

UC San Diego

UC San Diego Previously Published Works

Title

Innate-like functions of natural killer T cell subsets result from highly divergent gene programs

Permalink

<https://escholarship.org/uc/item/5k54s990>

Journal

Nature Immunology, 17(6)

ISSN

1529-2908

Authors

Engel, Isaac
Seumois, Grégory
Chavez, Lukas
[et al.](#)

Publication Date

2016-06-01

DOI

10.1038/ni.3437

Peer reviewed



Published in final edited form as:

Nat Immunol. 2016 June ; 17(6): 728–739. doi:10.1038/ni.3437.

Innate-like functions of natural killer T cell subsets result from highly divergent gene programs

Isaac Engel^{1,6}, Grégory Seumois^{1,6}, Lukas Chavez^{1,5,6}, Daniela Samaniego-Castruita¹, Brandie White¹, Ashu Chawla¹, Dennis Mock², Pandurangan Vijayanand^{1,3,7}, and Mitchell Kronenberg^{1,4,7}

¹La Jolla Institute for Allergy & Immunology, La Jolla, California, USA

²David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA

³Clinical and Experimental Sciences, Southampton NIHR Respiratory Biomedical Research Unit, Faculty of Medicine, University of Southampton, Southampton, UK

⁴Division of Biological Sciences, University of California San Diego, La Jolla, California, USA

Abstract

Natural killer T cells (NKT cells) have stimulatory or inhibitory effects on the immune response that can be attributed in part to the existence of functional subsets of NKT cells. These subsets have been characterized only on the basis of the differential expression of a few transcription factors and cell-surface molecules. Here we have analyzed purified populations of thymic NKT cell subsets at both the transcriptomic level and epigenomic level and by single-cell RNA sequencing. Our data indicated that despite their similar antigen specificity, the functional NKT cell subsets were highly divergent populations with many gene-expression and epigenetic differences. Therefore, the thymus ‘imprints’ distinct gene programs on subsets of innate-like NKT cells that probably impart differences in proliferative capacity, homing, and effector functions.

Invariant natural killer T cells (iNKT cells) have invariant rearrangement of the gene encoding the T cell antigen receptor (TCR) α -chain and they recognize glycolipid antigens

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

Correspondence should be addressed to P. V. (vijay@lji.org) or M.K. (mitch@lji.org).

⁵Present address: Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

⁶These authors contributed equally to this work.

⁷These authors jointly directed this work.

Accession codes. GEO: sequencing data, GSE74597.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS

I.E., G.S., P. V. and M.K. conceived of the work, designed, performed and analyzed experiments, and wrote the paper; I.E. performed all cell isolation, iNKT subset phenotyping and functional experiments with wild-type and HS V β -deficient mice; G.S. performed ‘micro-scaled’ RNA-Seq, ChIP-Seq and single-cell RNA-Seq experiments; D.S.-C. and A.C. assisted in the analysis of bulk and single-cell RNA-Seq data under the supervision of L.C.; B.W. assisted in single-cell sequencing experiments under the supervision of G.S. and P.V.; and D.M. provided statistical input and support and direction for the single-cell RNA-Seq analysis.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

presented by CD1d¹. A striking property of these cells is their ability to produce copious amounts of cytokines rapidly after TCR stimulation. A puzzling feature, however, is that in some cases λ NKT cells stimulate immunity and inflammation, while in others they are anti-inflammatory². One contributor to their divergent effects is probably the selective activation of functional subsets of λ NKT cells. CD4⁺ T lymphocyte subsets reactive to major histocompatibility complex class II have been categorized by the expression of signature transcription factors (TFs) and cytokines. The subsets of CD4⁺ T lymphocytes also have been reported for λ NKT cells; prominent among these are NKT1, NKT2 and NKT17 cells, analogous to the T_H1, T_H2 and T_H17 subsets of helper T cells, respectively³. An important difference, however, is that in mice these λ NKT cell subsets expressing the TCR α -chain variable region 14 (V α 14) differentiate in the thymus. Because there is relatively little information on the molecular bases for the formation and function of λ NKT cell subsets, here we have taken a comprehensive approach to define the transcriptional and epigenetic profiles of functional subpopulations of λ NKT cells.

RESULTS

Identification of λ NKT cell subsets by cell-surface markers

NKT1, NKT2 and NKT17 cells have been previously defined (Fig. 1a). To identify the global gene-expression and chromatin profiles of V α 14 λ NKT thymocyte subsets, we performed RNA sequencing (RNA-Seq), as well as chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) for the modification H3K27ac (histone H3 acetylated at Lys27), in purified populations (Supplementary Table 1). We defined the subsets on the basis of the expression of surface proteins^{4–6} (Online Methods and Supplementary Fig. 1) and demonstrated that cells isolated via these surface markers exhibited the expected expression pattern of signature TFs: NKT1 cells were T-bet⁺PLZF^{lo}ROR γ t[–]; NKT17 cells were ROR γ t⁺T-bet[–]PLZF^{int}; and the majority of NKT2 cells were PLZF^{hi}ROR γ t[–]T-bet[–] (Fig. 1b).

Bulk RNA-Seq revealed that the expression pattern of genes previously linked to the λ NKT cell subsets was consistent with those in published reports^{4,5}. For examples, genes encoding the key TFs T-bet (*Tbx21*), PLZF (*Zbtb16*) and ROR γ t (*Rorc*) were ‘preferentially’ expressed in the NKT1 subset, NKT2 subset and NKT17 subset, respectively (Fig. 1c, top). A similar subset-specific pattern was observed for cytokine-encoding genes, such as *Ifng*, *Il4*, *Il13* and *Il17a*, and for the cytokine-receptor-encoding genes *Il2rb*, *Il17rb* and *Il23r* (Fig. 1c, middle and bottom). We also found other examples of subset-specific gene expression (Supplementary Fig. 2). Enhancer profiles, identified as regions in these loci showing greater enrichment for H3K27ac than its abundance in other regions in the locus, generally were concordant with the gene-expression pattern, although in some cases, such as *Il17f* chromatin in NKT17 cells, chromatin-activation marks were present in the absence of detectable transcripts (Fig. 1c). This probably reflected chromatin that was poised for transcription but not actively expressed. Together these data suggested that our sorting strategy reliably identified functional subsets of λ NKT cells.

Subset heterogeneity revealed by single-cell analysis

Published evidence of the transcriptomes of single cells has revealed that individual cells within the ‘same’ population may differ substantially and that these differences might have important consequences for understanding the origin and functions of the population⁷. To capture the extent of cellular heterogeneity within a given λ NKT cell subset, we performed single-cell RNA-Seq of the three thymic λ NKT cell subsets and a precursor population⁸, CD24⁺ NKT0 cells. For each population, at least 44 cells were analyzed. Principal-component analysis (PCA) showed that single cells from each subset clustered together and formed spatially separated groups that reflected their highly divergent transcriptional states (Fig. 1d and Supplementary Video 1). The NKT2 subset appeared nearest to NKT0 cells, which suggested that this population included precursor cells, and it also displayed the greatest heterogeneity, which suggested that it included several differentiation states (Fig. 1d and Supplementary Video 1). Unsupervised hierarchical clustering of single cells based on the expression of 1,500 genes with the highest variance, which accounted for 34% of the total variance, also showed distinct separation of the populations (Supplementary Fig. 3a). Additionally, PCA of the bulk RNA-Seq data indicated three clearly distinct λ NKT cell populations (Supplementary Fig. 3b).

To further characterize the gene-expression program in λ NKT cell subsets, we performed pairwise comparison of bulk RNA-Seq data from NKT1 cells versus NKT2 cells, NKT2 cells versus NKT17 cells, and NKT1 cells versus NKT17 cells (Fig. 2a, Supplementary Fig. 4 and Supplementary Table 2). We classified the ‘differentially expressed genes’ into groups as follows (Supplementary Tables 2–4): ~276 transcripts with higher expression in the NKT1 subset than in the NKT2 or NKT17 subset (‘NKT1-enriched’), ~329 transcripts with higher expression in the NKT2 subset than in the NKT1 or NKT17 subset (‘NKT2-enriched’), and ~260 transcripts with higher expression in the NKT17 subset than in the NKT1 or NKT2 subset (‘NKT17-enriched’)⁴ (Fig. 2a and Supplementary Table 2). Bulk ChIP-Seq analysis showed that a total of ~10,000 *cis*-regulatory regions had differential enrichment for H3K27ac in the λ NKT cell subsets (NKT1 versus NKT2, NKT2 versus NKT17, and NKT1 versus NKT17)⁹ (Supplementary Fig. 5 and Supplementary Table 5), indicative of major changes in the genome-wide enhancer profile of the subsets.

The single-cell transcriptome profiles of a total of 203 λ NKT cells from the various subsets also showed significant differences in gene expression (Fig. 2b; genes with the most significantly differential expression, Supplementary Table 6). The majority of genes detected as being ‘subset-enriched’ by bulk RNA-Seq analysis were expressed in a greater fraction of the single cells in the corresponding subset and had higher expression on a per-cell basis (Fig. 2a–c and Supplementary Fig. 6). Within a subset, however, heterogeneity in gene expression was observed for some genes (Supplementary Figs. 7–9), including cases in which only a minority of the cells in a subset had high expression of a subset-specific transcript. Examples of this intra-subset heterogeneity included *Irf4* in NKT2 cells (Supplementary Fig. 2 (bulk sequence data) versus Supplementary Fig. 7 (single-cell data)), *Ccr2* in NKT17 cells (Supplementary Fig. 2 versus Supplementary Fig. 8), and *Ccl5* in NKT1 cells (Supplementary Fig. 2 versus Supplementary Fig. 9). Therefore, despite the heterogeneity found at the single-cell level, the results of bulk and single-cell RNA-Seq

analysis were consistent in showing three very distinct transcriptomes in $\mathcal{N}$ KT cell subsets, consistent with the major changes in their enhancer profiles.

Differences in homing molecules

Most $\mathcal{N}$ KT cells are tissue resident and noncirculating^{10,11}, with NKT1 cells being dominant in the liver, and the skin and other sites showing enrichment for NKT17 cells³. Consistent with their differential localization, the subsets differed in their expression of chemokine receptors (Fig. 2d,e and Supplementary Fig. 2). Transcripts encoding the chemokine receptor CCR6 (*Ccr6*) were exclusively present in the single NKT17 cells, although they were not in each of the NKT17 cells, despite the use of a monoclonal antibody to CCR6 to sort the population (Fig. 2d,e). That finding was consistent with the fact that even genes with high expression can be either ‘on’ or ‘off’ for several hours, due to the well-described ‘bursting’ phenomenon of gene expression in single cells¹². As for the differential expression of chemokine receptors, transcripts encoding CXCR3 (*Cxcr3*) and CCR5 (*Ccr5*) were present mainly in NKT1 cells, and transcripts encoding CCR4 (*Ccr4*) and CCR9 (*Ccr9*) were present mainly in NKT2 cells (Fig. 2 and Supplementary Fig. 2), and results obtained by flow cytometry supported those mRNA findings (Supplementary Fig. 10). The expression of integrin α - and β -subunits also showed subset ‘preferences’. NKT1 cells showed enrichment for transcripts encoding the α -subunit (α_1) and β -subunit (β_1) of the collagen-binding integrin VLA-1 (*Itga1* and *Itgb1*) (Supplementary Fig. 11), consistent with flow cytometry data indicating CD49A (the α_1 integrin subunit) was a marker for this subset (Supplementary Fig. 10). Integrins with a β_1 subunit (encoded by *Itgb1*) have been linked to the accumulation of CD8⁺ T cells in the liver, a site in which NKT1 cells are most abundant¹³. In contrast, *Itgb7* (encoding the β_7 integrin subunit), *Itgb4* (encoding the β_4 integrin subunit) and *Itgb5* (encoding the β_5 integrin subunit) had higher expression in NKT17 cells (Supplementary Fig. 11). Therefore, the $\mathcal{N}$ KT cell subsets expressed different molecules involved in migration and homing, which would provide a potential explanation for their differential tissue localization^{3,10,11,14}.

Precursors among NKT2 cells

The definition of NKT2 cells, based in part on the absence of expression of the activating NK cell receptor NK1.1, includes at least some of what also have been called ‘stage 2 $\mathcal{N}$ KT cells’, a population that might include precursors of stage 3, mature, NK1.1⁺ $\mathcal{N}$ KT cells³. Therefore, we sought to identify potential precursor cells among NKT2 thymocyte populations. Pathway analysis of ‘NKT2-enriched’ transcripts from the bulk RNA-Seq data revealed over-representation of genes encoding products involved in cell-cycle control, mitosis, signaling via the tumor suppressor p53, and DNA replication (Supplementary Table 4), which indicated that the NKT2 cells were actively undergoing cell division (Supplementary Fig. 12). Similar enrichment in NKT2 cells for the expression of transcripts encoding products involved in cell-cycle control was evident for the 59 such genes analyzed by single-cell RNA-Seq (Fig. 3a and Supplementary Fig. 3a). PCA of the single-cell RNA-Seq data (Fig. 1d and Supplementary Video 1) indicated that the NKT2 population could be divided into two clusters, with one showing expression of genes encoding products involved in the cell cycle (Fig. 3b). This cluster constituted 47% of the NKT2 cells; a smaller proportion of NKT17 and NKT0 cells (<15%) expressed such genes, and their expression

was absent in NKT1 cells (Fig. 3a,b). Overall, these data were consistent with published observations showing that NK1.1⁺ α NKT thymocytes, which are mostly NKT1 cells, are quiescent^{15–17}.

