

Transcriptional programs define molecular characteristics of innate lymphoid cell classes and subsets

Michelle L Robinette¹, Anja Fuchs², Victor S Cortez¹, Jacob S Lee³, Yaming Wang¹, Scott K Durum⁴, Susan Gilfillan¹, Marco Colonna¹ & the Immunological Genome Consortium⁵

The recognized diversity of innate lymphoid cells (ILCs) is rapidly expanding. Three ILC classes have emerged, ILC1, ILC2 and ILC3, with ILC1 and ILC3 including several subsets. The classification of some subsets is unclear, and it remains controversial whether natural killer (NK) cells and ILC1 cells are distinct cell types. To address these issues, we analyzed gene expression in ILCs and NK cells from mouse small intestine, spleen and liver, as part of the Immunological Genome Project. The results showed unique gene-expression patterns for some ILCs and overlapping patterns for ILC1 cells and NK cells, whereas other ILC subsets remained indistinguishable. We identified a transcriptional program shared by small intestine ILCs and a core ILC signature. We revealed and discuss transcripts that suggest previously unknown functions and developmental paths for ILCs.

The Immunological Genome (ImmGen) Project is a collaborative effort by immunologists and computational biologists who seek, through rigorously controlled protocols for data generation and analysis, to delineate gene-expression patterns of cell types across the immune system so as to better understand the immune response and comprehensively define its regulatory networks¹. In this context, we investigated the global gene-expression profiles of innate lymphoid cells (ILCs) from mice following the ImmGen Project's stringent standards¹.

ILCs are non-T, non-B lymphocytes present throughout the body that show enrichment in frequency at mucosal surfaces^{2,3}. Developing from an Id2⁺ common helper-like ILC progenitor⁴, three classes of ILCs, now known as ILC1, ILC2 and ILC3, have emerged that mirror helper T cells in both their cytokine-production profiles and their transcriptional circuitry^{2,3}. Functionally, T-bet⁺ ILC1 cells respond to interleukin 12 (IL-12), IL-15 and IL-18 to produce interferon- γ (IFN- γ); GATA-3⁺ ILC2 cells react to IL-33, IL-25 and thymic stromal lymphopoietin to produce type 2 cytokines, including IL-5 and IL-13; and ROR γ ⁺ ILC3 cells are activated by IL-1 and IL-23 to produce IL-22 and/or IL-17. As innate sources of distinct cytokines, ILCs have roles in early defense against infections, modulation of the adaptive immune response, the development of lymphoid tissue, and repair and homeostasis of tissues^{2,3}.

Whereas shared transcription factors, cell-surface markers and functional properties of cytokine production define classes, deviations in one or more categories by subpopulations define subsets of ILCs within the larger class. In the mouse, ILC2 seems to be the most homogeneous class, defined by expression of the IL-7 receptor (IL-7R; also called CD127), Sca-1 and the IL-33 receptor ST2. In comparison,

at least five subsets of mouse ILC3 cells have been reported, four of which are found in the greatest numbers during steady state in the adult small intestine, and one in the adult large intestine. These include CD4⁺ and CD4⁻ subsets of NKp46⁻ROR γ ⁺ lymphoid tissue-inducer (LTi)-like cells⁵; ROR γ ⁺T-bet⁺ receptor Notch-dependent NKp46⁺ ILC3 cells⁵⁻⁹; a potential ILC3-ILC1 transitional subset that has downregulated ROR γ ⁺, produces IFN- γ and expresses T-bet and larger amounts of the activating natural killer (NK) cell receptor NK1.1 compared to NKp46⁺ ILC3 cells ('ex-ROR γ ⁺ ILC3 cells')⁴; and IL17-producing ROR γ ⁺NKp46⁻ ILC3 cells in the large intestine¹⁰.

ILC1 cells have the most complicated and controversial distinction, both of the ILC class itself and between subsets within the class. ILC1 cells and NK cells have functional similarities, mainly IFN- γ production, and share expression of T-bet and many cell-surface markers, such as NKp46 and NK1.1. However, NK cells are currently thought to have more cytotoxic potential than ILC1 cells have. Surface expression of CD127 and the integrin subunit CD49a are used to distinguish ILC1 cells from NK cells in many but not all mouse tissues, as there is considerable diversity of ILC1 subsets among tissues². Lineage-tracing experiments have shown that ILC1 cells and NK cells originate from distinct progenitors^{4,11}, and mature NK cells are dependent on the transcription factor eomesodermin (Eomes), whereas ILC1 cells are not. However, immature NK cells also share many markers with ILC1 cells¹² and lack expression of Eomes^{12,13}.

