-like T cells, such that they promote the former and suppress the latter. In response to pre-TCR and TCR signals, Id proteins inhibit E protein activity to play a critical role in promoting the differentiation and positive selection of conventional αβ T cells.\[10-12\] The Id-mediated inhibition is both necessary and sufficient to activate gene expression programs critical for T cell maturation. In line with this, it has been shown that disruption of *Id2* and *Id3* impairs conventional αβ T cell development beyond the TCR checkpoint[^13]. Analogous to αβ T cell development, the function of Id3 in promoting conventional γδ T cell development has also been mapped downstream of the γδ TCR[^14]. In contrast, however, large populations of iNKT, γδNKT and innate variant T<sub>FH</sub> cells have been observed in the same *Id3*- and *Id2/Id3*-deficient animals, indicating a negative role for Id proteins in regulating innate-like T cell development[^15-21]. However, the mechanism that drives the development and expansion of these innate-like T cell populations in *Id*-deficient mice is still elusive. Given the contradictory nature of Id proteins in supporting conventional T cells and suppressing innate-like T cells, it is reasonable to predict that Id proteins control innate-like T cell development through a distinct, or additional mechanism, that is TCR signaling-independent. Id proteins have also been shown to modulate E protein activity throughout T cell development[^10]. Therefore, it remains to be determined whether Id-mediated suppression of these innate-like T cells is limited to their expansion upon selection and lineage commitment, or if it also influences their lineage specification prior to TCR selection.

We first investigated the role of E proteins, particularly E2A, in iNKT cell development in the absence of Id proteins. E2A ChIP-Seq revealed several genes that are important for iNKT cell
development, such as \textit{Tcf7}\textsuperscript{[22]}, \textit{Egr2}\textsuperscript{[23, 24]} and \textit{Bcl11b}\textsuperscript{[25]}, to be direct targets of E2A. This suggested that Id proteins suppress E2A-driven iNKT-specific genetic programs. We then examined any developmental competition between iNKT and γδNKT cells in \textit{Id3}-deficient mice, which have enhanced development of both these lineages\textsuperscript{[20, 21]} In the absence of γδ T cells, \textit{Id3}-deficient mice had increased iNKT commitment and development, supporting a role for Id3 in regulating the outcome of lineage competition between γδNKT and iNKT cells. Systematic analysis of the transcriptional profiles of expanded iNKT and γδNKT populations, along with the genome-wide E2A binding data, demonstrated that E2A drives downstream transcriptional programs that support both iNKT and γδNKT development in the absence of Id protein(s). We further explored the impact of \textit{Id} deficiency on innate-like T cell development at an earlier stage, i.e. prior to the pre-TCR checkpoint. We found that the deletion of Id proteins in \textit{pre-Tα-/-} mice promoted iNKT cell development without rescuing conventional αβ lineage development. Furthermore, this lack of Id proteins in the absence of pre-TCR signaling uniformly turned on an innate-like transcriptional program among all T cells, including DP and γδ T cells. Collectively, these studies provide compelling evidence to demonstrate the role of Id proteins in the suppression of iNKT and innate-like developmental programs through inhibition of E protein activity, prior to both TCR and pre-TCR selection checkpoints. Our findings also suggest that this suppression mechanism is essential for the predominance of conventional T cell development in the thymus.
Materials and Methods

Study design

Appropriate sample size was determined empirically for each individual experiment or technique in a manner such that statistical significance is reached and scientific rigor is achieved with minimum number of mice necessary. For most experiments, 3 to 5 mice were sufficient per genotype to statistically observe dramatic phenotype changes. Certain experiments depicting cell numbers are cumulative, and have greater number of mice. Each figure is representative of at least 3 sets of independent experiments. All references of replicates in this study are indicative of biological replicates, i.e. individual animals, and not technical replicates. All data points obtained are included in each experiment without any exclusion.

Mice

Id2<sup>f/f</sup>Id3<sup>f/f</sup>LckCre<sup>+</sup> (L-DKO), Id3<sup>−/−</sup> and Id3<sup>−/−</sup> TCR<sup>δ</sup><sup>−/−</sup> mice were generated as previously described.[15, 26] L-DKO pTα<sup>−/−</sup> mice were generated by breeding L-DKO mice with pTα<sup>−/−</sup> mice,[27] which were a generous gift from David L. Weist (Fox Chase Cancer Center, Philadelphia). All mice were bred in a specific pathogen-free facility of Duke University Division of Laboratory Animal Resources, and all procedures were performed according to protocols approved by the Institutional Animal Care and Use Committee.

Cell sorting and RNA extraction

All cells were sorted in FACS buffer using a MoFlo XDP cell sorter. Total mRNA from sorted cells was extracted using an RNAqueous Kit (Life Technology) according to manufacturer’s protocol.

Flow cytometry
Surface marker antibodies were used according to manufacturer’s protocol (Biolegend). Intracellular staining with PLZF antibody (eBioscience) was done using the Foxp3 staining buffer kit (eBioscience). CD1d tetramers were received from the Tetramer Facility of the National Institutes of Health. Stained samples were run on a FACSCanto II machine (BD Biosciences) and data was further analyzed with FlowJo software (Tree Star). Bar graphs were drawn using GraphPad Prism (GraphPad Software). Two-tailed student’s t-test was used for statistics, with p values less than 0.05 considered significant.

**ChIP-Seq analysis**

26 x 10⁶ iNKT and 30 x 10⁶ DP cells were sorted and pooled from multiple L-DKO mice for the E2A ChiP-Seq analysis. 2 x 10⁶ γδ T cells were also sorted from Id3 +/- mice for H3K4me2 analysis. iNKT (CD1dTet+ TCRβ+) and DP (CD1dTet- CD4+ CD8+) cells were sorted from 3-5 weeks old L-DKO mice. For H3K4me2 ChIP-Seq, cells were fixed with 1% formaldehyde, whereas for E2A ChIP-Seq, cells were additionally fixed with 1.5 mM EGS (ethylene glycol-bis(succinic acid N-hydroxsuccinimide ester)). Crosslinked cells were lysed, nuclei were extracted, and sonicated using Bioruptor Plus (Diagenode) and immunoprecipitated with E2A (V-18, Santa Cruz Biotechnology, Lot G0814) or H3K4me2 (Millipore, 07-030, Lot 2430486) antibody. After elution and reverse crosslinking, RNA and proteins were digested, followed by DNA purification using a ChIP DNA Clean and Concentrator kit (Zymoresearch). Libraries were prepared with the NEBNext primer set, which included applying ChIP DNA to end repair, A-tailing, adapter ligation and PCR amplification. Samples were cleaned and size selected by 8% PAGE or AMPure beads (Agencourt). Sequencing was done on HiSeq4000 platform (Illumina).

ChIP-Seq sequenced reads were aligned to the mm9 genome using Bowtie² software (version 1.1.2, parameters: --chunkmbs 128 --mm -m1 --best --strata -p4 -S -q). Peaks were called using
MACS\textsuperscript{29} (version 1.4.2, default parameters). Peaks were then annotated by the NGS: Peak Annotation tool on Nebula.\textsuperscript{30} Bed and wiggle files were generated by MACS for visualization using the Integrative Genomic Viewer.\textsuperscript{31} De-novo motif analysis was done using the findmotifs.pl program with HOMER\textsuperscript{32} (v4.7.2, 50 or 200 bp within each peak).