Although the proliferating NKT2 cells did not uniformly have markers of decreased maturity, interestingly, by single-cell RNA-Seq, a subset of NKT2 cells, the majority of which did express genes encoding products involved in the cell cycle, did express *Tbx21* (Supplementary Fig. 9). Some NKT2 cells contained T-bet protein (Fig. 1b), and a fraction of the T-bet⁺ NKT2 cells also expressed the T-bet target gene *Cxcr3*, although they retained expression of the NKT2 cell markers IL-17RB and ICOS (Supplementary Fig. 13). Therefore, we hypothesized that the T-bet⁺ NKT2 cells might represent a transitional population in the process of becoming NKT1 cells. A published study using *Il13* fate mapping indicated that all α NKT cells were derived from a precursor cell with *Il13* expression¹⁸. Although the bulk RNA-Seq data indicated that the abundance of *Il13* mRNA was greater in NKT2 cells, at the single-cell level, *Il13* transcripts were not present in NKT0 cells and were present in only one of the NKT2 cells (Supplementary Fig. 7b). This highly uneven expression in NKT2 cells raised questions about the amount and timing of *Il13* mRNA expression in α NKT thymocytes. Therefore, while several lines of evidence suggested that there might be a precursor of other subsets in the NKT2 population, definitive evidence showing that all α NKT cells pass through a true NKT2 stage is still lacking.

Genes related to NKT2 differentiation and function

Several genes of interest selectively and broadly expressed by NKT2 cells were revealed by both single-cell RNA-Seq (Fig. 3) and bulk RNA-Seq (Supplementary Fig. 12 and Supplementary Table 2). We identified by single-cell sequencing the 50 genes that were most differentially regulated in NKT2 cells (Fig. 3c and Supplementary Table 7). *Il6ra* (which encodes the cytokine receptor IL-6R α (CD126)) was expressed exclusively in NKT2 cells, as determined by bulk sequencing (Fig. 3d and Supplementary Table 2) and by single-cell sequencing (Fig. 3e and Supplementary Fig. 14), a result confirmed by flow cytometry (Supplementary Fig. 15). Signaling via IL-6R has been shown to induce expression of the transcription factor NFATc2 and its translocation to the nucleus and thus direct differentiation of naive CD4⁺ T cells into IL-4-producing effector T_H2 cells, even in the absence of canonical *in vitro* T_H2-polarizing signals¹⁹. Hence, it is possible that IL-6R signaling might be important for the differentiation of thymic NKT2 cells and for their high probability of *Il4* expression. Notably, the IL-6 pathway also induces expression of the cytokine-signaling suppressor SOCS1, which in turn inhibits signaling via interferon- γ (IFN- γ) and the T_H1 polarization of naive CD4⁺ T cells¹⁹. Thus, we hypothesized that the selective expression of *Il6ra* in thymic α NKT cells might restrain the NKT1 differentiation pathway and favor NKT2 differentiation.

NKT2 cells have been linked to mouse models of airway inflammation²⁰. We found enrichment in NKT2 cells for the transcripts *Plac8* (Fig. 3d,e) and *Fbln1* (Fig. 3e and Supplementary Fig. 12), which suggested the hypothesis that the proteins they encode might contribute to NKT2 cell-mediated lung inflammation. *Plac8*, whose expression showed the greatest enrichment in NKT2 cells (Fig. 3c and Supplementary Table 7), encodes PLAC8, a

cysteine-rich, secreted protein linked to defense against bacterial pathogens²¹ that is a key effector molecule for eliciting inflammatory contact hypersensitivity²², although a role for this molecule in asthma has not been demonstrated. *Fbln1* encodes fibulin-1, a cysteine-rich calcium-binding extracellular matrix molecule that has been associated with airway remodeling and asthma^{23,24}. Therefore, our transcriptomic analyses identified molecules potentially important for the differentiation and function of NKT2 cells, although further studies will be needed to verify this.

Genes driving NKT17 differentiation and function

RNA-Seq data indicated that several genes encoding products related to CD4⁺ T_H17 cell function, such as *Rorc*, *Ccr6*, *Il23r*, *Il17re* and *Il1r1* (ref. 25), had high expression in and were expressed almost exclusively in the NKT17 subset (Fig. 4 and Supplementary Figs. 2 and 16 (bulk RNA-Seq), and Supplementary Figs. 8 and 17 and Supplementary Table 8 (single-cell analysis)). Extensive enrichment for H3K27ac in NKT17 cells was in accordance with the RNA-Seq data (Figs. 1c and 4c and Supplementary Figs. 2 and 16). As for the NKT2 subset, expression of the T_H17 signature genes (*Rorc*, *Ccr6*, *Il23r*, *Il17re* and *Il1r1*) was not uniform in single NKT17 cells, not even that of *Rorc* (Fig. 4b). Interestingly, some genes encoding products not previously known to be associated with Λ NKT cell function, such as *Blk*, *Sox13*, *Maf*, *Serpinb1a*, *Aqp3*, *Sepp1* and others, showed nearly uniform high expression in NKT17 cells, with little or no expression in the other subsets (Fig. 4c,d and Supplementary Figs. 16 and 17).

Several transcripts that characterized the NKT17 subset are found in other IL-17-producing T cells. For example, *Blk* encodes a kinase of the Src family (Blk) that is important in B cell development²⁶. However, *Blk* deficiency has been shown to result in failure to generate IL-17-producing $\gamma\delta$ T cells²⁶, which would suggest that *Blk* might have a similar role in Λ NKT cells. The TFs SOX5 and c-MAF have been shown to work together to stimulate *Rorc* expression in CD4⁺ T_H17 cells²⁷, and the related SOXD-group protein SOX13 is required for the differentiation of IL-17-producing $\gamma\delta$ T cells²⁸. Notably, expression of *Maf* and *Sox13* transcripts (Fig. 4c,d and Supplementary Figs. 16 and 17), but not of *Sox5* transcripts (data not shown), was higher in NKT17 cells than in the other subsets, which would suggest that the mechanism for *Il17* expression in Λ NKT cells might resemble that in $\gamma\delta$ cells. *Serpinb1a* encodes a peptidase inhibitor (SERPINB1A) shown to be greater in abundance in T_H17 cells than in other helper T cell subsets, but SERPINB1A-deficient mice have enhanced population expansion of T_H17 cells and IL-17-producing $\gamma\delta$ T cells²⁹. Therefore, we hypothesized that SERPINB1A might be a negative feedback regulator of IL-17-producing T cells, including NKT17 cells. *Aqp3* also exhibited an NKT17-biased expression pattern (Fig. 4a,d and Supplementary Figs. 16 and 17). The product of *Aqp3* facilitates the transport of hydrogen peroxide, a process necessary for chemokine-mediated infiltration of CD4⁺ and CD8⁺ T cells into the skin and contact hypersensitivity³⁰. Because NKT17 cells are the principal subset of Λ NKT cells in the skin, we hypothesized that *Aqp3* might have a similar role in the localization of NKT17 cells to the dermis. In another possible example of feedback regulation, NKT17 cells also had increased expression of transcripts for the selenoprotein GPX1, which detoxifies hydrogen peroxide³¹ (Fig. 4d and Supplementary Fig. 16).

The expression of several other transcripts might be correlated with unique effector functions of NKT17 cells and other IL-17-secreting cells. NKT17 cells had increased expression of transcripts encoding another selenoprotein, SEPP1, a secreted protein that can affect pathogen clearance and inflammation^{32,33} and that has been shown to be ‘preferentially’ expressed in type 3 innate lymphoid cells (ILCs)³⁴ (Supplementary Table 9). Overall, there was a tendency for genes ‘preferentially’ expressed in an ILC subset to have increased expression in the corresponding Δ NKT cell subset, a trend that was observed for type 1 ILCs and NKT1 cells, as well as for type 3 ILCs and NKT17 cells (Supplementary Table 9). The similarities in subset-specific gene-expression programs therefore extended beyond T lymphocytes to include ILCs. NKT17 cells also had high expression of *Mmp25*, which encodes the metalloproteinase MT6-MMP (Fig. 4a,c,d and Supplementary Fig. 17). This protein can activate several chemokines that mediate neutrophil recruitment, but it also cleaves vimentin into a form that facilitates anti-inflammatory phagocytosis that removes neutrophils³⁵. On the basis of these findings and the function of IL-17 in the recruitment of neutrophils, we hypothesize that NKT17 cells might have both pro-inflammatory activities and anti-inflammatory activities related to neutrophil activity.

Cell survival- and cytotoxicity-related genes in NKT1 cells

Pathway analysis of ‘NKT1-enriched’ genes revealed significant over-representation of genes encoding products involved in regulating apoptosis (*Bcl2*, *Dapk2*, *Cdh1*, *Tnfrsf10*, *Tnfrsf1a*, *Tjp1*, *Gzmb*, *Ppp3cc* and *Camk2b*) and the pro-survival IL-2–IL-15 signaling pathway (*Il2rb*, *Stat4*, *Bcl2*, *Sgk1* and *Nr3c1*) (Supplementary Table 4), consistent with the long-term survival of NKT1 cells in the thymus. Moreover, seven genes encoding products of the GIMAP (‘GTPase-of-immunity-associated protein’) family, which have been linked to the regulation of apoptosis³⁶, showed increased transcription and marked locus-wide enrichment for H3K27ac in NKT1 cells (Fig. 5a). NKT1 cells also had increased expression of *Ms4a4b*, which encodes a product known to inhibit entry into the S-G2–M phases of the cell cycle and to prevent activation-induced cell death of T cells³⁷ (Fig. 5b,c and Supplementary Figs. 18 and 19). Consistent with published reports³⁸, we experimentally confirmed that NKT1 cells were indeed more resistant to apoptosis than were the other Δ NKT cell subsets (Supplementary Fig. 20). Together these data suggested that several anti-apoptotic and pro-survival mechanisms contributed to the abundance of NKT1 cells among Δ NKT cells in the C57BL/6 mouse thymus, although ‘preferential’ cell retention also might have contributed to this.

NKT1 cells had increased expression of *Socs2* transcripts, which encode a product (SOCS2) that suppresses the T_H2 responses of CD4⁺ T cells³⁹ and that might have a similar role in NKT1 cells. Pathway analysis also revealed increased expression of genes encoding products related to IL-12 signaling, such as *Il12rb2*, *Tbx21*, *Stat4*, *Ifng*, *Il18rap* and *Ccr5* (Fig. 5b,c, Supplementary Fig. 18 and Supplementary Table 10); NKT1 cells are known to respond to IL-12 (ref. 40). Population-wide and single-cell RNA-Seq data confirmed NKT1-subset-specific expression of many genes encoding NK cell receptors (Fig. 5c,d and Supplementary Fig. 19) and genes encoding products related to cytotoxicity, including *Gzma*, *Gzmb*, *Prf1* and *FasL*, as well as *Ctsw* and *Serpinb9*, which encode negative regulators of cytotoxicity (Supplementary Figs. 18 and 19), again suggestive of a possible

feedback mechanism. The restriction of these effector cells to the NKT1 subset indicated that cytolytic function was restricted mainly to this subset. Although NKT1 cells were the predominant cell type that expressed genes encoding members of the NK cell receptor family, NKT17 cells had increased expression of *Klrb1b* mRNA, which encodes the inhibitory receptor NKR-P1D, and *Klrb1f* mRNA, which encodes the activating receptor NKR-P1F (Fig. 5d and Supplementary Fig. 19). We also noted that the inhibitory receptor CD200 was expressed prominently in NKT2 cells (Fig. 3e and Supplementary Figs. 12 and 14). Therefore, we propose that each α NKT cell subset expresses particular inhibitory receptors needed to modulate autoreactivity. NKT1 cells also exclusively had high expression of XCL1 (Fig. 5b,d and Supplementary Fig. 19), a chemokine that has been shown to be important for the recruitment of CD8 α ⁺ dendritic cells expressing its receptor, XCR1, to sites of the immune response⁴¹. Therefore, we hypothesize that the ability to attract these antigen-presenting cells might be related to the ability of activated α NKT cells to promote the cross-presentation of antigens by dendritic cells⁴² and perhaps is related to the prevention of infections of thymic tissue. Therefore, our analysis indicated that the cytotoxic function of α NKT cells is probably confined to the NKT1 subset, although we hypothesize that inhibition of autoreactivity is a more general requirement that encompasses all subsets.