The breadth of polarization and relationships among ILC subsets has remained incompletely understood. To better understand the functional differences between ILC classes and reported subsets within a class, we discriminated seven populations of ILCs in the small intestine lamina propria (siLP) (NK cells, ILC1 cells, ILC2 cells

¹Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri, USA. ²Department of Surgery, Washington University School of Medicine, St. Louis, Missouri, USA. ³Merck Research Laboratories, Palo Alto, California, USA. ⁴Cancer and Inflammation Program, Center for Cancer Research, National Cancer Institute, Frederick, Maryland, USA. ⁵A list of members and affiliations appears at the end of the paper. Correspondence should be addressed to M.C. (mcolonna@wustl.edu).

Received 17 October 2014; accepted 22 December 2014; published online 26 January 2015; doi:10.1038/ni.3094

and four subsets of ILC3 cells); three additional subsets of ILC1 cells from liver, spleen and small intestine intraepithelial lymphocytes (siIELs); and two NK cell subsets from liver and spleen. Our findings provide a molecular definition of ILC classes and subsets, and also identify a core signature in ILCs distinct from that in NK cells. We also identify novel targets for future investigation and generate a comprehensive, high-quality and publically available resource of ILC transcriptomes.

RESULTS

Analysis of ILC frequency and diversity

We isolated all reported ILC subsets in the siLP of 6-week-old C57BL/6 male mice. We isolated ILC2 cells from wild-type C57BL/6 mice and isolated all other subsets from ROR γ ^{eGFP} reporter mice, which express enhanced green fluorescent protein (eGFP) driven by the gene encoding ROR γ . These subsets included CD127⁻ NK cells, CD127⁺ ILC1 cells, CD4⁻NKp46⁻ and CD4⁺NKp46⁻ LTi-like ILC3 cells, NKp46⁺ROR γ ^{hi} ILC3 cells and NKp46⁺ROR γ ^{lo} ILC3 cells (Fig. 1a). Notably, NKp46⁺ROR γ ^{hi} and NKp46⁺ROR γ ^{lo} ILC3 cells cannot be discriminated by intracellular staining of ROR γ in wild-type C57BL/6 mice, in which only one NKp46⁺ROR γ ⁺ subset is detectable; thus, the use of ROR γ ^{eGFP} reporter mice provided the unique opportunity to profile ROR γ ^{lo} and ROR γ ^{hi} subsets, as described¹⁴. As the ROR γ ^{lo} ILC3 subset had higher expression of NK1.1 than ROR γ ^{hi} ILC3 cells had (Supplementary Fig. 1a–c), we reasoned that this subset would show enrichment for converted or actively converting ‘ex-ROR γ ⁺’ ILC3 cells. We also profiled liver CD49b⁺TRAIL⁻ NK cells and CD49b⁻TRAIL⁺ ILC1 cells and spleen CD127⁻ NK cells and CD27⁺CD127⁺ ILC1 cells (Fig. 1a), the last of which have been reported but have not previously been called ‘ILC1 cells’¹⁵. Small intestine intraepithelial ILC1 cells were isolated from the intestinal epithelium as NK1.1⁺NKp46⁺ (Fig. 1a). These cells are phenotypically distinct from NK cells as a result of imprinting with transforming growth factor- β ¹⁶, although their developmental origin and transcriptional relationship to siLP ILC subsets remain unclear.

Cytospins showed that ILC subsets were morphologically pure lymphoid populations (Fig. 1b). We assessed the frequency of these 12 ILC subsets within the lymphocyte populations of the small intestine, liver and spleen in naive mice (Fig. 1c). ILCs were most abundant in the siLP, followed by liver and spleen; siIEL populations had the lowest frequency of ILCs. Next we sorted these populations according to the ImmGen Project’s standardized protocol for data generation and analyzed gene expression by means of whole-mouse genome array. Principal-component analysis (PCA) showed a greater degree of diversity generated by ILC2 and ILC3 subsets than by ILC1 cells (Fig. 1d and Supplementary Fig. 1e). In hierarchical clustering, three pairs of ILC subsets were computationally indistinguishable (Fig. 1e): splenic and liver NK cells, ROR γ ^{hi}NKp46⁺ and ROR γ ^{lo}NKp46⁺ ILC3 cells, and CD4⁺NKp46⁻ and CD4⁻NKp46⁻ LTi-like ILC3 cells were intermixed. All ILC1 subsets clustered separately from NK cells from the same tissue. However, siIEL ILC1 cells, siLP NK cells and siLP ILC1 cells clustered in a separate branch of the dendrogram from liver and spleen NK cells and ILC1 subsets. We concluded that beyond classical polarizations, environmental factors in the small intestine differentiated intestinal subsets from those in liver and spleen.