\textbf{RNA-Seq analysis}

\textit{iNKT (CD1dTet$^+$ TCR$\beta^+$), \gamma\delta\textsuperscript{NKT} (TCR$\gamma\delta^+$ CD3$^+$) and DP (TCR$\gamma\delta^-$ CD4$^+$ CD8$^+$) cells were sorted from 4-5 weeks old WT (total 2 x 10$^5$ \gamma\delta T, 2 x 10$^5$ iNKT cells, 2 x 10$^6$ DP cells), Id3$^{-}$ (total 15 x 10$^6$ \gamma\delta T cells) or L-DKO (total 15 x 10$^6$ iNKT cells, 15 x 10$^6$ DP cells) mice. After RNA extraction using the RNAqueous kit, and quality was assessed using the Agilent Bioanalyzer RNA Pico chip. Ribosomal RNA was depleted using the RiboErase method from Kapa Biosystems. In short, 1ug of total RNA was hybridized with 1ug of hybridization oligos tiling the 18s, 28s, 5.8s and mitochondrial rRNA sequences. Each sample was then RNaseH treated to degrade complementary rRNA sequence. The product was cleaned and purified using 2.2X AMPure beads (Agencourt). The cleaned product was DNase treated to degrade the DNA oligo mix. The remaining rRNA depleted samples were then purified using 2.2X AMPure XP beads. The Kapa Stranded RNA-Seq Kit was used to generate stranded Illumina sequencing libraries (Kapa Biosystems). RNA from was fragmented at 94°C for 6 minutes. Briefly, RNA was hybridized to random primers, followed by first-strand cDNA synthesis, second-strand cDNA synthesis with marking, A-tailing, ligation of Illumina paired-end adapters with 8 bp barcodes, and 9 cycles of PCR amplification. Reactions were purified with Agencourt AMPure XP beads where necessary. Libraries were multiplexed in equimolar amounts, and sequenced as paired-end 50-bp reads using a HiSeq2500 platform (Illumina).}
A second, independent round of RNASeq was done with DP (TCRγδ+ CD4+ CD8+) cells sorted from 5 weeks old WT (9 x 10^5 cells), pTaKO (4.5 x 10^5 cells), L-DKO (9 x 10^5 cells) and L-DKO pTaKO (9 x 10^5 cells) mice. 150 bp paired-end sequencing was done on the HiSeq2500 platform (Illumina). All subsequent data analysis was done using the steps outlined above.

RNA-Seq sequencing reads were first trimmed using Trimmomatic.[33] Read alignment was done using Tophat and expression quantification was done using Cufflinks.[34] Log2 transformed FPKM (fragments per kilobase exon-model per million reads mapped) were used for downstream analyses. Further filtering of low quality genes, Principal Component Analysis, statistical analysis and visualizations were done using R.[35] Pathway analysis was done using the Molecular Signatures Database (MSigDB) v5.2.[36]

**Innate gene signature and network analysis**

Raw microarray expression data was requested and downloaded from Immgen for selected subsets: preT_DN3A_Th (DN3a), preT_DN3B_Th (DN3b), T_DN4_Th (DN4), T_DP_Th (DP), T_4SP69+_Th (post-selection CD4SP), NKT_44-NK1_1-_Th (stage 0 and 1 iNKT cells), Tgd_Th (total thymic γδ T cells), Tgd_vg1+vd6+24ahi_Th (immature Vγ1.1Vδ6.3 cells).

Average gene expression among DN3a, DN3b, DN4, DP and CD4SP cells was assumed to be the reference conventional αβ T cell population. Total thymic γδ T cells were considered as reference for conventional γδ T cell population. Fold change in expression for iNKT and γδNKT cells was calculated with respect to the reference conventional αβ and γδ T cell populations respectively. Genes that had more than 1.5 fold upregulation or 0.6 fold downregulation among both iNKT and γδNKT cells were considered to represent the “innate-like gene signature”. These moderately relaxed fold change parameters allowed us to ensure that maximal numbers of appropriate genes were captured in this analysis. 189 genes were therefore identified from these
specific expression patterns among WT iNKT and γδNKT cells. Additionally, we also included 7
other genes - Tcf3 (E2A), Id2, Id3, Lef1, Sox13, Blk and Sox4 – which have been reported to
play important roles in iNKT and γδNKT lineage development, but did not have expression
patterns that fit our criteria, i.e. being significantly upregulated or downregulated in both cell
types as compared to reference populations. The total 197 genes constituted our innate-like gene
signature, derived from Immgen and literature.

111 of the 197 signature genes were found to be dysregulated in Id3-/- γδ and/or in L-DKO iNKT
cells. Other known interactions between these 111 genes were retrieved from GeneMania.[37] 83
of the 111 genes were also identified as E2A targets, which had E2A binding to the enhancer,
promoter, intragenic, intergenic or downstream regions of these genes, as annotated by Nebula.
These interactions, ChIP-Seq targets and gene expression patterns of the 111 genes were
represented as a network using Cytoscape 3.4.0.[38]

**Gene Set Enrichment Analysis (GSEA)**

The GSEA[36, 39] desktop application (v2.0) was used to analyze the log2FPKM expression
patterns in L-DKO and L-DKO pTα DP samples. 9245 genes that were unchanged in L-DKO DP
samples as compared to WT DP samples were included in this analysis. Enrichment in L-DKO
pTα DP samples over L-DKO samples was determined using weighted, log2(ratio of classes)
parameters and 1000 permutations. The innate-gene signature included 196 genes derived from
Immgen, and listed in Supplementary Table 2. The iNKT development and maturation gene
set[40] (Msigdb gene set M18517) and inflammatory responses gene sets (Msigdb gene set
M5932) were downloaded from Msigdb and used as is.

**Correlation analysis**
To determine correlation with Zbtb16 expression, Pearson and Spearman correlation coefficients were determined for all genes across six samples, including replicates of WT DP, pTαKO DP, L-DKO pTαKO DP and L-DKO DP, as derived from RNA-Seq analysis. Genes with both coefficients greater than or equal to 0.7 were considered to be positively correlated, and those with both coefficients less than or equal to -0.7 were considered to be negatively correlated with Zbtb16 expression. Scatter plots were generated using a custom R script.
Results

**E2A regulates downstream transcription factors critical for iNKT cell development**

Several studies have previously demonstrated that the lack of Id proteins in Id2^fl/fl^Id3^fl/fl^LckCre+ (L-DKO, or LckCre-mediated double knockout) mice gives rise to a significantly large iNKT cell population.\(^{[41]}\) Therefore, we hypothesized that uninhibited E2A activity may induce genes important for the iNKT developmental program. We performed E2A ChIP-seq analysis using L-DKO DP and L-DKO iNKT cells as representative populations prior to, and after CD1d-mediated selection respectively. We found that the number of E2A peaks in L-DKO DP cells was much greater than in L-DKO iNKT cells (**Figure 1a**). However, many of the E2A targets in L-DKO iNKT cells were also occupied by E2A in DP cells. These shared targets likely drive and sustain iNKT fate during and after TCR selection, such that E2A is continually bound in both DP and iNKT cells. Next, we performed de novo motif analysis of E2A binding sites to identify transcription factors that potentially cooperate with E2A in regulating the iNKT developmental program (**Figure 1b**). These identified factors included RUNX1, TCF7, LEF1 and GATA3, which are known to be critical for iNKT development.\(^{[8, 22, 42, 43]}\) RORγt, which is important for regulating survival of DP cells and, as consequence, for the distal iNKT TCRα rearrangement,\(^{[44]}\) was found to be a co-factor only in DP cells, but not in iNKT cells. ChIP-Seq analysis also revealed E2A peaks in L-DKO DP cells at these cooperating transcription factor loci, indicating that E2A can directly regulate these factors to support the downstream iNKT program (**Figure 1c**). E2A peaks at various loci, including at RORγt, were also shared between our mutant L-DKO cells and previously characterized peaks in DN3, DN4 cells with Wild Type (WT) E protein activity and a T cell line with E protein overexpression, suggesting that E2A might regulate similar targets in developing T cells in WT mice (**Supplementary Figure 1a**).\(^{[11]}\)
E2A also bound to the promoter and/or enhancer regions of several genes that have been reported to be critical for iNKT cell development and function (Figure 1d). This included genes such as Tcf7, Bcl11b, Sox4 and Gzma. While direct regulation of PLZF (encoded by Zbtb16) by E2A has been reported previously, we found E2A binding to intragenic, rather than putative regulatory regions of Zbtb16 in L-DKO iNKT or DP cells (Supplementary Figure 1b). It is possible that there is an indirect regulation of PLZF through another downstream E2A target in this context. We found E2A bound to the Egr2 promoter (Supplementary Figure 1b), which has also been shown to drive PLZF expression in iNKT cells, making it a likely mediator. Overall, these findings indicate directed regulation of the iNKT developmental program by E2A and other cooperating transcription factors in the absence of Id proteins. More importantly, the comparison of E2A targets between L-DKO DP and L-DKO iNKT cells suggest a possible engagement of E2A in orchestrating an iNKT transcription signature prior to TCR expression.