Distinct gene-expression program of NKT0 cells

NKT0 cells are a rare subset, but we were able to correctly assign differentially expressed, NKT0-subset-specific transcripts and to eliminate contaminating CD4⁺CD8⁺ (double-positive) thymocytes in the single-cell RNA-Seq, on the basis of expression of *Traj18* (which encodes the TCR α -chain joining region 18) and other transcripts (Supplementary Fig. 21). NKT0 cells displayed a gene-expression program that included genes encoding TFs, tetraspanins and other cell-surface proteins (Fig. 6, Supplementary Fig. 21 and Supplementary Table 11). Some genes that exhibited differential expression in NKT0 cells included two whose expression was dynamically regulated in positively selected cells: *Itm2a* (which encodes the integral membrane protein ITM2A) and *Ccr9* (Fig. 2e and Supplementary Fig. 7a). ITM2A expression is increased during the transition of thymocytes from CD4⁺CD8⁺ to CD4⁺ or CD8⁺ (single positive) and hence is a marker of T cell positive selection⁴³. CCR9 is expressed by CD4⁺CD8⁺ thymocytes and is important for thymocyte localization, but its expression is downregulated as cells mature⁴⁴. Other notable genes exhibiting differential expression in NKT0 cells were *Ldhb* and *Gsn* (Fig. 6 and Supplementary Fig. 21). *Ldhb* encodes lactate dehydrogenase b, a critical enzymatic activator of the glycolysis pathway that allows the utilization of lactate as an alternative nutrient source for oxidative metabolism⁴⁵. *Gsn* encodes gelsolin, a calcium-dependent regulatory protein involved in cytoplasmic actin-filament transitions, a process required for cellular locomotion and for the creation of a signaling complex at the immunological synapse⁴⁶. In agreement with a published report⁴⁷, expression of the high-mobility-group-box protein and transcription factor LEF1 was increased in NKT0 cells (Fig. 6 and Supplementary Fig. 21). LEF1 has been reported to be required for the transition of NKT0 cells to the other stages by inducing the expression of *Cd127* (which encodes the cytokine receptor IL-7R α (CD127)) and *Myc* (which encodes the transcription factor c-Myc)⁴⁷. Although NKT0 cells were present at an abundance of only approximately 1,000 cells per

thymus, NKT0 cells were shown by single-cell analysis to have elevated expression of transcripts characteristic of recently positively selected cells but also elevated expression of transcripts encoding products that support increased cellular metabolism, homing and motility, which might give insight into their function.

TF-binding motifs in *i*NKT cell subsets

During differentiation, lineage-driving TFs leave a ‘footprint’ on the enhancer landscape of a cell, often by increasing the abundance of H3K27ac at enhancers. Therefore, we carried out *in silico* analysis of enhancer profiles to predict transcriptional regulators that might contribute to the setting up of gene-expression programs in *i*NKT cell subsets⁹. As expected, enhancers that gained H3K27ac in NKT17 and NKT1 cells showed significant enrichment, by ChIP-Seq analysis, for binding motifs for ROR γ t and T-bet, respectively (Fig. 7a and Supplementary Table 12). ROR γ t-binding motifs were present in nearly 10% of all the enhancers that gained H3K27ac specifically in NKT17 cells, and they displayed the greatest enrichment over background relative to that of any other TF-binding motif ($P < 1 \times 10^{-40}$), which indicated that ROR γ t was in fact the dominant TF driving NKT17 differentiation, although there was evidence for the involvement of other TFs, such as Revrb (Fig. 7a and Supplementary Table 12). Enhancers in NKT1 cells showed enrichment for binding motifs for the transcription factors RUNX and Eomes, although *Eomes* transcripts were not detectable (Fig. 7a and Supplementary Table 12), which suggested low or transient expression or the activity of a transcription factor with a similar motif. Enhancers that gained H3K27ac in NKT2 cells showed enrichment for binding motifs for both the GATA family and the IRF family of TFs (Fig. 7a and Supplementary Table 12). In addition, enhancers in NKT2 cells showed substantial enrichment for binding motifs for the KLF, SOX and TCF families of TFs (Fig. 7a and Supplementary Table 12). KLF13 and KLF2 have been linked to the development of NKT2 cells in BALB/c mice and C57BL/6 mice, respectively¹⁷; however, among genes encoding the KLF family of TFs, we found that only *Klf10* expression was higher in NKT2 cells than in the other subsets (Fig. 7b), which suggested that KLF10 might be another, previously unknown participant in NKT2 differentiation. Therefore, our analysis of binding motifs associated with active enhancers identified candidate TFs driving the differentiation and/or function of *i*NKT cell subsets.

An enhancer essential for *i*NKT cell IL-4 expression

Comparison of the enhancer profiles of different cell types can aid in the identification of enhancers that are critical for cell-type-specific gene expression. We took this approach to predict enhancers that increase the probability of *Il4* expression in thymic NKT2 cells. Expression of *Il4* mRNA was increased in NKT2 cells in the bulk sequencing data and in single cells (Fig. 3e and Supplementary Table 7), and among the enhancers in the extended *Il4* locus, only the hypersensitivity site (HS) V enhancer showed an increase in H3K27ac selectively in NKT2 cells relative to the abundance of this mark at the HS V enhancer in NKT1 cells (Fig. 7c). Follicular helper T cells (T_{FH} cells) and T_H2 cells, but not basophils or eosinophils, utilize the HS V enhancer to regulate *Il4* transcription but, notably, HS V is also non-redundant only in T_{FH} cells⁴⁸. We observed a nearly complete loss of innate memory-like CD8⁺ cells that expressed Eomes in the thymus of HS V-deficient BALB/c mice (Fig. 7d). Given the established role of thymic-*i*NKT-cell IL-4 in promoting the development of

these thymocytes⁴, we reasoned that like T_{FH} cells, λ NKT cells might require the HS V enhancer for *Il4* transcription. Splenic λ NKT cells from HS V-deficient mice produced significantly less IL-4 after antigenic stimulation (Fig. 7e), in support of the hypothesis that the HS V enhancer has a critical role in regulating IL-4 production by thymic and peripheral λ NKT cells.

DISCUSSION

Our ‘micro-scaled’ RNA-Seq analyses revealed an extensive and unexpected diversity in the genome-wide transcriptional program of three cytokine-polarized thymic λ NKT cell subsets. Furthermore, nearly 10,000 potential enhancer regions showed gain or loss of H3K27ac, indicative of a stable gene-expression program preset in the epigenome of λ NKT cell subsets in the thymus. We did not find unequivocal evidence for additional thymic λ NKT cell subsets, such as NKT_{TH} cells⁴⁹ or NKT10 cells⁵⁰, even at the single-cell level. Our cell-sorting strategy might have excluded these additional subsets or, alternatively, they might be induced in the periphery only after antigenic stimulation.

We performed single-cell RNA-Seq analysis because such analysis might reveal fundamental biological principles and it addresses the substantial challenges in systematically describing allegedly homogenous cell populations⁷. Broadly, most of the subset-enriched genes from bulk RNA-Seq analysis showed one of three distinct single-cell gene-expression patterns: many of the λ NKT cell subset signature genes (for example, *Rorc*, *Il23r*, *Nkg7* and *Klrd1*) showed a nearly exclusive single-cell expression pattern in their corresponding λ NKT cell subset, although not necessarily in every cell of that subset. In contrast, several subset-enriched genes (for example, *Ccr4*, *Cxcr3*, *Tbx21*, *Il4* and *Ccr2*) showed overlapping expression in other subsets, although with a greater fraction of cells in one subset expressing the target gene. A third pattern was observed for several subset-specific or subset-enriched genes, particularly those encoding cytokines and chemokines (for example, *Ifng*, *Il13*, *Il10*, *Il21*, *Ccl5* and *Cxcl10*), which showed very high expression in only a small number of the cells (<10–15%) in the corresponding λ NKT cell subset. An important concern about single-cell RNA-Seq analysis is the possibility of false-negative results. The published sensitivity of single-cell RNA-Seq assays ranges from one mRNA molecule per cell to ten mRNA molecules per cell⁵¹, so we potentially failed to detect some transcripts. However, most of the genes that showed subset enrichment had high expression and thus were not likely to have been missed due to low expression.

Notably, the single-cell sequencing methodology allowed us to identify heterogeneity, most prominently in NKT2 cells, on the basis of the expression of genes encoding products related to the cell cycle. We readily distinguished genes, such as *Il4*, *Plac8*, *Il6ra*, *Cd200* and *Tnfsf11*, whose expression truly defined molecular features of the NKT2 subset regardless of cell-cycle state. The single-cell analysis also allowed us to demonstrate that cells that expressed *Tbx21* and *Cxcr3* also expressed cell-cycle-related genes to a considerable extent, although they did not account for all the cycling NKT2 cells. Although further experiments will be needed to demonstrate that the mixed-phenotype NKT2-NKT1 cells are precursors, the coordinated expression of cell-cycle-related transcripts and, to some extent, typical NKT1 cell transcripts would be more consistent with the existence of a true transitional

population, rather than ‘transcriptional bursting’, which would show a more random expression pattern in comparisons of different genes.

Our study combined single-cell RNA-Seq, ‘micro-scaled’ RNA-Seq and H3K27ac ChIP-Seq assays to comprehensively elucidate the differentiation and function of thymic α NKT cell subsets. We demonstrated these subsets were substantially different in their expression of not only TFs and cytokine but also molecules involved in homing, inhibitory receptors that presumably help to modulate autoreactivity, and also effector molecules that provide host defense and the regulation of inflammation. The means by which cells in a population with highly similar specificities might attain such different fates remain to be determined. Regardless of that, the candidate molecules described here provide new insights into α NKT cell biology, and the set of α NKT-cell-subset-enriched transcripts defined here will probably allow better characterization of peripheral α NKT cell subsets in the context of infection, cancer, autoimmune and allergic diseases.

ONLINE METHODS

Mice

C57BL/6J and BALB/cJ mice were purchased from Jackson Laboratories. HS V-deficient mice on a BALB/cJ background have been described previously⁴⁸. All studies were approved by the Institutional Animal Care and Use Committee at the La Jolla Institute for Allergy and Immunology.

In vivo antigen challenge

α -galactosyl ceramide (α -GalCer or KRN7000) was supplied as a lyophilizate in vehicle and provided by Kirin Hakko Kirin California. It was resuspended at 200 μ g/ml in water and then diluted to 10 μ g/ml in PBS. 0.2 ml of 10 μ g/ml α -GalCer was injected i.p., and mice were sacrificed and organs harvested 90 min later.

α NKT cell subset isolation

For RNA isolation from thymic α NKT cell subsets, both as bulk populations and for single-cell analysis, thymus cell suspensions prepared from five female C57BL/6J mice (approximately 5 weeks of age) were enriched for α NKT cells by negative selection of cells using biotinylated antibodies against CD8b.2 (53–5.8, BD Biosciences), CD19 (1D3, Tonbo Biosciences), and TER-119 (TER-119, Tonbo Biosciences) together with Rapidspheres (StemCell technologies #19860) and either the Big Easy (StemCell technologies #18001) or Easy eight magnets (StemCell technologies #18103) and protocols from Stem Cell Technologies. The remaining cells were then suspended at 10^8 /ml, incubated with 1 μ g/ml of Streptavidin A (Sigma-Aldrich) and stained using an 11 -parameter panel of reagents that included tetramers of CD1d loaded with either α -GalCer or PBS57 combined with streptavidin-brilliant violet (BV)421 (produced in-house or obtained from the NIH Tetramer Core Facility), anti-CD44-V500 (IM7) and anti-NK1.1-PE-Cy7 (PK136) (both from BD Biosciences), live/dead yellow (#L34959) and anti-CD4-Qdot 605 (RM4-5) (both from ThermoFisher Scientific), anti-CD103-FITC (2E7), anti-CD24-PerCPCy5.5 (M1/69), anti-CD27-APC (LG.7F9), and TCR β -APC-eF780 (H57-597) (all from Affymetrix), and anti-

CD8 α -AF700 (53–6.7) and anti-CCR6-PE (29-2L17) (both from BioLegend). λ NKT subsets were sorted using a FACS Aria III (BD Biosciences), selecting V α 14 λ NKT cells, defined as CD8 α ⁺, CD1d- α -GalCer tetramer-binding, TCR β ^{int} into NKT1, NKT2 and NKT17 subsets according to previously published surface expression profiles. These are as follows: NKT1, CD24^{lo}TCR β ^{lo}NK1.1^{hi}CD27^{hi}CCR6⁺; NKT2, CD24^{lo}TCR β ^{hi}NK1.1⁺CD27^{hi}CD4^{hi}; NKT17, CD24^{lo}CD27^{lo}CD4⁺CCR6⁺CD103^{hi} (refs. 4–6) (Supplementary Fig. 1). Cells were lysed in Trizol LS (Life Technologies) and RNA prepared using protocols and reagents from Life Technologies and Qiagen. For the single-cell sort, we started with three thymuses, and also collected, in parallel with the other λ NKT subsets, a population of CD8 α ⁺, CD1d- α -GalCer tetramer-binding, TCR β ^{int}, CD24^{hi}, CD44^{lo}, NK1.1⁺ cells designated as 'NKT0', which probably represent a very early stage of λ NKT cell development⁵². For collecting λ NKT cell subsets for ChIP-Seq analysis, the protocol above was used with the following modifications: 15 thymi were used per sort, biotinylated anti-CD24 (M1/69, BD Biosciences) was included in the enrichment protocol together with the other biotinylated antibodies listed above, anti-ICOS-PerCP-Cy5.5 (C398.4A, BioLegend) replaced anti-CD24-PerCP-Cy5.5 in the staining panel, NKT1 cells were defined as NK1.1^{hi}ICOS^{lo}CD27^{hi}CCR6⁺, NKT2 cells were defined as ICOS^{hi}NK1.1⁺CD27^{hi}CD4^{hi}, and NKT17 cells were defined as CD27^{lo}CD4⁺CCR6⁺CD103^{hi}. ICOS was found by both RNA-Seq and flow cytometry analyses to be expressed at much higher levels in NKT2 and NKT17 cells than in NKT1 cells (Supplementary Fig. 13, Supplementary Tables 2 and 5 and data not shown). Each sort was used to collect all of the populations in parallel. CD 1d tetramers and live/dead yellow were used at a dilution of 1:200, while all antibodies were used at a concentration of 0.5–1 μ g/ml.