Expression profiles of individual ILC subsets

We assessed the gene-expression profile of each sorted subset (Fig. 2a–h) by identifying characteristic transcripts that were expressed at least twofold or fourfold higher in the index subset than in all other profiled subsets. Heat maps demonstrate the extent to which the identified

transcripts were specifically expressed by the index subset (Fig. 2a–h). We found the greatest number of characteristic transcripts in the siLP ILC2 cells, with 100 transcripts showing expression more than twofold higher than that in other subsets (Supplementary Table 1) and 34 transcripts showing expression more than fourfold higher than that in other subsets (Fig. 2a). These included expected transcripts, such as *Il13*, *Il5*, *Il4*, *Il9r* and *Il17rb* (which encodes the IL-25 receptor)^{17–19}, as well as several transcripts not previously shown to be expressed by ILC2 (to our knowledge), including *Rxrg*, *Pparg*, *Mc5r*, *Dgat2* and *Alox5* (Fig. 2a). The transcriptional repressor RXR γ , which binds the vitamin A metabolite 9-*cis* retinoic acid²⁰, has not been previously described in ILC2 cells. However, vitamin A is known to directly inhibit ILC2 differentiation by an unknown mechanism²¹, which suggests that RXR γ may mediate this effect. We also found previously unrecognized expression of the gene encoding the transcription factor PPAR γ (Fig. 2a), which forms heterodimers with repressor retinoid X receptors and transcriptionally mediates lipid homeostasis²⁰. ILC2 cells expressed several other genes encoding molecules linked to lipid metabolism. These included genes encoding *Dgat2*, which mediates the final reaction of triglyceride synthesis; *Mc5r*, a receptor that causes lipid mobilization in adipocytes; and *Alox5*, a lipoxygenase that catalyzes synthesis of leukotriene A4 (Fig. 2a). ILC2 cells are known to regulate the cellular immune system within visceral adipose tissues²², but our data suggested that ILC2 cells might also directly sense lipids and produce lipid mediators.

NKp46⁻ LTi-like ILC3 subsets expressed the second greatest number of characteristic transcripts. As suggested by clustering analysis (Fig. 1e), CD4⁺ and CD4⁻ LTi-like cells had overlapping gene-expression patterns. CD4⁺ LTi-like cells expressed only four transcripts more than twofold higher when compared to all other subsets, one of which was *Cd4*, encoding the monomorphic coreceptor CD4; in contrast, no transcripts had characteristic higher expression in CD4⁻ LTi-like ILC3 cells (Fig. 2b). However, 65 transcripts had expression more than twofold higher in both subsets together relative to other subsets (Supplementary Table 1), and 9 of these transcripts were expressed more than fourfold higher than in other subsets (Fig. 2b). Eight additional transcripts were expressed at least fourfold higher than in other subsets by either CD4⁺ LTi-like or CD4⁻ LTi-like ILC3 cells, with differences in expression among LTi-like ILC3 subsets probably due to replicate variation (Fig. 2b). Transcripts expressed at higher levels by both LTi-like ILC3 subsets included those encoding the chemokine receptor CCR6, which has been used as a marker for LTi-like ILC3 cells⁷, and the chemokine receptor CXCR5 (Fig. 2b). We also found several genes not previously described in LTi-like ILC3 cells (to our knowledge), including *Gucy1a3*, *Cntn1*, *Slc6a7*, *Cacna1g* and *Nrp1* (Fig. 2b). *Gucy1a3* encodes the α -subunit of the soluble guanylate cyclase receptor, which transduces signals from nitric oxide; however, we found no expression in LTi-like ILC3 cells of the other components of the guanylate cyclase receptor (data not shown). *Cntn1*, which encodes a glycosylphosphatidylinositol-linked member of the immunoglobulin family, is best known for its role in regulating axonal guidance and neural system development²³ and has not previously been studied in an immune context. The L-proline transporter encoded by *Slc6a7* and the voltage-gated calcium channel encoded by *Cacna1g* are similarly atypical and are not expressed in other cells of the immune system (data not shown). We concluded that LTi-like ILC3 cells specifically expressed several transcripts that are unique in the immune system, including those encoding putative factors involved in neural crosstalk.