**Absence of γδ T cells enhances iNKT cell development in Id3-deficient mice**

In the previous sections, we investigated the mechanisms that drive iNKT cell development in Id2/Id3-double deficient mice. Interestingly, Id3-deficient mice have been described to support the expansion of both γδNKT and iNKT cells. We have also previously shown that varying activity levels of E proteins can differentially influence lineage outcomes between γδNKT and iNKT cells. Additionally, recent publications have demonstrated the sharing of transcriptional programs between these innate-like lineages, particularly between γδNKT and NKT2 cells. Given the transcriptional and functional similarities between these two innate-like populations, and their expansion in Id-deficient mice, we wanted to examine if Id proteins suppressed these lineages equally or differentially, particularly in the developmental window prior to TCR expression.
selection. We therefore tested lineage competition in Id3-deficient mice by eliminating γδ lineage development and expansion.

We found a significant increase in the iNKT population in Id3−/−TCRδ−/− mice as compared to Id3−/− mice (Figure 2a-c). There was a modest but significant increase in iNKT lineage-committed stage 0 cells in Id3−/−TCRδ−/− mice as compared to Id3−/− mice (Figure 2d, e). It has been reported that iNKT cells and γδNKT cells compete for a thymic niche, based on the reduction in iNKT cells upon expansion of γδNKT cells. In contrast, another study has reported that a reduction in iNKT cells does not lead to a corresponding increase in γδNKT cells. In order to separate ongoing T cell development from homeostatic expansion associated with void space, we examined pre-weaning pups that have not yet undergone full expansion and stabilization of the thymic architecture. A large increase in the iNKT population was observed again in pre-weaning age Id3−/−TCRδ−/− mice (Figure 2f, g). Age matched TCRδ−/− mice, which lack total γδ T cells but are sufficient in Id3, also did not show any corresponding increase in iNKT cells. Overall, this data suggests that Id proteins control the outcome of the lineage competition between γδNKT and iNKT cells. Since the former arise at the DN3 stage and the latter at the DP stage, it also indicates that an early lineage checkpoint may be shared between them.

**E2A drives a gene network that promotes iNKT and innate-like fate in the absence of Id proteins**

Since both iNKT and γδNKT cell populations were amplified in the absence of Id protein(s), and we identified E2A as a potential regulator of iNKT-specific transcription factors, we wanted to determine whether E2A controls the transcription programs critical for either or both of these innate-like T cell lineages. We have previously characterized the predominantly enhanced γδNKT development in Id3−/− mice, and iNKT development in L-DKO mice. Therefore, we
performed RNA-Seq analysis of Id3−/− γδ (predominantly γδNKT cells), L-DKO iNKT and L-DKO DP cells, and their wild type (WT) counterparts. Interestingly, both Principal Component Analysis (PCA) and clustering analysis highlighted transcriptional similarity between L-DKO iNKT and Id3−/− γδNKT cells (Figure 3a, b). There were 1611 and 1865 genes dysregulated by at least two fold in L-DKO iNKT and Id3−/− γδNKT cells respectively, as compared to their WT controls (Figure 3c). Among these, more than 400 genes shared similar expression patterns in the two cell types (Figure 3c, d). Pathway analysis of dysregulated genes revealed involvement of several pathways, including IL2/STAT5, p53 and MTORC signaling (Figure 3e). We also found enrichment of genes that were dysregulated in T-to-natural killer (ITNK) cells upon Bcl11b deletion (Figure 3e).[52]

In order to specifically identify any dysregulation of innate-specific genes, we decided to filter differentially expressed genes in L-DKO iNKT and/or Id3−/− γδNKT cells, i.e. a total of 3060 genes, against an innate-like gene signature. We first searched for the reference innate-like signature genes from publicly available Immgen data.[53] We hypothesized that the genes that are upregulated or downregulated significantly in both iNKT and γδNKT cells over other T cell populations in the thymus, would be unique to these innate-like lineages and most likely be important for their development and/or function. Upon TCR selection, the divergence of innate-like lineages from conventional T cells is primarily represented by CD44NK1.1− stage 0/1 iNKT cells and CD24hi immature γδNKT cells. These cells are also phenotypically most similar to the expanding populations in Id-deficient mice. Therefore, we compared the gene expression in these WT innate-like T cells against multiple WT conventional T cell populations, including DN3a, DN3b, DN4, DP, post-selection CD4SP and thymic γδ T cells. This analysis resulted in 189 signature genes which are significantly overexpressed or repressed in both WT iNKT and
γδNKT populations. We also added 7 other genes to this list that are reported in literature to be important for their development, but are not significantly overexpressed or repressed in these populations (Supplementary Table I). When this final innate-like gene set was compared to the genes identified in our RNA-Seq analysis, we found more than 50% (111 genes) to be dysregulated by at least two fold in either or both of the cell populations in Id-deficient mice (Figure 4a, Supplementary Table II). Importantly, E2A directly bound to many (83 of 111 genes) of these differentially expressed signature genes. This data suggested that E2A plays a profound role in regulating the expression of genes critical for innate-like T cell development.

In order to further delineate the role of E proteins in regulating the developmental programs of iNKT and γδNKT cells through these downstream mediators, we divided the 111 genes into 3 groups based on their expression profiles in L-DKO iNKT and Id3−/− γδNKT cells (Figure 4b, Supplementary Table II). Groups 1 and 2 included “biased” genes that were upregulated or downregulated by a significantly larger magnitude in one population as compared to the other. On the other hand, group 3 included “common” genes that were significantly and similarly upregulated or downregulated in both innate-like populations. These categories broadly represent genetic programs that are potentially important for the lineage-specific development of iNKT (group 1) and γδNKT cells (group 2), as well as for overall innate-like T cell specification (group 3). By combining known interactions between these genes with our RNA-Seq and ChIP-Seq data, we created a network map with the three groups of genes demarcated (Figure 4c). The distribution of E2A targets across all three groups strongly reaffirmed the role of E2A in orchestrating innate-like T cell developmental programs. Our previous observation of diminished iNKT and γδNKT populations in Id2Δf/Δf Id3Δf/Δf E2AΔf/Δf HEBΔf/Δf LckCre+ (or Q-KO) mice that lack E
protein activity further supports the pivotal role of E proteins in the development of these cells.\textsuperscript{[15]}

**Lack of Id proteins promotes innate-like T cell development independent of pre-TCR signaling**

All observations thus far suggested the promotion of innate-like T cell fate by uninhibited E protein activity. The E2A occupancy of innate-like genes at DP cells and the lineage competition between iNKT and γδNKT cells further prompted us to consider the possibility that Id proteins may regulate the fate choice for innate-like T cells early in T cell development. We therefore tested the requirement of pre-TCR signaling in the development of innate-like T cells in Id-deficient mice. A functional pre-TCR signal at the DN3 stage is critical for conventional αβ T cell development. Mice deficient in pre-Tα, are therefore known to have restricted T cell development, with the majority of cells blocked at the DN stage.\textsuperscript{[27]} pTα\textsuperscript{−/−} mice are also known to completely lack iNKT cells.\textsuperscript{[54]} We crossed pTα\textsuperscript{−/−} mice to L-DKO to generate Id2\textsuperscript{fl/fl}Id3\textsuperscript{fl/fl}LckCre\textsuperscript{+} pTα\textsuperscript{−/−} (L-DKO pTα\textsuperscript{−/−}) mice.