Flow cytometry

All cell-staining experiments were performed on thymus or spleen suspensions from mice of between 5 and 9 weeks of age. All mice used were C57BL/6J, with the exception of the HS V-deficient mice, which were on a BALB/c background. For some experiments, thymic λ NKT cells were enriched by magnetic selection as described for λ NKT cell subset isolation for RNA-Seq. CD1d tetramers loaded with α -GalCer were prepared in our laboratory as previously described⁵³, while tetramers loaded with the α -GalCer analog PBS57 were obtained from the NIH Tetramer Core Facility, and used at a dilution of 1:200 or 1:400. In addition to the antibodies and other reagents used for λ NKT isolation described above, we also used anti-PLZF-AF647 (R17–809), anti-T-bet-PE-CF594 (O4–46), anti-IFN γ -PE-CF594 (XMG1.2), anti-CD49A-PE (Ha31/8), anti-IL-6RA-PE (D7715A7), anti-IL-12-RB1-PE (114), and anti-Runx3-PE (R3–5G4) (all from BD Biosciences), anti-ROR γ t-PerCPeF710 (B2D), anti-Eomes-AF647 (Dan11mag), anti-IL-4-PE-Cy7 (BVD6–24G2) (all from Affymetrix), anti-CCR4-PE (2G12), anti-CXCR3-PE (CXCR3–173), anti-PD-1-BV605 (29F.1A12), and anti-Bcl-2-PE (Bcl/10C4) (all from BioLegend), and anti-IL-17RB-AF488 (752101, R&D Systems). All antibodies were used at concentration of between 0.5 μ g/ml and 1 μ g/ml. Staining for intracellular antigens was performed using reagents and protocols from the Transcription Factor (TF) Buffer Set (BD Biosciences). For intracellular cytokine staining, surface-stained cells were incubated in BD Fix (BD Biosciences) for 10 min at 37 °C before continuing with the TF Buffer Set protocol. Stained samples were analyzed using either LSRII or Fortessa flow cytometers (BD Biosciences) and FlowJo

software (Treestar). All flow cytometry data presented in this manuscript reflect reproducible differences that were consistent with the RNA-Seq data. All antibodies were validated for use in flow cytometry according to data provided from the manufacturers in the accompanying product data sheets; additional validation is also available online (<http://1degreebio.org>).

Apoptosis assay

Thymocyte suspensions were prepared from individual thymi of C57BL/6J mice 8 weeks of age, and enriched for α NKT cells as described for α NKT subset isolation for RNA-Seq analysis. Enriched thymus suspensions were resuspended at 5×10^6 cells/ml in RPMI containing 10% FBS, 50 μ M 2-mercaptoethanol and pen-streptomycin-glutamine (Gemini #400–110), and then cultured in 6 well T.C. plates for 4 h at 37 °C before harvesting and staining. Cells were stained for surface proteins, then washed and stained with AnnexinV-FITC (BD Biosciences #556420) or AnnexinV-APC (BD Biosciences #550475) together with Live-Dead Yellow (ThermoFisher Scientific #L34959) in Annexin binding buffer as described in protocols supplied by BD Biosciences.

Micro-scaled RNA-Seq

Total RNA was purified using a miRNAeasy kit (Qiagen) and quantified as described previously⁹. 10–15 ng of purified total RNA was used for poly(A) mRNA selection (Poly(A) Purist Mag kit, Life Technologies). Poly(A)-selected mRNA was amplified with a whole transcriptome amplification sequencing technology kit (SEQR, Sigma-Aldrich) as per the manufacturer's recommendations. 1 μ g of this amplified cDNA was treated with the restriction enzyme (SEQR, Sigma-Aldrich) to remove the primer sequences and then purified using AMPure XP beads (Beckman Coulter). The efficiency of removal of SEQR primer sequences was assessed by quantitative PCR. From this step, 250 ng purified DNA was diluted with Tris EDTA buffer pH 8.0 to obtain a total volume of 65 μ l. Diluted cDNA was sonicated with a Covartis E220 multiplex sonicator to generate 100- to 250-bp DNA fragments. Approximately 250 ng DNA was used for preparing a standard SOLiD sequencing library (5500 SOLiD fragment 48 library core kit and fragment library barcode adaptors 1–96). Following emulsion PCR samples were sequenced on the 5500 SOLiD sequencer to obtain 35-bp single end reads (SOLiD EZ Bead E120 system kits).

Micro-scaled H3K27ac ChIP-Seq

Formaldehyde-fixed cell pellets were snap-frozen in liquid nitrogen and stored at –80 °C before the ChIP assay. Frozen cells were lysed in 120 μ l of lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS, 1 mM phenyl methane sulfonylfluoride, 20 mg/ml sodium butyrate, proteinase inhibitor cocktail; Sigma), and chromatin was sheared by sonication, using a BioRuptor (Diagenode), to generate 100- to 500-bp DNA fragments. Chromatin was diluted in 1 ml of radioimmunoprecipitation assay (RIPA) buffer and immunoprecipitation was done overnight at 4 °C by incubating 1 μ l of H3K27ac antibody (rabbit polyclonal IgG, lot# Gr167929-1, Abcam) pre-coated on 10 μ l protein A-coated magnetic beads (Invitrogen). Immunocomplexes were captured and washed, and ChIP DNA was eluted as described previously⁹. 1 ng of ChIP DNA was PCR amplified for 18–22 cycles using whole-genome amplification (WGA) primers (WGA-SEQX, Sigma) following the manufacturer's

instructions. Approximately 100 ng amplified DNA was used for preparing a standard Illumina sequencing library (TruSeq Nano DNA HT Sample Preparation Kit, Illumina). Libraries were sequenced on the HiSeq2500 Illumina platform to obtain 50-bp single-end reads (TruSeq Rapid Kit, Illumina). Samples that failed quality control steps, as previously described⁹, were eliminated from further downstream steps and analysis.

Single-cell RNA-Seq

This assay was adapted from the Smart-seq2 protocol⁵⁴. Single cells were sorted by flow cytometry into 96-well PCR plates containing 4 µl cell lysis buffer with RNAase inhibitor (Takara). The sort stream was adjusted carefully to ensure that the cells landed in the liquid interface. RNA loss was minimized by performing on-plate RNA extraction, reverse-transcription and whole transcriptome pre-amplification to generate ~1–10 ng of cDNA. Given that primary λ NKT cells express much lower amounts of RNA per cell (0.3–1 pg) than do the cell lines that were used in the published protocol⁵⁴ (~10 pg/cell), we increased the concentration of primers used (0.1 µM) and the number of PCR cycles for pre-amplification (22 cycles) so that an adequate amount of cDNA was generated without primer dimers and also to avoid over-amplification. Multiple quality-control steps were introduced to ensure that consistency was maintained during the procedure for all samples. Samples that failed quality-control steps as described⁵⁴, were eliminated from further downstream steps and analysis. Standard qPCR was performed for housekeeping genes to ensure comparable amplification of all single-cell samples. Barcoded Illumina sequencing libraries (Nextera XT library preparation kit, Illumina) were generated using the automated platform (Biomek FXP). Libraries were sequenced on the HiSeq2500 Illumina platform to obtain 50-bp single end reads (TruSeq[®] Rapid Kit, Illumina).

Bulk RNA-Seq analysis

Bulk RNA-Seq data were mapped against the mouse mm9 reference genome using tophat⁵⁵ (v1.4.1., -library-type fr-secondstrand -C) and the RefSeq gene annotation downloaded from the UCSC Genome Bioinformatics Site. Sequencing read coverage per gene was counted using HTSeq-count (-m union -s yes -t exon -i gene_id, <http://www-huber.embl.de/users/anders/HTSeq/>). To identify differential gene expression between λ NKT subsets, we performed negative binomial tests for pairwise comparisons of the NKT1, NKT2 and NKT17 cells employing the Bioconductor package DESeq. We considered genes differentially expressed between two λ NKT cell subsets when the DESeq analysis resulted in a Benjamini-Hochberg adjusted *P* value of < 0.05 (5% FDR) and a change in expression of over twofold. A total of 2,538 unique differentially expressed genes (DEGs) were identified in any of the three pairwise cell type comparisons (Supplementary Table 2). Next, we assigned an overall state to each distinct DEG based on the combined results in the three cell-type comparisons. In each comparison, there are three possible outcomes (up, down, or non-significant). Our 2,538 DEGs showed 18 different patterns of outcomes (Supplementary Table 3). To identify the gene-expression program that is 'preferentially' active in each λ NKT subset, we restricted the DEGs into three groups: ~276 'NKT1-enriched' genes that showed increased expression in the NKT1 subset compared to both the NKT2 and NKT17 subsets; ~329 'NKT2-enriched' genes that showed increased expression in the NKT2 subset compared to both the NKT1 and NKT17 subsets; and ~260 'NKT17-

enriched' genes that showed increased expression in the NKT17 subset compared to both the NKT1 and NKT2 subsets. Gene sets obtained for each of three categories (as above) were analyzed using DAVID (Database for Annotation, Visualization and Integrated Discovery; Version 6.750)⁵⁶. Biological process and pathways enriched for genes in our data set are presented (Supplementary Table 5). The number of samples for RNA-Seq and ChIP-Seq analysis was determined based on previous analysis⁹.

ChIP-Seq analysis

H3K27ac ChIP-Seq data were mapped against the mouse mm9 reference genome using bowtie v.0.12.7 ($-n\ 2\ -m\ 1$). The sam files obtained were converted and merged into separated bam files using samtools v.0.1.18. As described previously⁹, H3K27ac enrichment values, presented as reads per kilobase per million mapped (RPKM), were calculated based on reads in non-overlapping, consecutive 500-bp windows ($\sim 6 \times 10^6$ in total) across the whole genome. For visualization of the ChIP-Seq data in public genome browsers, sequencing coverage was calculated at genome wide 500-bp windows after extending each read to a length of 250 bp along the sequencing direction using Bioconductor package MEDIPS v.1.10.0 (extend = 250, uniq = F, window_size = 500, BSgenome = 'BSgenome.Hsapiens.UCSC.hg19')^{57,58}, and the resulting coverage profiles (RPKM) were exported as wiggle files. To identify *NKT* cell subset-specific differentially enriched regions (DERs), we grouped H3K27ac ChIP-Seq data from each specific subset (NKT1 ($n = 4$), NKT2 ($n = 4$) and NKT17 ($n = 3$) cells) and calculated differential enrichment for each subset by modeling a negative binomial distribution for the data dispersion using Bioconductor package MEDIPS, as described previously⁹. To improve the statistical power of the analysis, windows with an average of less than eight mapped reads across all samples were excluded from differential analysis. We identified a total of 10,353 unique DERs (Benjamini-Hochberg adjusted P value, < 0.1): 2,866 DERs for NKT1 vs. NKT2 cells (1,519 up in NKT1 cells and 1,347 up in NKT2 cells), 5,016 DERs for NKT1 vs. NKT17 cells (3,475 up in NKT1 cells and 1,541 up in NKT17 cells), 4,727 DERs for NKT2 vs. NKT17 cells (3,374 up in NKT2 cells and 1,353 up in NKT17 cells) (Supplementary Table 6).

Detecting enrichment of TF-binding motifs at NKT subset-specific enhancers

We examined if there was a significant enrichment in known transcription factor (TF)-binding motifs in the six groups of DERs defined above. In each DER category, we retrieved the genomic sequence corresponding to the DERs, and used the findMotifsGenome.pl of the HOMER software package⁵⁹ with the following parameter settings: region size 500 bp, motif length 8mer to 15mer (Fig. 7a and Supplementary Table 11). The motifs with P values less or equal to 1.00×10^{-3} and ratio of target sequences with motif vs. background sequences with motif > 1.1 were defined as significantly enriched (Supplementary Table 11).

Single-cell RNA-Seq analysis

Single-cell RNA-Seq data were mapped against the mouse mm10 reference genome using tophat⁵⁵ (v1.4.1., -library-type fr-secondstrand) and the RefSeq gene annotation downloaded from the UCSC genome Bioinformatics Site. The Qlucore Omics Explorer 3.1 software package was used for visualization of single-cell RNA-Seq data. Qlucore Omics software converts imported *bam* files into read counts that were normalized across the samples using

the TMM (trimmed mean of M-values) method⁶⁰. The normalization also accounts for sequencing depth and RNA production levels of the samples. Comparison with other normalization methods⁶¹ showed no major differences in the results reported. Briefly, the level of gene expression in single cells was not overly sensitive to normalization procedure. Ideally, the normalization process removes the systematic technical bias between samples and minimizes the biological effect on the final results. Due to large number of zero counts with single-cell RNA-Seq, we used an alternative normalization method employing the upper quantiles of the distribution using the BioConductor package EDASeq⁶². This method rescales the counts by the 75th percentile value. We did not find biological differences in the results for differential expression from using the TMM method (default method used by Qlucore Omics software, data not shown). PCA of normalized data was performed by using the default setting: mean = 0, variance = 1 normalization in the Qlucore Omics software. PCA plot density pattern with four distinct clusters was observed with a wide range of settings for filtering by variance. We did PCA based on 4,780 genes that passed filtering by variance set at the level of 0.59 (Fig. 1d and Supplementary Video 1). Unsupervised hierarchical clustering of 203 single-cell transcriptomes based on the expression of genes ($n = 1,500$) with the highest variance, which accounted for 34% of the total variance, was performed using DESeq package functions and custom scripts on R. Other PCA-based plots, heat maps and scatter plots were constructed using default settings in the Qlucore Omics software, and these plots were used for visualization, at a single-cell level, of λ NKT cell subset-specific genes of interest defined from the bulk RNA-Seq data analysis. Differential gene expression analysis across single cells in the four λ NKT cell subsets was carried out using Qlucore Omics software, which performs a multi-group comparison using the ANOVA F-test.