Despite the complete absence of iNKT cells in pTα\textsuperscript{−/−} mice, to our surprise, we found a robust iNKT population in L-DKO pTα\textsuperscript{−/−} mice (Figure 5a-c). These iNKT cells also expressed high levels of PLZF, which is closely associated with the innate-like lineage gene expression program (Figure 5d). It is known that pTα\textsuperscript{−/−} mice have an increase in γδ T cells,\textsuperscript{[27]} and L-DKO pTα\textsuperscript{−/−} mice showed a similar increase in the γδ population compared to WT mice (Figure 5e).

However, the γδ T cells in L-DKO pTα\textsuperscript{−/−} mice were predominantly Vγ1.1\textsuperscript{+}Vδ6.3\textsuperscript{+} and uniformly upregulated PLZF, reflecting a specific increase in innate-like γδNKT cells in these mice (Figure 5f-i).

As expected, L-DKO pTα\textsuperscript{−/−} mice still had a profound block in conventional αβ T cell development due to the lack of pre-TCR signaling (Figure 6a, b). Interestingly, despite the pre-
TCR block in L-DKO pTα−/− mice, the deletion of Id proteins seemed to partially rescue the development of DP cells (Figure 6a, c). Upon careful investigation, however, we found that these DP cells upregulate PLZF (Figure 6d). These PLZFhi DP cells, however, did not recognize the CD1d tetramer, indicating that they are not iNKT cells that aberrantly upregulate CD4 and CD8. Total thymocytes from L-DKO pTα−/− mice also displayed a prevalent innate-like phenotype, as indicated by their PLZF expression pattern (Figure 6e). Thus, blocking conventional αβ T lineage development with pTα-deficiency revealed a pre-TCR independent pathway that drives innate-like T cell, including iNKT, lineage development in Id2/Id3-deficient mice.

**Initiation of an innate-like T cell transcriptional program in the absence of Id proteins and conventional T cell development**

Our previous observation suggested that pre-TCR signaling and Id protein activity is necessary to enforce conventional T cell fate, such that the absence of both gave rise to predominantly innate-like T cell populations in the thymus. We further verified the innate-like phenotype of PLZFhi DP cells in L-DKO pTα−/− mice by RNA-Seq analysis. 28 genes out of the 196 genes in the innate-like gene signature defined previously (Supplementary Table II) were found to be upregulated by more than 2 fold in L-DKO pTα−/− DP cells, as compared to WT DP cells (Figure 6f). The gene expression pattern for L-DKO DP was significantly different from L-DKO pTα−/− DP cells, including lower expression of key innate-like signature genes, such as Zbtb16 and Gzma. Gene Set Enrichment Analysis (GSEA) verified enrichment of the innate-like gene signature, as well as genes associated with iNKT cell development and inflammatory responses in L-DKO pTα−/− DP cells as compared to L-DKO DP cells (Figure 6g). RNA-Seq analysis of
the PLZF^hi^ DP cells also revealed these cells had undergone TCRα rearrangement with a fairly broad V-J usage (data not shown).

Since PLZF is currently one of the only well-defined innate-like transcription factors, we decided to alternatively explore the innate-like transcription program by identifying genes that are positively or negatively correlated with PLZF expression. The idea is that genes with similar expression patterns can be expected to function together and/or be involved in similar biological processes. We computed pair-wise correlations between the expression of Zbtb16 and all other genes across WT, pTα^-/-, L-DKO and L-DKO pTα^-/- DP cells (Supplementary Table III). A scatter plot of 465 positive and 460 negatively correlated genes revealed clear differences in expression patterns between L-DKO pTα^-/- and L-DKO DP cells (Figure 6h). Most genes upregulated in L-DKO pTα^-/- DP cells were positively correlated with Zbtb16, whereas most genes that were downregulated in these cells had negative correlation with Zbtb16. Interestingly, a significant fraction of positively correlated genes (308 genes) had a higher expression in L-DKO pTα^-/- DP cells. On the other hand, a significant fraction of negatively correlated genes (283 genes) had a higher expression in L-DKO DP cells. These analyses demonstrate the early initiation and adoption of an innate-like transcriptional program specifically in DP cells in L-DKO pTα^-/- mice. Therefore, we observe a complete diversion of cells towards innate-like T cell fate as a result of uninhibited E protein activity and lack of conventional T cells in the thymus.

Overall, these results suggest that Id proteins are potent suppressors of innate-like T cell fate at the pre-TCR checkpoint.
In order to better understand the opposing roles of Id proteins in regulating conventional and innate-like T cell development, in this paper, we examined their role in suppressing innate-like T cells, including iNKT and γδNKT cells, prior to TCR selection. Our genetic studies clearly revealed a developmental behavior of the iNKT lineage that is inconceivable with our current understanding of these cells, whose lineage fate choice normally does not start until the TCR selection stage. Our finding represents the first case where iNKT lineage development can be distinguished from the conventional αβT lineage as early as the pre-TCR checkpoint, albeit under a unique genetic background. The combined genome-wide E2A binding and transcriptional data provides a direct explanation to the several previous reports of expansion of these innate-like populations in the absence of Id proteins. As more genes are increasingly reported to be important for innate T cell development and effector functions, our data also serves as a repository to examine the role of E2A in their regulation distinctly or commonly in both iNKT and γδNKT populations, particularly during early stages of their development.

Our E2A ChIP-Seq analysis of Id2/Id3-deficient (L-DKO) DP and iNKT cells placed E2A as either collaborator or upstream regulator of genes and transcription factors known to be important for iNKT cell development. The regulation of these gene targets at the DP stage, prior to the TCR-mediated selection of iNKT cells, indicated early lineage specification events for iNKT cells, as regulated by E2A. We also observed shared binding patterns of E2A in L-DKO DP and iNKT cells, with those previously documented in DN3 and DN4 cells with no Id deficiency. This suggests that several E2A targets identified in Id-deficient mice may also be similarly regulated in WT mice to promote innate-like T cell development.
It has been shown that the effector programs of iNKT2 cells closely resemble that of γδNKT cells,[49] even though these two lineages are considered to independently diverge at the DP and DN3 stages respectively. The effector iNKT cells expanded in Id2/Id3-deficient mice most closely resembles the iNKT2 lineage.[21, 46] Clustering analysis of our RNA-Seq data indicated transcriptional similarity between both WT and Id-deficient iNKT and γδNKT cells. The lineage competition between these innate-like populations revealed by Id3−/−TCRδ−/− mice further hints towards the developmental lineage proximity between these two lineages. Due to the typically neonatally restricted development of γδNKT cells, it is unfeasible to verify this phenotype using bone-marrow chimeras.[19] However, the increase in stage 0 iNKT cells in the absence of γδNKT cells suggests that these lineages might have shared innate-like precursors, whose lineage choice is regulated by Id proteins.[55] This data also challenges the stochastic TCR-driven iNKT developmental model at the DP stage, and warrants re-examination of the lineage origin of iNKT cells.