Comparison of genes uniquely upregulated in ILCs and NKT subsets

Lists of genes found to be upregulated uniquely in one of the three λ NKT cell subsets (NKT1, NKT2 or NKT17) were compared to lists of genes uniquely upregulated in type 1 ILCs, type 2 ILCs and both type 3 ILCs and LT α i-like cells compared to all other ILC subsets, using online Venn diagram software (<http://www.bioinformatics.lu/venn.php>), in order to identify genes commonly upregulated in particular ILC and λ NKT cell subsets. Significant associations between specific ILC and NKT subsets were determined using a hypergeometric test: $h(x; N, n, k) = \frac{[kC_x][N-kC_{n-x}]}{[NC_n]}$, where h = hypergeometric probability, x = number of genes upregulated in both a specific λ NKT cell subset and an ILC subset, N = total number of genes with significant expression in either the λ NKT or the ILC subset, n = number of genes upregulated in the λ NKT cell subset, and k = number of genes upregulated in the ILC subset.

Statistical analysis

Unpaired Student's t -tests were performed using Prism 6.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

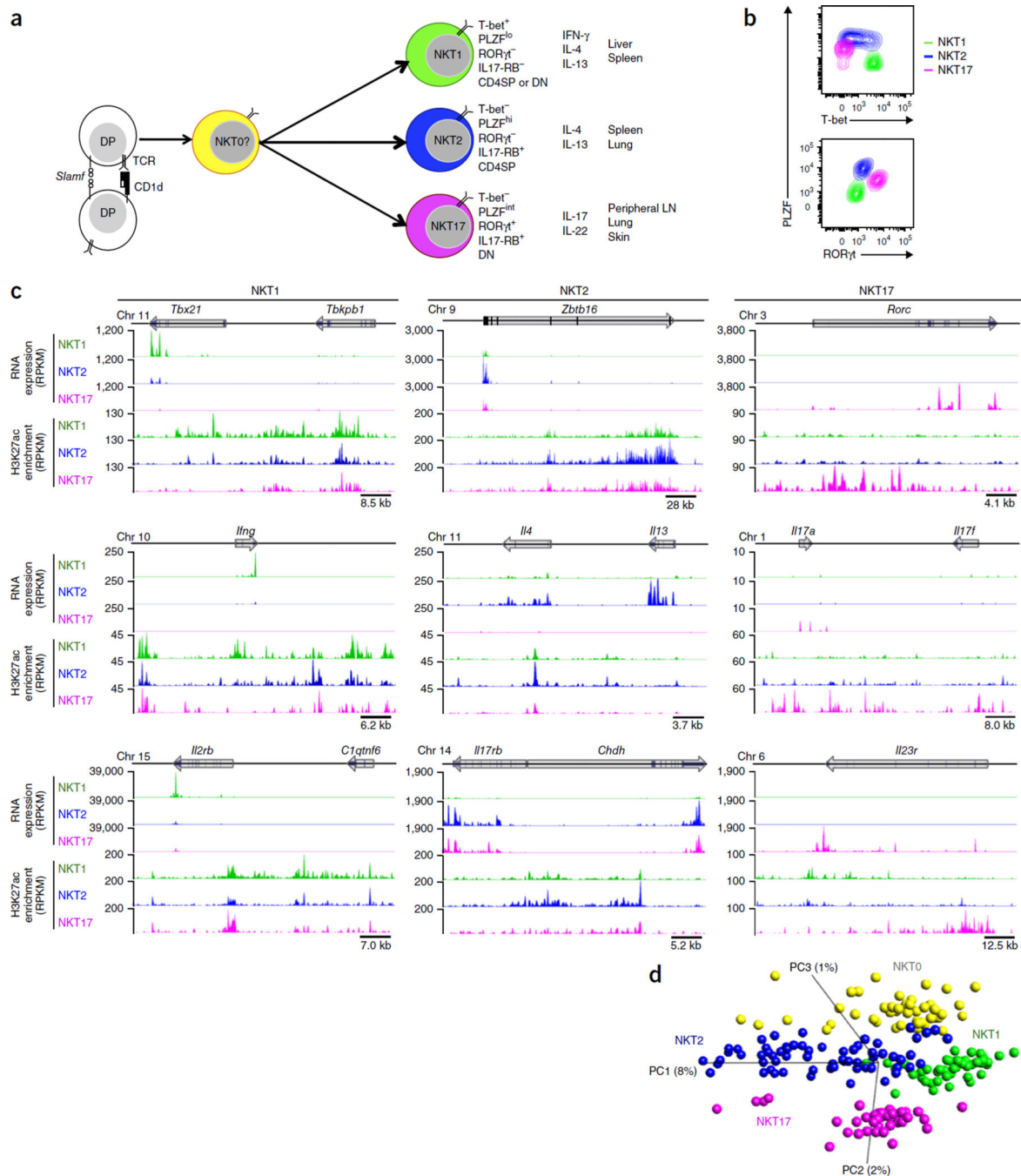
We thank C. Kim, L. Nosworthy and K. Van Gundt for assisting with single cell sorting; J. Day for assistance with next generation sequencing; Z. Fu, A.M. Gholami and J. Greenbaum for help with processing and analysis of sequencing data; and K.M. Ansel (University of California, San Francisco) and A. Rao (La Jolla Institute for Allergy & Immunology) for HS V-deficient mice. Supported by the Alexander von Humboldt Foundation (L.C.), The William K. Bowes Jr Foundation (P.V.) and the US National Institutes of Health (R01 HL114093 to P.V.; U19 AI100275 to P.V.; and R37 AI71922 and R01 AI105215 to M.K.).

References

1. Chandra S, Kronenberg M. Activation and Function of iNKT and MAIT Cells. *Adv. Immunol.* 2015; 127:145–201. [PubMed: 26073984]
2. Matsuda JL, Mallevaey T, Scott-Browne J, Gapin L. CD1d-restricted iNKT cells, the ‘Swiss-Army knife’ of the immune system. *Curr. Opin. Immunol.* 2008; 20:358–368. [PubMed: 18501573]
3. Constantinides MG, Bendelac A. Transcriptional regulation of the NKT cell lineage. *Curr. Opin. Immunol.* 2013; 25:161–167. [PubMed: 23402834]
4. Lee YJ, Holzapfel KL, Zhu J, Jameson SC, Hogquist KA. Steady-state production of IL-4 modulates immunity in mouse strains and is determined by lineage diversity of iNKT cells. *Nat. Immunol.* 2013; 14:1146–1154. [PubMed: 24097110]
5. Watarai H, et al. Development and function of invariant natural killer T cells producing T_H2- and T_H17-cytokines. *PLoS Biol.* 2012; 10:e1001255. [PubMed: 22346732]
6. Doisne JM, et al. Skin and peripheral lymph node invariant NKT cells are mainly retinoic acid receptor-related orphan receptor γ ⁺ and respond preferentially under inflammatory conditions. *J. Immunol.* 2009; 183:2142–2149. [PubMed: 19587013]
7. Sandberg R. Entering the era of single-cell transcriptomics in biology and medicine. *Nat. Methods.* 2014; 11:22–24. [PubMed: 24524133]
8. Benlagha K, Wei DG, Veiga J, Teyton L, Bendelac A. Characterization of the early stages of thymic NKT cell development. *J. Exp. Med.* 2005; 202:485–492. [PubMed: 16087715]
9. Seumois G, et al. Epigenomic analysis of primary human T cells reveals enhancers associated with T_H2 memory cell differentiation and asthma susceptibility. *Nat. Immunol.* 2014; 15:777–788. [PubMed: 24997565]
10. Thomas SY, et al. PLZF induces an intravascular surveillance program mediated by long-lived LFA-1-ICAM-1 interactions. *J. Exp. Med.* 2011; 208:1179–1188. [PubMed: 21624939]
11. Lynch L, et al. Regulatory iNKT cells lack expression of the transcription factor PLZF and control the homeostasis of T_{reg} cells and macrophages in adipose tissue. *Nat. Immunol.* 2015; 16:85–95. [PubMed: 25436972]
12. Shalek AK, et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature.* 2014; 510:363–369. [PubMed: 24919153]
13. John B, Crispe IN. Passive and active mechanisms trap activated CD8⁺ T cells in the liver. *J. Immunol.* 2004; 172:5222–5229. [PubMed: 15100260]
14. Lee YJ, et al. Tissue-specific distribution of iNKT cells impacts their cytokine response. *Immunity.* 2015; 43:566–578. [PubMed: 26362265]
15. Benlagha K, Kyin T, Beavis A, Teyton L, Bendelac A. A thymic precursor to the NK T cell lineage. *Science.* 2002; 296:553–555. [PubMed: 11968185]
16. Berzins SP, McNab FW, Jones CM, Smyth MJ, Godfrey DI. Long-term retention of mature NK1.1⁺ NKT cells in the thymus. *J. Immunol.* 2006; 176:4059–4065. [PubMed: 16547241]
17. Engel I, Kronenberg M. Transcriptional control of the development and function of Va14i NKT cells. *Curr. Top. Microbiol. Immunol.* 2014; 381:51–81. [PubMed: 24839184]
18. O’Brien TF, et al. Cytokine expression by invariant natural killer T cells is tightly regulated throughout development and settings of type-2 inflammation. *Muc. Immunol.* 2015 Sep 9.
19. Diehl S, et al. Induction of NFATc2 expression by interleukin 6 promotes T helper type 2 differentiation. *J. Exp. Med.* 2002; 196:39–49. [PubMed: 12093869]

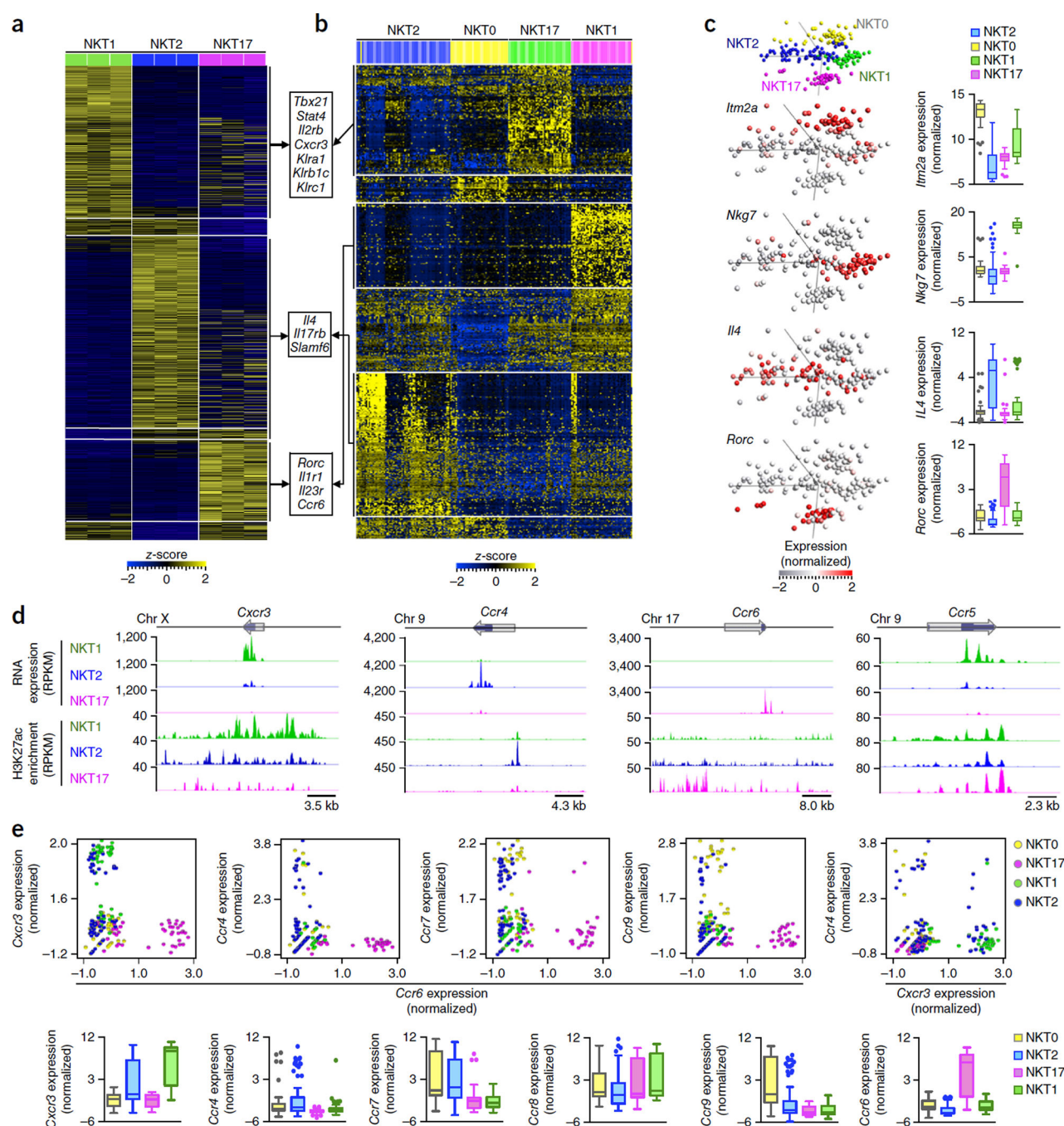
20. Iwamura C, Nakayama T. Role of NKT cells in allergic asthma. *Curr. Opin. Immunol.* 2010; 22:807–813. [PubMed: 21115332]
21. Ledford JG, Kovarova M, Koller BH. Impaired host defense in mice lacking ONZIN. *J. Immunol.* 2007; 178:5132–5143. [PubMed: 17404296]
22. Ledford JG, Kovarova M, Jania LA, Nguyen M, Koller BH. ONZIN deficiency attenuates contact hypersensitivity responses in mice. *Immunol. Cell Biol.* 2012; 90:733–742. [PubMed: 22249203]
23. Lau JY, et al. Fibulin-1 is increased in asthma--a novel mediator of airway remodeling? *PLoS One.* 2010; 5:e13360. [PubMed: 20967215]
24. Clinton GM, et al. Estrogens increase the expression of fbulin-1, an extracellular matrix protein secreted by human ovarian cancer cells. *Proc. Natl. Acad. Sci. USA.* 1996; 93:316–320. [PubMed: 8552629]
25. Engel I, Zhao M, Kappes D, Taniuchi I, Kronenberg M. The transcription factor Th-POK negatively regulates Th17 differentiation in V α 14i NKT cells. *Blood.* 2012; 120:4524–4532. [PubMed: 23034280]
26. Laird RM, Laky K, Hayes SM. Unexpected role for the B cell-specific Src family kinase B lymphoid kinase in the development of IL-17-producing $\gamma\delta$ T cells. *J. Immunol.* 2010; 185:6518–6527. [PubMed: 20974990]
27. Tanaka S, et al. Sox5 and c-Maf cooperatively induce Th17 cell differentiation via ROR γ t induction as downstream targets of Stat3. *J. Exp. Med.* 2014; 211:1857–1874. [PubMed: 25073789]
28. Malhotra N, et al. Immunological Genome Project Consortium. A network of high-mobility group box transcription factors programs innate interleukin-17 production. *Immunity.* 2013; 38:681–693. [PubMed: 23562159]
29. Zhao P, Hou L, Farley K, Sundrud MS, Remold-O'Donnell E. SerpinB1 regulates homeostatic expansion of IL-17 $^{+}$ $\gamma\delta$ and CD4 $^{+}$ Th17 cells. *J. Leukoc. Biol.* 2014; 95:521–530. [PubMed: 24249741]
30. Hara-Chikuma M, et al. Chemokine-dependent T cell migration requires aquaporin-3-mediated hydrogen peroxide uptake. *J. Exp. Med.* 2012; 209:1743–1752. [PubMed: 22927550]
31. Esworthy RS, Ho YS, Chu FF. The Gpx1 gene encodes mitochondrial glutathione peroxidase in the mouse liver. *Arch. Biochem. Biophys.* 1997; 340:59–63. [PubMed: 9126277]
32. Barrett CW, et al. Selenoprotein P influences colitis-induced tumorigenesis by mediating stemness and oxidative damage. *J. Clin. Invest.* 2015; 125:2646–2660. [PubMed: 26053663]
33. Burk RF, Hill KE. Selenoprotein P-expression, functions, and roles in mammals. *Biochim. Biophys. Acta.* 2009; 1790:1441–1447. [PubMed: 19345254]
34. Robinette ML, et al. Immunological Genome Consortium. Transcriptional programs define molecular characteristics of innate lymphoid cell classes and subsets. *Nat. Immunol.* 2015; 16:306–317. [PubMed: 25621825]
35. Starr AE, Bellac CL, Dufour A, Goebeler V, Overall CM. Biochemical characterization and N-terminomics analysis of leukolysin, the membrane-type 6 matrix metalloprotease (MMP25): chemokine and vimentin cleavages enhance cell migration and macrophage phagocytic activities. *J. Biol. Chem.* 2012; 287:13382–13395. [PubMed: 22367194]
36. Nitta T, Takahama Y. The lymphocyte guard-IANs: regulation of lymphocyte survival by IAN/ GIMAP family proteins. *Trends Immunol.* 2007; 28:58–65. [PubMed: 17196432]
37. Xu H, et al. MS4a4B, a CD20 homologue in T cells, inhibits T cell propagation by modulation of cell cycle. *PLoS One.* 2010; 5:e13780. [PubMed: 21072172]
38. Seino K, Harada M, Taniguchi M. NKT cells are relatively resistant to apoptosis. *Trends Immunol.* 2004; 25:219–221. [PubMed: 15099560]
39. Knosp CA, et al. SOCS2 regulates T helper type 2 differentiation and the generation of type 2 allergic responses. *J. Exp. Med.* 2011; 208:1523–1531. [PubMed: 21646394]
40. Tyznik AJ, Verma S, Wang Q, Kronenberg M, Benedict CA. Distinct requirements for activation of NKT and NK cells during viral infection. *J. Immunol.* 2014; 192:3676–3685. [PubMed: 24634489]
41. Lei Y, Takahama Y. XCL1 and XCR1 in the immune system. *Microbes Infect.* 2012; 14:262–267. [PubMed: 22100876]

42. Gottschalk C, Mettke E, Kurts C. The role of invariant natural killer T cells in dendritic cell licensing, cross-priming, and memory CD8⁺ T cell generation. *Front. Immunol.* 2015; 6:379. [PubMed: 26284065]
43. Kirchner J, Bevan MJ. ITM2A is induced during thymocyte selection and T cell activation and causes downregulation of CD8 when overexpressed in CD4⁺CD8⁺ double positive thymocytes. *J. Exp. Med.* 1999; 190:217–228. [PubMed: 10432285]
44. Uehara S, et al. Premature expression of chemokine receptor CCR9 impairs T cell development. *J. Immunol.* 2006; 176:75–84. [PubMed: 16365398]
45. Doherty JR, Cleveland JL. Targeting lactate metabolism for cancer therapeutics. *J. Clin. Invest.* 2013; 123:3685–3692. [PubMed: 23999443]
46. Witke W, et al. Hemostatic, inflammatory, and fibroblast responses are blunted in mice lacking gelsolin. *Cell.* 1995; 81:41–51. [PubMed: 7720072]
47. Carr T, et al. The transcription factor lymphoid enhancer factor 1 controls invariant natural killer T cell expansion and Th2-type effector differentiation. *J. Exp. Med.* 2015; 212:793–807. [PubMed: 25897173]
48. Vijayanand P, et al. Interleukin-4 production by follicular helper T cells requires the conserved Il4 enhancer hypersensitivity site V. *Immunity.* 2012; 36:175–187. [PubMed: 22326582]
49. Chang PP, et al. Identification of Bcl-6-dependent follicular helper NKT cells that provide cognate help for B cell responses. *Nat. Immunol.* 2012; 13:35–43. [PubMed: 22120117]
50. Sag D, Krause P, Hedrick CC, Kronenberg M, Wingender G. IL-10-producing NKT10 cells are a distinct regulatory invariant NKT cell subset. *J. Clin. Invest.* 2014; 124:3725–3740. [PubMed: 25061873]
51. Islam S, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods.* 2014; 11:163–166. [PubMed: 24363023]
52. Godfrey DI, Stankovic S, Baxter AG. Raising the NKT cell family. *Nat. Immunol.* 2010; 11:197–206. [PubMed: 20139988]
53. Sidobre S, Kronenberg M. CD1 tetramers: a powerful tool for the analysis of glycolipid-reactive T cells. *J. Immunol. Methods.* 2002; 268:107–121. [PubMed: 12213347]
54. Picelli S, et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods.* 2013; 10:1096–1098. [PubMed: 24056875]
55. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009; 25:1105–1111. [PubMed: 19289445]
56. Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 2009; 4:44–57. [PubMed: 19131956]
57. Chavez L, et al. Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage. *Genome Res.* 2010; 20:1441–1450. [PubMed: 20802089]
58. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics.* 2014; 30:284–286. [PubMed: 24227674]
59. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell.* 2010; 38:576–589. [PubMed: 20513432]
60. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010; 11:R25. [PubMed: 20196867]
61. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics.* 2010; 11:94. [PubMed: 20167110]
62. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Gen. Biol.* 2014; 15:550.

**Figure 1.**

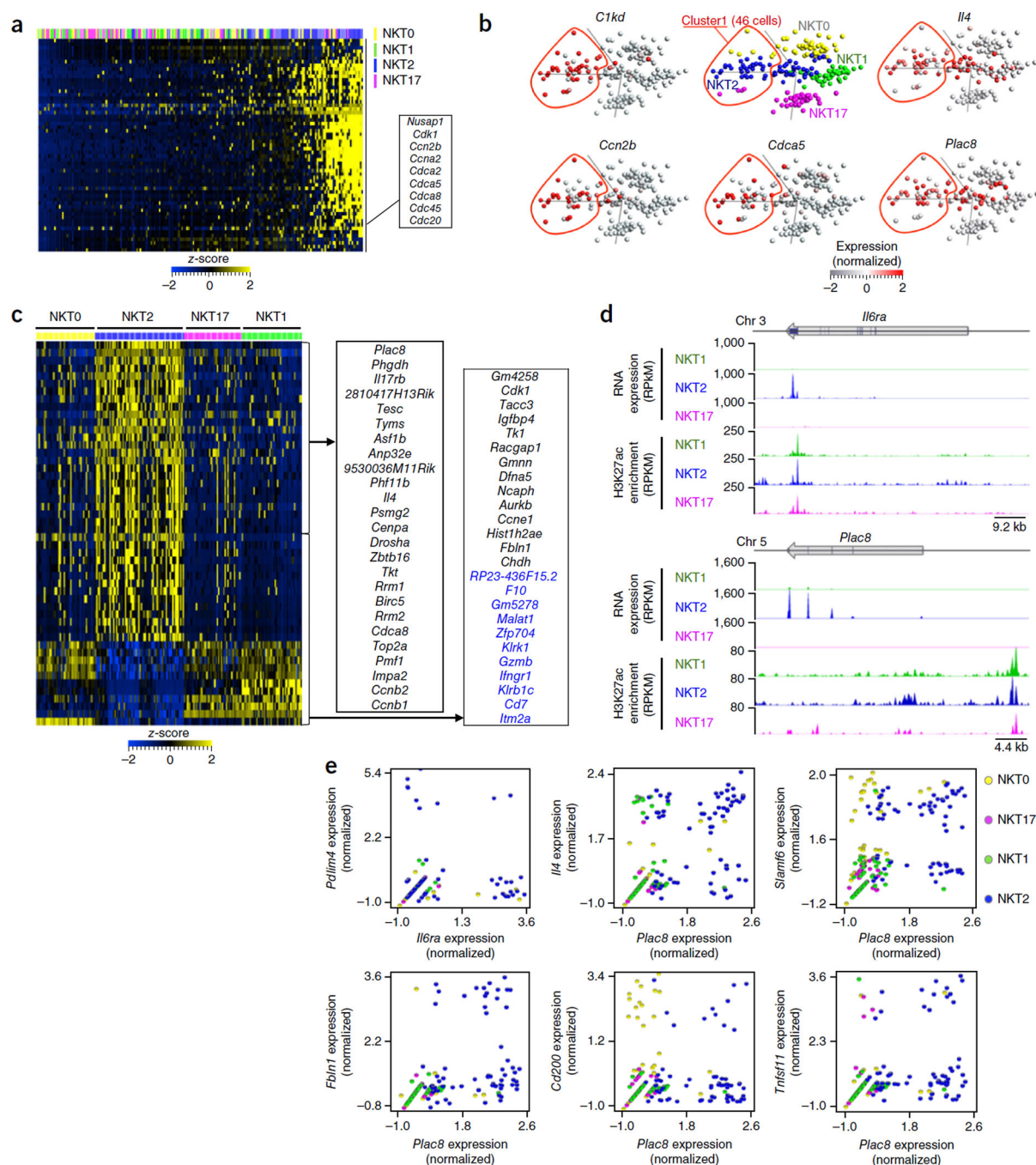
Transcriptional and enhancer profiles of λ NKT cell subsets. **(a)** Developmental pathways, known markers and properties of the NKT0 (yellow), NKT1 (green), NKT2 (blue) and NKT17 (magenta) subsets of λ NKT cells (same subset colors throughout). ‘*Slamf*’ indicates *Slamf1* and *Slamf6*, which encode members of the signaling lymphocytic-activation molecule (SLAM) family; DP, double positive (CD4⁺CD8⁺); LN, lymph node. **(b)** Flow cytometry analyzing expression of the key transcription factors PLZF, T-bet and RORγt in electronically gated λ NKT thymocyte cell subsets from 6-week-old female C57BL/6J mice,

defined by a sorting strategy based on differential surface-antigen expression (Supplementary Fig. 1). **(c)** Bulk analyses of λ NKT cell subset-specific expression and H3K27ac-enrichment patterns for genes encoding transcription factors (top), cytokines (middle) and cytokine receptors (bottom) in NKT1, NKT2 and NKT17 subsets from 5-week-old C57BL/6J female mice, along with UCSC gene tracks (above plots). RPKM, reads per kilobase per million mapped. **(d)** PCA of 203 single-cell transcriptomes that distinguish λ NKT cell subsets from 5-week-old C57BL/6J female mice: each symbol represents an individual cell; numbers in parentheses indicate percent variance in each principal component (PC1–PC3). Data are from one experiment representative of two separate experiments with five independent samples **(b)** or are from two experiments with three to four pooled biological replicates, each generated from a pool of thymi from five mice (RNA-Seq) or fifteen mice (ChIP-Seq) **(c)**, or one experiment with samples obtained from one sorting of thymi from three mice, processed in two technical batches **(d)**.