Our study also revealed E2A-mediated transcription programs that support the development of γδNKT and iNKT lineages. This analysis identified E2A as an upstream regulator of genes critical for iNKT lineage differentiation, including Zbtb16, Slamf6 and Egr2 (Figure 3e).[3, 4, 6, 23, 24, 56-59] Among other target genes, we also found a dramatic downregulation of Bcl11b in γδNKT cells, which is needed to suppress a Natural Killer (NK) program in conventional T cells.[60] This indicated a possible role for Bcl11b in the lineage choice between conventional γδT and γδNKT cell development. Genes that are associated with NTK1 and NKT17 cytokine profiles, such as Ifngr1 and Rorc, were found to be significantly downregulated in both Id3−/−γδNKT and L-DKO iNKT cells. These findings demonstrate the involvement of E2A in preferentially driving γδNKT and NKT2 lineage development in the absence of Id proteins.
In this study, we also uncovered a pre-TCR independent pathway for the development of iNKT and innate-like T cells using our L-DKO pTα−/− mice. Interestingly, the expanded innate-like phenotype in the L-DKO pTα−/− mice was not limited to only iNKT and γδNKT cells, but also a novel innate-like PLZFhi DP population. The early induction of an innate-like transcriptional program among all developing T cells in these mice with high E protein activity demonstrates the necessity of Id proteins in suppressing innate-like fate choice. The expanded innate-like populations in our Id-deficient mice allowed us to uncover the regulation of innate-like T cell lineage development prior to TCR selection. However, the divergence of iNKT and innate-like T cells from conventional T cells prior to TCR selection has also been proposed in other mouse models with physiological levels of E protein activity.[55, 61] It is likely that the depletion of Id proteins unleashes the “early”, pre-TCR-independent developmental program for iNKT and other innate-like T cells, that can usually be observed in much smaller frequencies in other studies. Consequently, we have also observed heterogeneous innate-like αβ T cell lymphomas in Id2/Id3 deficient mice that are derived from iNKT, CD1dTet− or TFH cells.[16, 62]

PLZF undoubtedly plays key roles in innate-like T cell development and effector function. However, the transcriptional regulation of PLZF itself remains ambiguous. While some studies have found strong agonistic signaling to be sufficient to induce PLZF expression, [5, 24, 63] others have reported early and stable suppression of PLZF in conventional T cells, which can’t be reversed by TCR signaling alone.[64] The function of PLZF itself, in promoting innate-like effector genes, has also been demonstrated to be independent of agonistic TCR signals.[65, 66] This implies that unveiling the transcriptional regulation of PLZF and lineage specification of innate-like fate in developing T cells is critical for advancing our understanding of innate-like T cells. Our data suggests that Id-mediated inhibition of E2A is able to suppress the downstream
activation of PLZF and innate-like transcription programs in majority of developing T cells as early as the pre-TCR selection step. This supports a layered, rather than a parallel, developmental structure that coordinates the distinct developmental fates of innate and conventional T cells in the thymus.\[45\]

Although innate-like T cells represent only a small fraction of the thymic population, their indispensable roles in mounting rapid immune responses in different contexts warrants a holistic understanding of the regulation of their concurrent development with conventional T cells in the thymus. In this study, we characterized E2A-driven transcriptional programs that promote innate-like T cell development prior to TCR selection and independent of pre-TCR signaling, which are otherwise suppressed by Id proteins. Not surprisingly, phylogenetic analysis of innate-like T cells and their associated transcription factors indicates that these cells emerged much earlier than conventional T cells in the course of evolution\[45, 67\]. Hence, we propose that innate-like lineage specification precedes conventional αβ T cells in the thymus, and that evolutionary pressures necessitated Id-mediated suppression to ensure the predominance of conventional αβ T cells. Our data also suggests that the collaborative action of Id proteins and pre-TCR signaling delays the fate choice of iNKT and innate-like T cells to drive conventional T cell development in the thymus.

**Acknowledgments**

We thank Drs. A. Lasorella and A. Lavarone (Columbia University) for sharing the Id2f/f strain. We thank the Duke Cancer Center Flow Cytometry Facility for assistance in cell sorting, Duke Cancer Center Sequencing Facility for assistance in Ion Torrent sequencing, UC San Diego IGM Genomics Centre for assistance in ChIP-Seq sequencing, Duke Center for Genomic and Computational Biology for assistance in RNA-Seq sequencing, and the NIH tetramer facility for...
providing the CD1d tetramer. We thank Drs. M. Krangel, M. Ciofani, B. Mathey-Prevot and Mr. Zachary Carico (Duke University) for their expertise and feedback.

References

1. Godfrey D. I., Uldrich A. P., McCluskey J., Rossjohn J. and Moody D. B. 2015. The burgeoning family of unconventional T cells. *Nat Immunol*. 16: 1114-1123

2. Alonzo E. S. and Sant’Angelo D. B. 2011. Development of PLZF-expressing innate T cells. *Curr Opin Immunol*. 23: 220-227

3. Savage A. K., Constantinides M. G., Han J., Picard D., Martin E., Li B., Lantz O. and Bendelac A. 2008. The transcription factor PLZF directs the effector program of the NKT cell lineage. *Immunity*. 29: 391-403

4. Kovalovsky D., Uche O. U., Eladad S., Hobbs R. M., Yi W., Alonzo E., Chua K., Eidson M., Kim H. J., Im J. S., Pandolfi P. P. and Sant’Angelo D. B. 2008. The BTB-zinc finger transcriptional regulator PLZF controls the development of invariant natural killer T cell effector functions. *Nat Immunol*. 9: 1055-1064

5. Kreslavsky T., Savage A. K., Hobbs R., Gounari F., Bronson R., Pereira P., Pandolfi P. P., Bendelac A. and von Boehmer H. 2009. TCR-inducible PLZF transcription factor required for innate phenotype of a subset of gammadelta T cells with restricted TCR diversity. *Proc Natl Acad Sci U S A*. 106: 12453-12458

6. Mao A. P., Constantinides M. G., Mathew R., Zuo Z., Chen X., Weirauch M. T. and Bendelac A. 2016. Multiple layers of transcriptional regulation by PLZF in NKT-cell development. *Proc Natl Acad Sci U S A*. 113: 7602-7607

7. Gapin L., Matsuda J. L., Surh C. D. and Kronenberg M. 2001. NKT cells derive from double-positive thymocytes that are positively selected by CD1d. *Nat Immunol*. 2: 971-978

8. Egawa T., Eberl G., Taniuchi I., Benlagha K., Geissmann F., Hennighausen L., Bendelac A. and Littman D. R. 2005. Genetic evidence supporting selection of the Valpha14i NKT cell lineage from double-positive thymocyte precursors. *Immunity*. 22: 705-716

9. Azuara V., Levraud J. P., Lembezat M. P. and Pereira P. 1997. A novel subset of adult gamma delta thymocytes that secretes a distinct pattern of cytokines and expresses a very restricted T cell receptor repertoire. *Eur J Immunol*. 27: 544-553

10. Engel I. and Murre C. 2001. The function of E- and Id proteins in lymphocyte development. *Nat Rev Immunol*. 1: 193-199
11. 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

12. Miyazaki M., Miyazaki K., Chen S., Itoi M., Miller M., Lu L. F., Varki N., Chang A. N., Broide D. H. and Murre C. 2014. Id2 and Id3 maintain the regulatory T cell pool to suppress inflammatory disease. *Nat Immunol.* 15: 767-776

13. Jones-Mason M. E., Zhao X., Kappes D., Lasorella A., Iavarone A. and Zhuang Y. 2012. E protein transcription factors are required for the development of CD4(+) lineage T cells. *Immunity.* 36: 348-361