**Figure 2.**

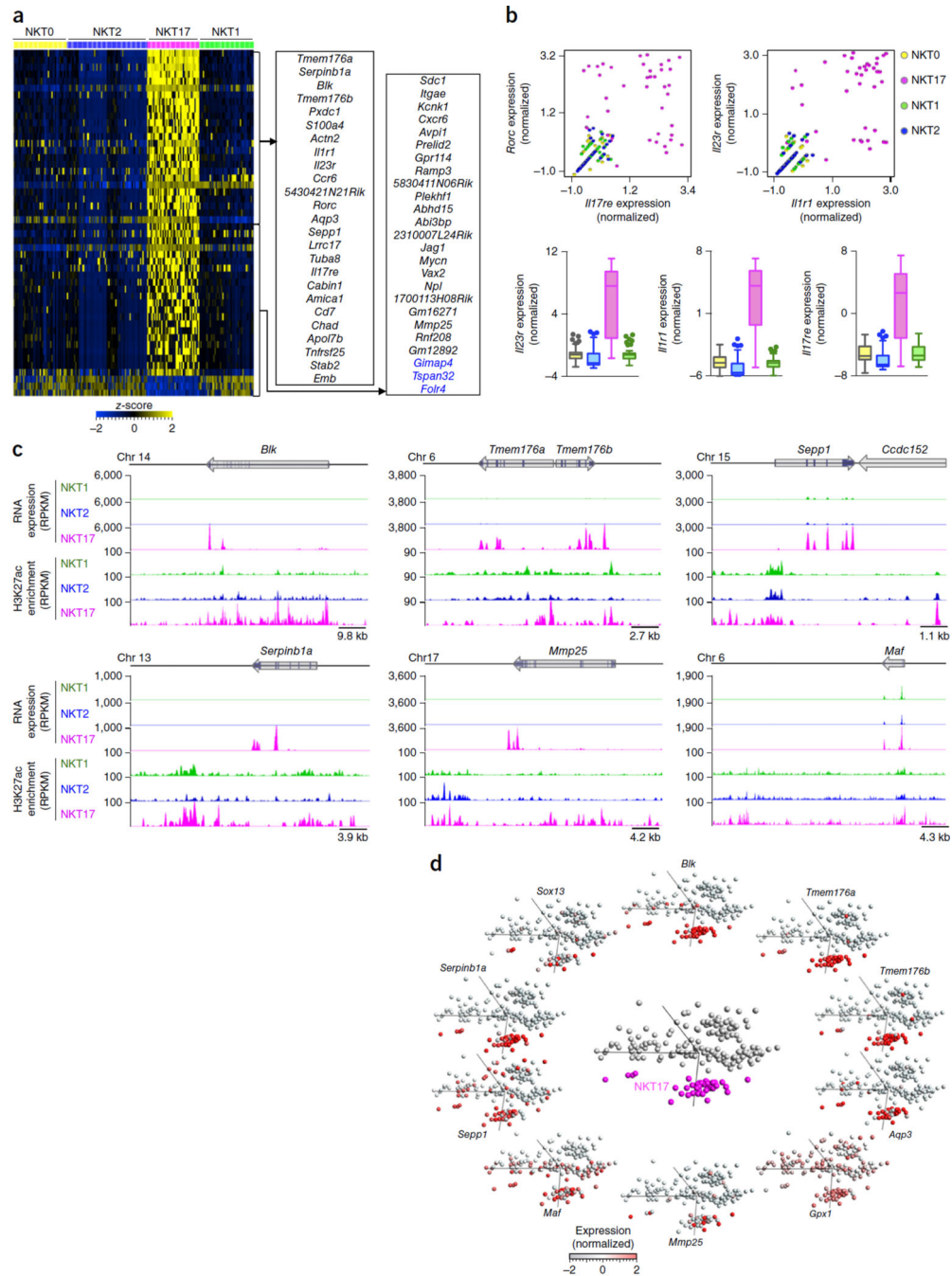
Bulk and single-cell RNA-Seq profiling of iNKT cell subsets. **(a)** Bulk sequencing analysis, showing row-wise z-scores of normalized read counts for each differentially expressed gene (rows) obtained from pairwise comparisons of NKT1, NKT2 and NKT17 cells from 5-week-old C57BL/6J female mice, and independent bulk RNA-Seq assays (columns; each bar at top is one assay of nine total). Right, genes with expression enriched in a subset, as assessed by both bulk sequencing and single-cell sequencing. **(b)** Single-cell RNA-Seq analysis of the 520 genes (rows) most differentially expressed among iNKT cell subsets from 5-week-old

C57BL/6J female mice, and row-wise z -scores of normalized read counts from 203 single cells (columns; each bar at top represents a single cell (subset colors as in Fig. 1)); single cells and genes were ordered on the basis of unsupervised hierarchical clustering. Left, as in **a**. **(c)** PCA of iNKT cell subsets from 5-week-old C57BL/6J female mice (top; as in Fig. 1d) and expression of select subset-specific genes (below) in each cell at top, presented as normalized gene expression (key; mean = 0 and variance = 1) (left), and expression of those genes in single cells from each subset (key), presented as ‘box and whisker’ plots (mean and interquartile range of expression) (right). **(d)** Bulk RNA-Seq analysis and H3K27ac-enrichment patterns for genes encoding chemokine receptors in NKT1, NKT2 and NKT17 subsets from 5-week-old C57BL/6J female mice (presented as in Fig. 1c). **(e)** Single-cell RNA-Seq analysis of the expression ($\log_2$ normalized counts) of various genes (axes) in cells from 5-week-old C57BL/6J female mice, presented as scatterplots (top row; each symbol represents an individual cell) or as in **c**, right (bottom row). Data are from one experiment with three pooled biological replicates, each generated from a pool of thymi from five mice (**a**, **d** (RNA-Seq)), one experiment with samples obtained from one sorting of thymi from three mice, processed in two technical batches (**b,c,e**), or one experiment with four pooled biological replicates, each generated from a pool of thymi from fifteen mice (**d** (ChIPseq)).

**Figure 3.**

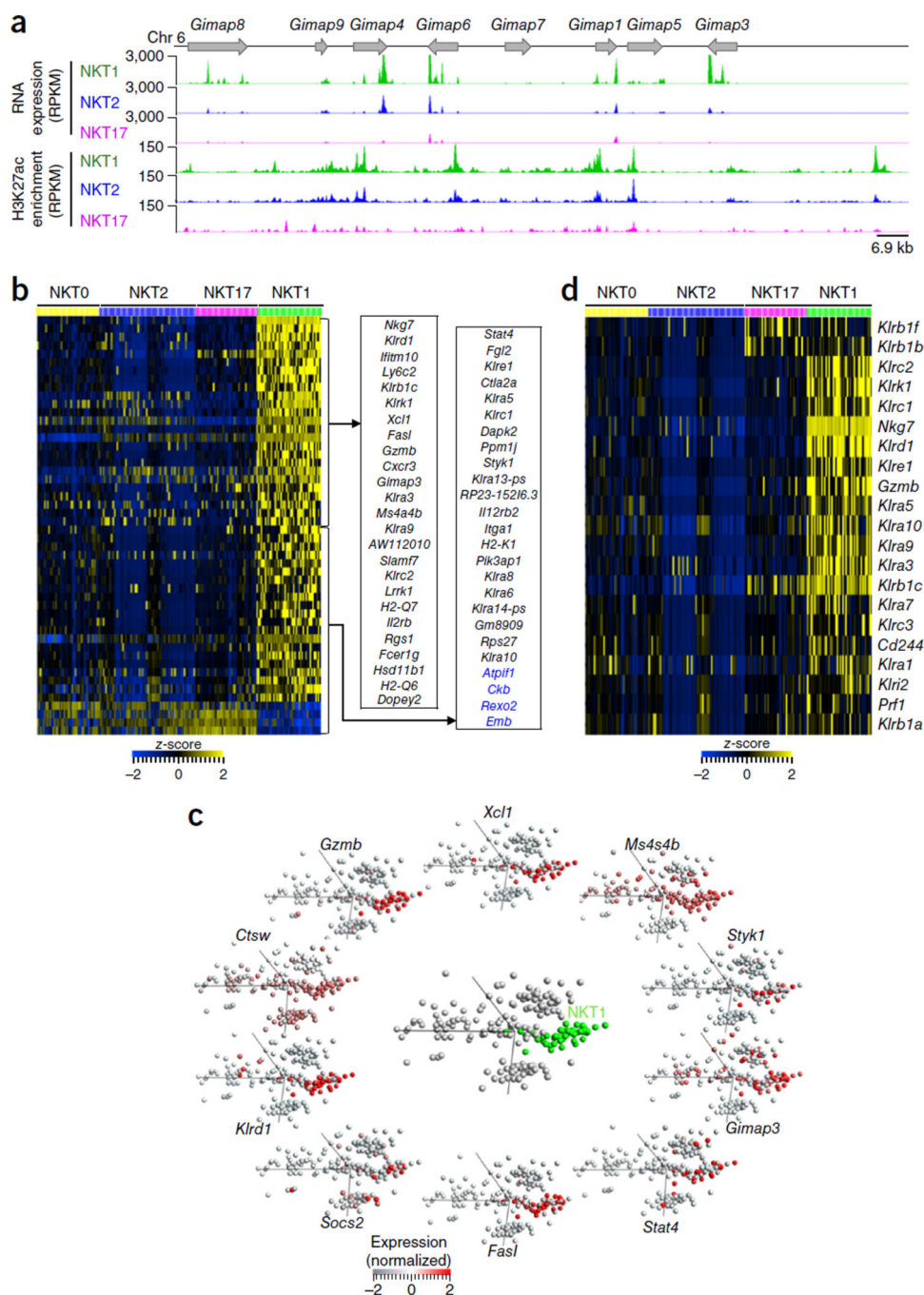
NKT2-enriched transcripts. **(a)** Single-cell RNA-Seq assays ($n = 203$) of cells from 5-week-old C57BL/6J female mice, showing row-wise z-scores (key) of library-size-normalized read counts (columns; bars at top as in Fig. 2b) for the expression of 51 genes (rows) encoding products involved in the cell cycle, with samples (single cells) and genes ordered on the basis of unsupervised hierarchical clustering. Right, selected genes. **(b)** PCA of α NKT cell subsets from 5-week-old C57BL/6J female mice (top middle; as in Fig. 1d) and single-cell RNA-Seq analysis of genes encoding products involved in the cell cycle and broadly NKT2-

enriched genes (presented as in Fig. 2c, left); red outlines (cluster 1) indicate individual cells that tend to express genes encoding products involved in the cell cycle. **(c)** Single-cell RNA-Seq assays ($n = 203$) of cells from 5-week-old C57BL/6J female mice, showing row-wise z -scores (presented as in a) for the 50 genes (rows) with the most significantly differential expression in the NKT2 subset relative to their expression in other subsets, with genes ordered on the basis of change in expression (top (greatest change) to bottom (least change)). Right, genes with expression enriched in the NKT2 subset (black, increased; blue, decreased). **(d)** Bulk RNA-Seq analysis and H3K27ac-enrichment patterns for *Il6ra* and *Plac8* in NKT1, NKT2 and NKT17 subsets from 5-week-old C57BL/6J female mice (presented as in Fig. 1c). **(e)** Single-cell RNA-Seq analysis of the expression of various genes in cells from 5-week-old C57BL/6J female mice (presented as in Fig. 2e, top). Data are from one experiment with one sorting of thymi pooled from three mice and processed in two technical batches (**a–c,e**) or are from two experiments with three to four pooled biological replicates, each generated from a pool of thymi from five mice (RNAseq) or fifteen mice (ChIPseq) (**d**).