14. Lauritsen J. P., Wong G. W., Lee S. Y., Lefebvre J. M., Ciofani M., Rhodes M., Kappes D. J., Zuniga-Pflucker J. C. and Wiest D. L. 2009. Marked induction of the helix-loop-helix protein Id3 promotes the gammadelta T cell fate and renders their functional maturation Notch independent. *Immunity.* 31: 565-575

15. Li J., Wu D., Jiang N. and Zhuang Y. 2013. Combined deletion of id2 and id3 genes reveals multiple roles for e proteins in invariant NKT cell development and expansion. *J Immunol.* 191: 5052-5064

16. Miyazaki M., Miyazaki K., Chen S., Chandra V., Wagatsuma K., Agata Y., Rodewald H. R., Saito R., Chang A. N., Varki N., Kawamoto H. and Murre C. 2015. The E-Id protein axis modulates the activities of the PI3K-AKT-mTORC1-Hif1a and c-myc/p19Arf pathways to suppress innate variant TFH cell development, thymocyte expansion, and lymphomagenesis. *Genes Dev.* 29: 409-425

17. Alonzo E. S., Gottschalk R. A., Das J., Egawa T., Hobbs R. M., Pandolfi P. P., Pereira P., Nichols K. E., Koretzky G. A., Jordan M. S. and Sant'Angelo D. B. 2010. Development of promyelocytic zinc finger and ThPOK-expressing innate gamma delta T cells is controlled by strength of TCR signaling and Id3. *J Immunol.* 184: 1268-1279

18. Verykokakis M., Boos M. D., Bendelac A. and Kee B. L. 2010. SAP protein-dependent natural killer T-like cells regulate the development of CD8(+) T cells with innate lymphocyte characteristics. *Immunity.* 33: 203-215

19. Verykokakis M., Boos M. D., Bendelac A., Adams E. J., Pereira P. and Kee B. L. 2010. Inhibitor of DNA binding 3 limits development of murine slam-associated adaptor protein-dependent "innate" gammadelta T cells. *PLoS One.* 5: e9303

20. Zhang B., Lin Y. Y., Dai M. and Zhuang Y. 2014. Id3 and Id2 act as a dual safety mechanism in regulating the development and population size of innate-like gammadelta T cells. *J Immunol.* 192: 1055-1063
21. Verykokakis M., Krishnamoorthy V., Iavarone A., Lasorella A., Sigvardsson M. and Kee B. L. 2013. Essential functions for ID proteins at multiple checkpoints in invariant NKT cell development. *J Immunol*. 191: 5973-5983

22. Berga-Bolanos R., Zhu W. S., Steinke F. C., Xue H. H. and Sen J. M. 2015. Cell-autonomous requirement for TCF1 and LEF1 in the development of Natural Killer T cells. *Mol Immunol*. 68: 484-489

23. Hu T., Gimferrer I., Simmons A., Wiest D. and Alberola-Ila J. 2011. The Ras/MAPK pathway is required for generation of iNKT cells. *PLoS One*. 6: e19890

24. Seiler M. P., Mathew R., Liszewski M. K., Spooner C. J., Barr K., Meng F., Singh H. and Bendelac A. 2012. Elevated and sustained expression of the transcription factors Egr1 and Egr2 controls NKT lineage differentiation in response to TCR signaling. *Nat Immunol*. 13: 264-271

25. Uddin M. N., Sultana D. A., Lorentsen K. J., Cho J. J., Kirst M. E., Brantly M. L., Califano D., Sant'Angelo D. B. and Avram D. 2016. Transcription factor Bcl11b sustains iNKT1 and iNKT2 cell programs, restricts iNKT17 cell program, and governs iNKT cell survival. *Proc Natl Acad Sci U S A*. 113: 7608-7613

26. Belle I., Mahlios J., McKenzie A. and Zhuang Y. 2014. Aberrant production of IL-13 by T cells promotes exocrinopathy in Id3 knockout mice. *Cytokine*. 69: 226-233

27. Fehling H. J., Krotkova A., Saint-Ruf C. and von Boehmer H. 1995. Crucial role of the pre-T-cell receptor alpha gene in development of alpha beta but not gamma delta T cells. *Nature*. 375: 795-798

28. Langmead B., Trapnell C., Pop M. and Salzberg S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10: R25

29. 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

30. Boeva V., Lermine A., Barette C., Guillouf C. and Barillot E. 2012. Nebula--a web-server for advanced ChIP-seq data analysis. *Bioinformatics*. 28: 2517-2519

31. Thorvaldsdottir H., Robinson J. T. and Mesirov J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 14: 178-192

32. 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
33. Bolger A. M., Lohse M. and Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30: 2114-2120

34. Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D. R., Pimentel H., Salzberg S. L., Rinn J. L. and Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 7: 562-578

35. Team R. C. 2013. R: A language and environment for statistical computing.

36. Subramanian A., Tamayo P., Mootha V. K., Mukherjee S., Ebert B. L., Gillette M. A., Paulovich A., Pomeroy S. L., Golub T. R., Lander E. S. and Mesirov J. P. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 102: 15545-15550

37. Warde-Farley D., Donaldson S. L., Comes O., Zuberi K., Badrawi R., Chao P., Franz M., Grouios C., Kazi F., Lopes C. T., Maitland A., Mostafavi S., Montojo J., Shao Q., Wright G., Bader G. D. and Morris Q. 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 38: W214-220

38. Shannon P., Markiel A., Ozier O., Baliga N. S., Wang J. T., Ramage D., Amin N., Schwikowski B. and Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 13: 2498-2504

39. Mootha V. K., Lindgren C. M., Eriksson K. F., Subramanian A., Sihag S., Lehar J., Puigserver P., Carlsson E., Ridderstrale M., Laurila E., Houpts N., Daly M. J., Patterson N., Mesirov J. P., Golub T. R., Tamayo P., Spiegelman B., Lander E. S., Hirschhorn J. N., Altshuler D. and Groop L. C. 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 34: 267-273

40. Matsuda J. L., Zhang Q., Ndonye R., Richardson S. K., Howell A. R. and Gapin L. 2006. T-bet concomitantly controls migration, survival, and effector functions during the development of Valpha14i NKT cells. *Blood*. 107: 2797-2805

41. Roy S. and Zhuang Y. 2015. Orchestration of invariant natural killer T cell development by E and Id proteins. *Crit Rev Immunol*. 35: 33-48

42. Kim P. J., Pai S. Y., Brigl M., Besra G. S., Gumperz J. and Ho I. C. 2006. GATA-3 regulates the development and function of invariant NKT cells. *J Immunol*. 177: 6650-6659

43. Carr T., Krishnamoorthy V., Yu S., Xue H. H., Kee B. L. and Verykokakis M. 2015. The transcription factor lymphoid enhancer factor 1 controls invariant natural killer T cell expansion and Th2-type effector differentiation. *J Exp Med*. 212: 793-807

44. D'Cruz L. M., Knell J., Fujimoto J. K. and Goldrath A. W. 2010. An essential role for the transcription factor HEB in thymocyte survival, Tcra rearrangement and the development of natural killer T cells. *Nat Immunol*. 11: 240-249
45. Kang J. and Malhotra N. 2015. Transcription factor networks directing the development, function, and evolution of innate lymphoid effectors. *Annu Rev Immunol*. 33: 505-538

46. D'Cruz L. M., Stradner M. H., Yang C. Y. and Goldrath A. W. 2014. E and Id proteins influence invariant NKT cell sublineage differentiation and proliferation. *J Immunol*. 192: 2227-2236

47. Ueda-Hayakawa I., Mahlios J. and Zhuang Y. 2009. Id3 restricts the developmental potential of gamma delta lineage during thymopoiesis. *J Immunol*. 182: 5306-5316