**Figure 4.**

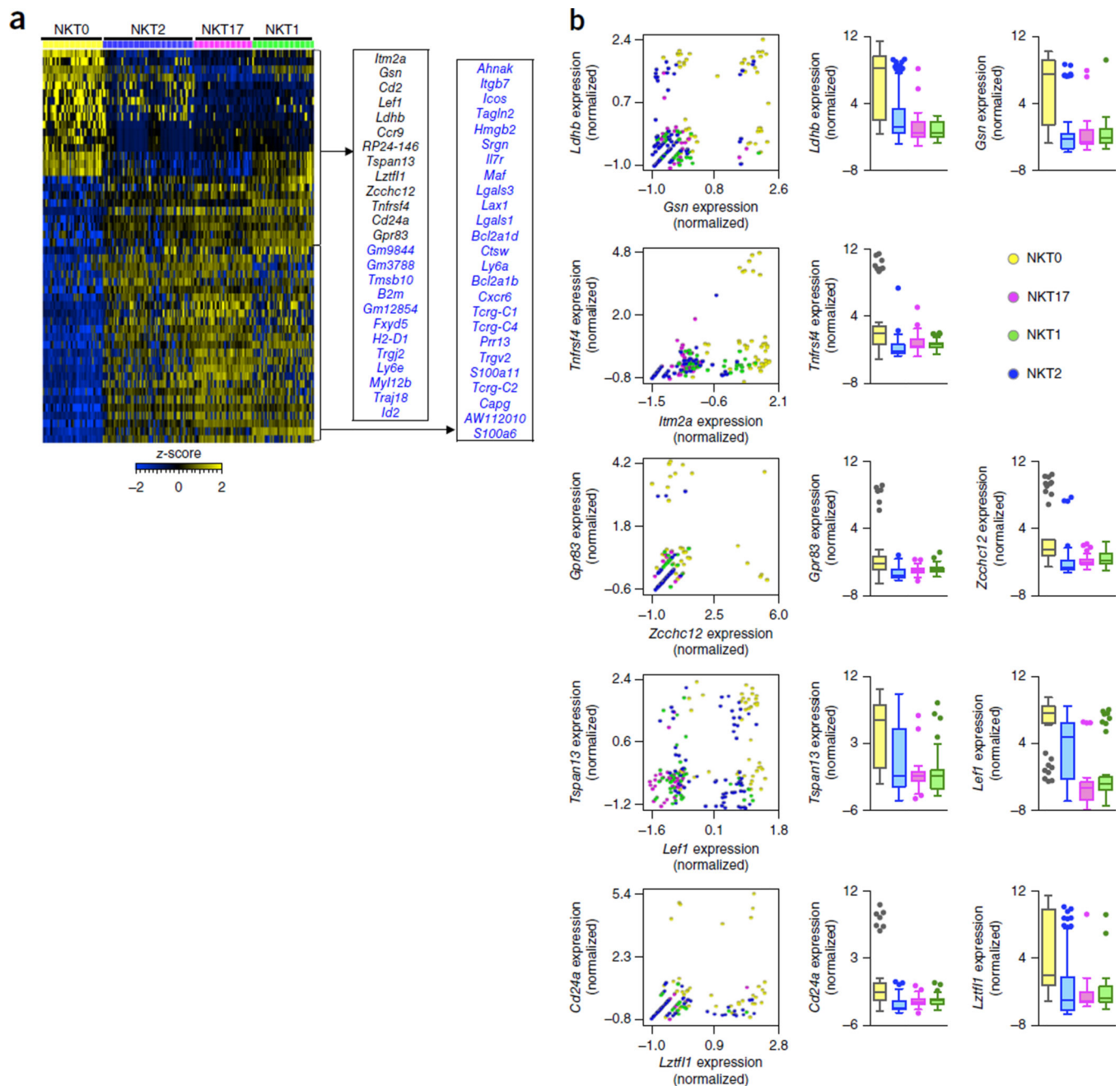
NKT17-enriched transcripts. **(a)** Single-cell RNA-Seq assays ($n = 203$) of cells from 5-week-old C57BL/6J female mice, showing row-wise z-scores (key) of library-size-normalized read counts (columns; bars at top as in Fig. 2b) for the 50 genes with the most significantly differential expression in the NKT17 subset relative to their expression in other subsets (presented as in Fig. 3c). **(b)** Single-cell RNA-Seq analysis ($\log_2$ normalized counts) of various genes in cells from 5-week-old C57BL/6J female mice (presented as in Fig. 2e). **(c)** Bulk RNA-Seq analysis and H3K27ac-enrichment patterns for various genes in NKT1,

NKT2 and NKT17 subsets from 5-week-old C57BL/6J female mice (presented as in Fig. 1c). **(d)** PCA of α NKT cell subsets from 5-week-old C57BL/6J female mice (center; as in Fig. 1d, with only NKT17 in color (magenta)) and single-cell RNA-Seq analysis of NKT17-specific transcripts (presented as in Fig. 2c, left). Data are from one experiment with one sorting of thymi pooled from three mice and processed in two technical batches (**a–c**) or are from two experiments with three to four pooled biological replicates, each generated from a pool of thymi from five mice (RNAseq) or fifteen mice (ChIPseq) (**d**).

**Figure 5.**

NKT1-enriched transcripts. **(a)** Bulk RNA-Seq analysis and H3K27ac-enrichment patterns for genes encoding members of the GIMAP family in NKT1, NKT2 and NKT17 subsets from 5-week-old C57BL/6J female mice (presented as in Fig. 1c). **(b)** Single-cell RNA-Seq assays ($n = 203$) of cells from 5-week-old C57BL/6J female mice, showing row-wise z-scores (key) of library-size-normalized read counts (columns; bars at top as in Fig. 2b) for the 50 genes (rows) with the most significantly differential expression in the NKT1 subset relative to their expression in other subsets. Right, genes with expression enriched in the

NKT1 subset (genes ordered as in Fig. 3c). **(c)** PCA of λ NKT cell subsets in 5-week-old C57BL/6J female mice (center; as in Fig. 1d, with only NKT1 in color (green)), and single-cell RNA-Seq analysis of NKT1-subset-specific transcripts (presented as in Fig. 2c, left). **(d)** Single-cell RNA-Seq assays ($n = 203$) as in **a** of cells from 5-week-old C57BL/6J female mice, analyzing the expression of 21 genes encoding NK cell receptors (rows; right margin), with hierarchical clustering. Data are from two experiments with three to four pooled biological replicates, each generated from a pool of thymi from five mice (RNAseq) or fifteen mice (ChIPseq) **(a)**, or are from one experiment with one sorting of thymi pooled from three mice and processed in two technical batches **(b–d)**.

**Figure 6.**

Gene expression in NKT0 cells. **(a)** Single-cell RNA-Seq assays ($n = 203$) of cells from 5-week-old C57BL/6J female mice, showing rowwise z-scores (key) of library-size-normalized read counts (columns; bars at top as in Fig. 2b) for the 50 genes (rows) with the most significantly differential expression in the NKT0 subset relative to their expression in other subsets. Right, genes with expression enriched in the NKT0 subset (genes ordered as in Fig. 3c). **(b)** Single-cell RNA-Seq analysis of the expression (log₂ normalized counts) of various genes (axes) in cells from 5-week-old C57BL/6J female mice (presented as in Fig.

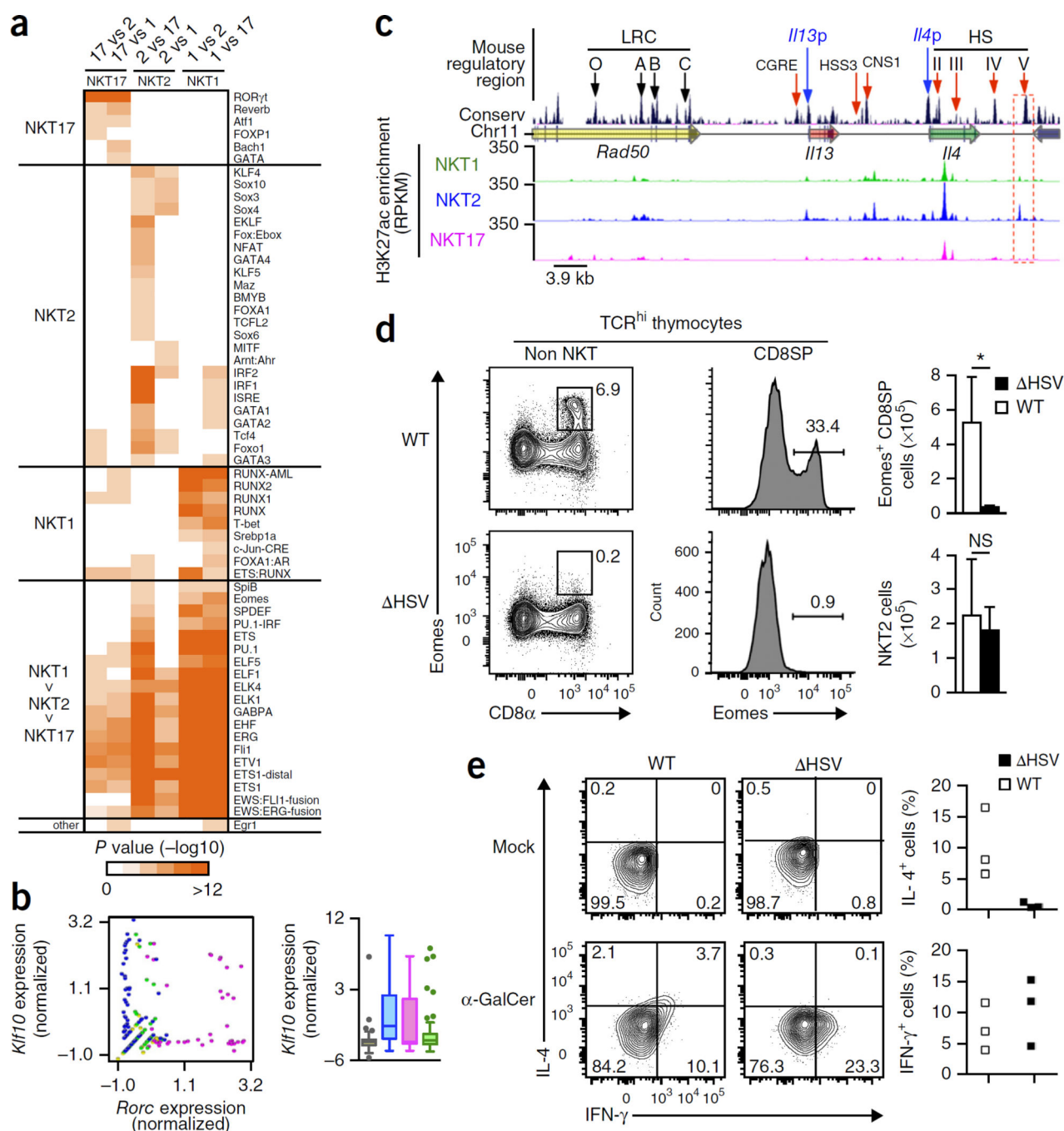
2e). Data are from one experiment with one sorting of thymi pooled from three mice and processed in two technical batches.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Figure 7.**

TF-binding motifs in enhancers linked to Δ NKT cell subset differentiation. **(a)** Significance (key, bottom) of the enrichment in binding motifs for various TFs (right margin) in enhancers that gained H3K27ac in each Δ NKT cell subset relative to the abundance of H3K27ac in other subsets (pairwise comparisons, top: 17, NKT17; 2, NKT2; 1, NKT1), in cells from 5-week-old C57BL/6J female mice. **(b)** Single-cell RNA-Seq analysis of the expression ($\log_2$ normalized counts) of *Klf10* and *Rorc* in cells from 5-week-old C57BL/6J female mice (presented as in Fig. 2e). **(c)** H3K27ac-enrichment patterns across the extended

Ii4 locus in λ NKT cell subsets from 5-week-old C57BL/6J female mice (presented as in Fig. 1c), along with the location of locus-control regions (LCR: O, A, B, C), the *Ii13* promoter (*Ii13p*), the conserved GATA-3-response element (CGRE), the *Ii4* promoter (*Ii4p*), DNase I-hypersensitivity site S3 (HSS3), the conserved noncoding sequence CNS1, and DNase I-hypersensitivity sites (HS) II–V (top), as well as UCSC multispecies conservation tracks (above plot); red dashed line indicates the HS V enhancer. **(d)** Flow cytometry of TCR^{hi} non- λ NKT cell thymocytes (Non NKT) from wild-type mice (WT) and HS V-deficient mice (Δ HSV), stained for Eomes and CD8 α (left), or of TCR^{hi} CD4[–]CD8⁺ thymocytes (CD8SP) from female wild-type BALB/c mice (top) and age-matched HS V-deficient mice (bottom), stained for Eomes (middle). Numbers adjacent to outlined areas (left) indicate percent Eomes⁺CD8 α ⁺ cells; numbers above bracketed lines (middle) indicate percent Eomes⁺ cells among the CD8SP thymocytes. Far right, quantification of Eomes⁺ CD8SP cells (top) and NKT2 cells (bottom) in the thymi of wild-type and HS V-deficient mice, on the basis of surface phenotype. NS, not significant ($P = 0.5827$); $*P = 0.0029$ (unpaired Student's *t*-test). **(e)** Flow cytometry of splenic λ NKT cells obtained from wild-type and HS V-deficient mice and left unstimulated (Mock) or challenged *in vivo* with α -galactosyl ceramide (α -GalCer), followed by intracellular staining for IL-4 and IFN- γ (left), and expression of IL-4 and IFN- γ in λ NKT cells after *in vivo* challenge with α -galactosyl ceramide (right). Each symbol (right) represents an individual mouse. Data are from one experiment with four pooled biological replicates, each generated from a pool of thymi from fifteen mice (**a,c**), one experiment with one sorting of thymi pooled from three mice and processed in two technical batches (**b**), or two independent experiments with five mice (**d,e** (three challenged with α -galactosyl ceramide; two mock-treated); average and standard error in **d**, right).