48. Bezman N. A., Kim C. C., Sun J. C., Min-Oo G., Hendricks D. W., Kamimura Y., Best J. A., Goldrath A. W., Lanier L. L. and Immunological Genome Project C. 2012. Molecular definition of the identity and activation of natural killer cells. *Nat Immunol*. 13: 1000-1009

49. Lee Y. J., Starrett G. J., Lee S. T., Yang R., Henzler C. M., Jameson S. C. and Hogquist K. A. 2016. Lineage-Specific Effector Signatures of Invariant NKT Cells Are Shared amongst gammadelta T, Innate Lymphoid, and Th Cells. *J Immunol*. 197: 1460-1470

50. Pereira P. and Boucontet L. 2012. Innate NKTgammadelta and NKTalphabeta cells exert similar functions and compete for a thymic niche. *Eur J Immunol*. 42: 1272-1281

51. Sandrock I., Zietara N., Lyszkiewicz M., Oberdorfer L., Witzlau K., Krueger A. and Prinz I. 2015. MicroRNA-181a/b-1 Is Not Required for Innate gammadelta NKT Effector Cell Development. *PLoS One*. 10: e0145010

52. Li P., Burke S., Wang J., Chen X., Ortiz M., Lee S. C., Lu D., Campos L., Goulding D., Ng B. L., Dougan G., Huntly B., Gottgens B., Jenkins N. A., Copeland N. G., Colucci F. and Liu P. 2010. Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. *Science*. 329: 85-89

53. Heng T. S., Painter M. W. and Immunological Genome Project C. 2008. The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol*. 9: 1091-1094

54. Eberl G., Fehling H. J., von Boehmer H. and MacDonald H. R. 1999. Absolute requirement for the pre-T cell receptor alpha chain during NK1.1+ TCRalphabeta cell development. *Eur J Immunol*. 29: 1966-1971

55. Kisielow J., Tortola L., Weber J., Karjalainen K. and Kopf M. 2011. Evidence for the divergence of innate and adaptive T-cell precursors before commitment to the alphabeta and gammadelta lineages. *Blood*. 118: 6591-6600

56. Chung B., Aoukaty A., Dutz J., Terhorst C. and Tan R. 2005. Signaling lymphocytic activation molecule-associated protein controls NKT cell functions. *J Immunol*. 174: 3153-3157
57. Nichols K. E., Hom J., Gong S. Y., Ganguly A., Ma C. S., Cannons J. L., Tangye S. G., Schwartzberg P. L., Koretzky G. A. and Stein P. L. 2005. Regulation of NKT cell development by SAP, the protein defective in XLP. *Nat Med*. 11: 340-345

58. Pasquier B., Yin L., Fondaneche M. C., Relouzat F., Bloch-Queyrat C., Lambert N., Fischer A., de Saint-Basile G. and Latour S. 2005. Defective NKT cell development in mice and humans lacking the adapter SAP, the X-linked lymphoproliferative syndrome gene product. *J Exp Med*. 201: 695-701

59. Lazarevic V., Zullo A. J., Schweitzer M. N., Staton T. L., Gallo E. M., Crabtree G. R. and Glimcher L. H. 2009. The gene encoding early growth response 2, a target of the transcription factor NFAT, is required for the development and maturation of natural killer T cells. *Nat Immunol*. 10: 306-313

60. Li L., Leid M. and Rothenberg E. V. 2010. An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. *Science*. 329: 89-93

61. Dashtsoodol N., Shigeura T., Aihara M., Ozawa R., Kojo S., Harada M., Endo T. A., Watanabe T., Ohara O. and Taniguchi M. 2017. Alternative pathway for the development of Valpha14+ NKT cells directly from CD4-CD8- thymocytes that bypasses the CD4+CD8+ stage. *Nat Immunol.*

62. Li J., Roy S., Kim Y. M., Li S., Zhang B., Love C., Reddy A., Rajagopalan D., Dave S., Diehl A. M. and Zhuang Y. 2017. Id2 Collaborates with Id3 To Suppress Invariant NKT and Innate-like Tumors. *J Immunol*. 198: 3136-3148

63. Dutta M., Kraus Z. J., Gomez-Rodriguez J., Hwang S. H., Cannons J. L., Cheng J., Lee S. Y., Wiest D. L., Wakeland E. K. and Schwartzberg P. L. 2013. A role for Ly108 in the induction of promyelocytic zinc finger transcription factor in developing thymocytes. *J Immunol*. 190: 2121-2128

64. Zhang S., Laouar A., Denzin L. K. and Sant’Angelo D. B. 2015. Zbtb16 (PLZF) is stably suppressed and not inducible in non-innate T cells via T cell receptor-mediated signaling. *Sci Rep*. 5: 12113

65. Kovalovsky D., Alonzo E. S., Uche O. U., Eidson M., Nichols K. E. and Sant’Angelo D. B. 2010. PLZF induces the spontaneous acquisition of memory/effector functions in T cells independently of NKT cell-related signals. *J Immunol*. 184: 6746-6755

66. Savage A. K., Constantinides M. G. and Bendelac A. 2011. Promyelocytic leukemia zinc finger turns on the effector T cell program without requirement for agonist TCR signaling. *J Immunol*. 186: 5801-5806

67. Edholm E. S., Banach M. and Robert J. 2016. Evolution of innate-like T cells and their selection by MHC class I-like molecules. *Immunogenetics*. 68: 525-536
1. This work has been supported by National Institute of Health grants GM R01 GM059638, and 1 P01 AI102853 awarded to Dr. Yuan Zhuang.

2. Id: Inhibitor of DNA binding

3. iNKT - invariant Natural Killer T cell

4. CD1dTet - CD1d tetramer

5. L-DKO - Id2^{f/f}Id3^{f/f}LckCre^{+}

6. TKO - Id2^{f/f}Id3^{f/f}LckCre^{+}CD1d^{-/-}

7. PLZF - Promyelocytic Leukemia Zinc Finger

8. WT - Wild Type
Figure Legends

Figure 1. E2A activates and collaborates with transcription factors critical for iNKT cell development.

(a) Number of genes with E2A binding in L-DKO DP and/or L-DKO iNKT cells. (b) De novo motif analysis of E2A peaks in L-DKO DP or iNKT samples with p-values (corresponding to DP samples for shared motifs, or for respective samples for unique motifs). (c) E2A peaks in L-DKO DP and L-DKO iNKT cells at loci for motifs identified in (b). (d) E2A peaks at selective loci. Tracks include L-DKO DP and iNKT samples, along with their input control (i.e. not immunoprecipitated with E2A antibody), and H3K4me2 methylation in Id3-/- γδ T cells, which is indicative of transcription factor binding. Important E2A peaks are highlighted by boxes within each panel. Solid black lines below the input control tracks indicate significant (p value less than $10^{-5}$) peaks called by MACS. Length of the genome depicted in each panel is indicated on the top right. Numbers at the top of each track indicate the maximum peak height.

Figure 2. Absence of γδ T cells amplifies iNKT cell population in Id3-/- mice.

(a) Representative distribution of iNKT (CD1dTet+) versus γδ T cells in 20 day old B6 (n=4), Id3-/- (n=4), Id3-/-TCRδ-/- (n=4) and L-DKO (n=4) mice. (b) Percentage and (c) numbers of iNKT cells in 3-5 weeks old B6 (n=5), Id3-/- (n= 7), Id3-/-TCRδ-/- (n=13) and L-DKO (n=14) mice. (d) Percentage and (e) numbers of stage 0 iNKT cells in 2-5 weeks old B6 (n=3), Id3-/- (n=3) and Id3-/-TCRδ-/- (n=5) mice. (f) Representative distribution of iNKT (CD1dTet+) versus γδ T cells in 13 day old B6 (n=3), TCRδ-/- (n=3), Id3-/- TCRδ-/- (n=5) and L-DKO (n=4) mice. (g) Percentage of iNKT cells in 13 day old mice shown in (f). Error bars represent SEM. Statistical significance is represented by p values (* < 0.05, ** < 0.005, *** < 0.0005, n.s > 0.05).
Figure 3. Transcriptional dysregulation in iNKT and γδNKT cells in the absence of Id proteins.

(a) Principal Component Analysis of RNA-Seq expression data. (b) Hierarchical clustering analysis based on RNA-Seq expression patterns. (c) Genes found to be differentially expressed (fold change > 2) in L-DKO DP, L-DKO iNKT and Id3−/− γδ samples, as compared to their WT counterparts. Numbers indicate unique or shared gene dysregulation between samples. (d) Expression patterns of differentially expressed genes identified in (c) in L-DKO iNKT and Id3−/− γδ T cells compared to WT iNKT and WT γδ T cells respectively. Genes with upregulation or downregulation in both cell types are highlighted. (e) Pathway analysis of genes that were more than 2 fold dysregulated in L-DKO iNKT or Id3−/− γδ T cells. Top pathways from hallmark and curated gene sets are displayed with the percentage of genes that were shared with the respective pathways, along with the p values. Gene sets representing upregulation or downregulation of genes in T-to-natural killer (ITNK) cells upon Bcl11b deletion, as curated by Msigdb (gene sets M2088 and M2086), are highlighted in red.

Figure 4. E2A regulates innate-like transcriptional programs in the absence of Id proteins.

(a) Flowchart representing the analysis pipeline used to derive the E2A-driven innate-like T cell program depicted in (c). (b) Expression patterns (fold change in mutant cells compared to WT) of genes in the 3 groups shown in (c), based on more than 2 fold dysregulation in one or more cell types, and part of the innate-gene signature derived from WT Immgen data. (c) Network analysis combining RNA-Seq expression and E2A ChIP-Seq data for innate-like signature genes. Genes in the “γδNKT” or “iNKT-biased” groups are colored according to their expression in Id3−/− γδ T cells and L-DKO iNKT cells respectively. Genes in the “common” group are
represented by their average expression in d3γδ T cells and L-DKO iNKT cells. ChIP-Seq binding of E2A (encoded by Tcf3 gene) to gene targets is represented by green lines. Other interactions between gene targets, classified by GeneMania as co-expression, physical interaction or co-regulation, are represented by orange, pink and purple lines respectively.

**Figure 5. Lack of Id proteins promotes iNKT cell and γδNKT development independent of pre-TCR signaling.**

(a) Representative iNKT cells in 3-5 week old Cre- (n=4), pTα-/- (n=10) and L-DKO pTα-/- (n=11) mice. (b) Percentage and (c) numbers of iNKT cells in 3-5 week old mice. (d) Representative PLZF expression levels in iNKT cells from 3-5 week old Cre- (n=3), L-DKO (n=3) and L-DKO pTα-/- (n=4) mice. (e) Total number of γδ T cells in 3-5 week old Cre- (n=4), pTα-/- (n=6), L-DKO pTα-/- (n=10) mice. (f) Representative γδNKT (Vγ1.1Vδ6.3+) populations in 3-5 week old pTα-/- (n=5), L-DKO pTα-/- (n=7) mice. (g) Representative PLZF expression levels in γδ T cells from 3-5 week old Cre- (n=3), L-DKO (n=3) and L-DKO pTα-/- (n=4) mice. (h) Percentages and (i) numbers of γδNKT cells in mice shown in (f).

**Figure 6. Adoption of innate-like transcription signature by DP cells in the absence of pre-TCR signaling and conventional T cells.**

(a) Representative thymocytes in 3-5 week old Cre-, pTα-/- and L-DKO pTα-/- mice as shown by CD4 and CD8 staining. (b) Total number of thymocytes in in 3-5 week old mice. (c) Number of DP cells in mice shown in (a). 5 to 13 mice included for each genotype shown in (a-c). (d, e) Representative PLZF expression levels in (d) DP cells and (e) total live cells from 3-5 week old Cre- (n=3) and L-DKO pTα-/- (n=4) mice. (f) Gene expression patterns (log2 fold change with respect to WT DP cells) for innate-like signature genes that are dysregulated by at least 2 fold in
L-DKO or L-DKO pTα<sup>+</sup> DP cells compared to WT DP cells. (g) Enriched gene sets in L-DKO pTα<sup>+</sup> DP cells over L-DKO DP cells pertaining to innate-like gene signature (top left), iNKT development and maturation (top right) and inflammatory responses (bottom), using Gene Set Enrichment Analysis (GSEA). (h) Gene expression patterns (log2 fold change in L-DKO or L-DKO pTα<sup>+</sup> DP cells with respect to WT DP cells) for genes that are positively (red) or negatively (blue) correlated with Zbtb16 expression. Numbers indicate number of positively or negatively correlated genes in each quadrant or section. Error bars represent SEM. Statistical significance is represented by p values (* < 0.05, ** < 0.005, *** < 0.0005, n.s > 0.05).
Figure 2

(a) Flow cytometry plots showing CD1d tetramer expression on TCRγδ cells in different genotypes: B6 Control, Id3⁻/⁻, Id3⁻/⁻ TCRδ⁻/⁻, and L-DKO.

(b) Bar graph showing the percentage of INKT cells in different genotypes: B6, Id3⁻/⁻, Id3⁻/⁻ TCRδ⁻/⁻, and L-DKO.

(c) Bar graph showing the number of INKT cells in different genotypes: B6, Id3⁻/⁻, Id3⁻/⁻ TCRδ⁻/⁻, and L-DKO.

(d) Bar graph showing the percentage of total cells in different genotypes: B6, Id3⁻/⁻, Id3⁻/⁻ TCRδ⁻/⁻.

(e) Bar graph showing the number of cells in different genotypes: B6, Id3⁻/⁻, Id3⁻/⁻ TCRδ⁻/⁻.

(f) Flow cytometry plots showing CD1d tetramer expression on TCRδ cells in different genotypes: B6 Control, TCRδ⁻/⁻, Id3⁻/⁻ TCRδ⁻/⁻, and L-DKO.

(g) Bar graph showing the percentage of INKT cells in different genotypes: B6, TCRδ⁻/⁻, Id3⁻/⁻ TCRδ⁻/⁻, and L-DKO.
RNA-Seq > 2 fold change
L-DKO iNKT & Id3/- γδ cells
3060 genes

Filter

Innate-like T cell gene signature
Immgen (189) + literature (7)

E2A-driven innate-like T cell program
111 genes
(83 direct targets)

Fig. 4
Fig. 6

(a) Flow cytometry plots showing the CD4 and CD8 expression levels in Cre- Control, pTa-/-, and L-DKO pTa-/- conditions. The graphs compare thymocyte counts.

(b) Bar graph illustrating the thymocyte counts: Cre-, pTa-/-, and L-DKO pTa-/-.

(c) Graph showing the number of DP cells per 10^6 cells for pTa-/- and L-DKO pTa-/- conditions.

(d, e) Flow cytometry histograms for DP cells and Live cells showing the distribution of PLZF expression in Cre-, Cre+iNKT, and L-DKO pTa-/- conditions.

(f) Heatmap displaying the enrichment of genes with INNATE_GENE_SIGNATURE, MATSUDA_NATURAL_KILLER_DIFFERENTIATION, and HALLMARK_INFLAMMATORY_RESPONSE.

(g) Enrichment plots for the signature genes.

(h) Scatter plot showing the Log2FC of L-DKO pTa-/- vs WT and L-DKO vs WT conditions.