The remaining ILC subsets had fewer candidate characteristic markers, probably because of multiple comparisons with other subsets in

the same class. As indicated by PCA, siLP NKp46⁺RORγ^{hi} ILC3 cells and siLP NKp46⁺RORγ^{lo} ILC3 cells had overlapping gene-expression profiles. Sixteen transcripts showed expression that was twofold

higher in NKp46⁺RORγ^{hi} ILC3 cells than in all other profiled ILCs except NKp46⁺RORγ^{lo} ILC3 cells, although a heat map revealed that most of these genes were also expressed at lower levels by other siLP

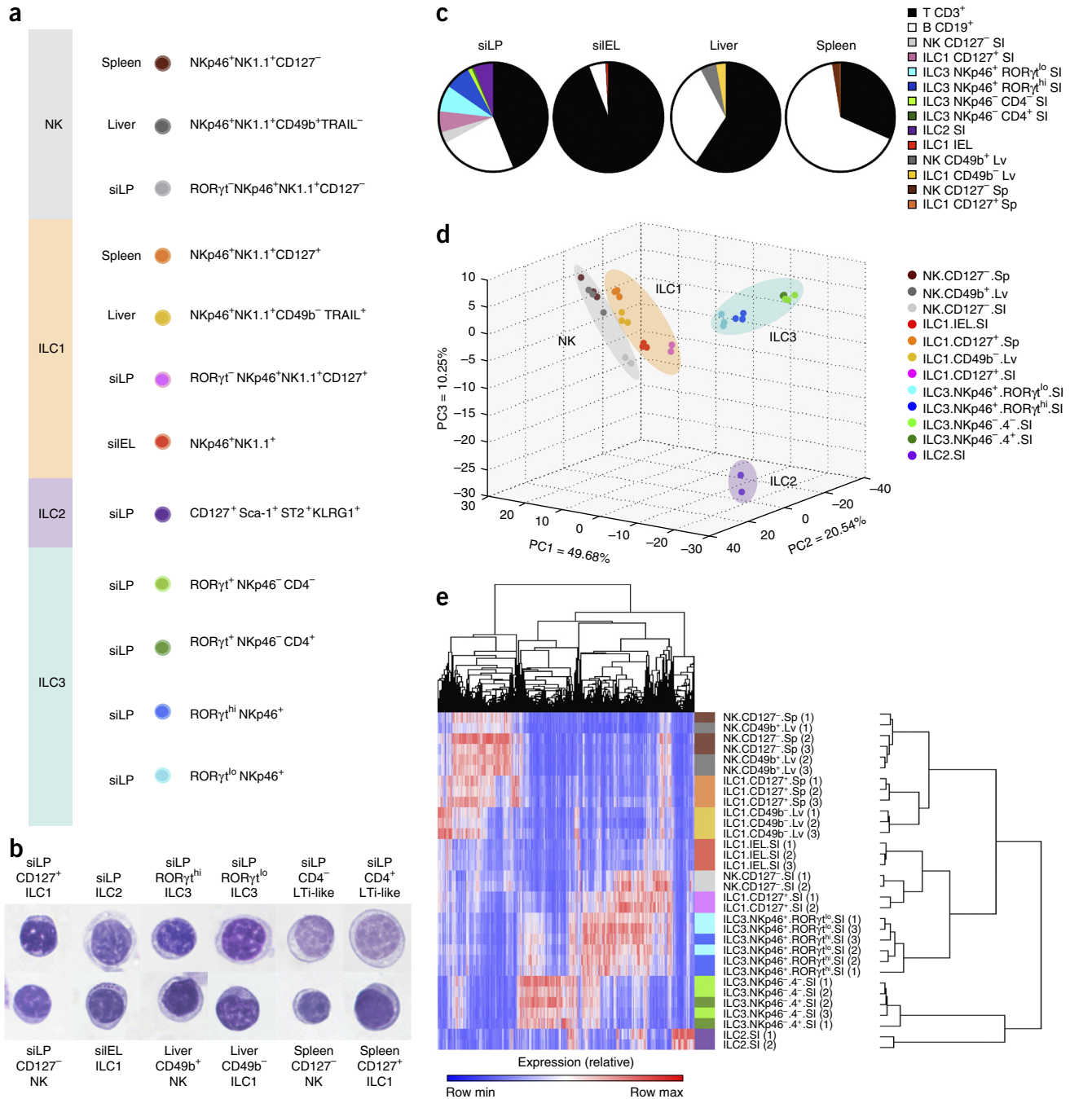


Figure 1 Analysis of ILC frequency and diversity. **(a)** Sorting strategy for array analysis after gating on live CD45⁺ and CD3⁻CD19⁻ cells. **(b)** Cytospins of cells using sorting strategies presented in **a** for each population (above and below images). Original length (of each panel), 14.6 μm. **(c)** Flow cytometry of cells from various tissues, showing the frequency of ILCs among the lymphocyte population. The gating strategy in **a** was used to distinguish CD3⁺ T cells, CD19⁺ B cells and ILCs within the same sample, except the siLP ILC2 cell marker KLRG1 and the spleen ILC1 cell marker CD27 were excluded and liver ILCs were distinguished with NKp46, CD49b and CD49a. T, T cell; B, B cell; NK, NK cell; SI, small intestine; Lv, liver; Sp, spleen. **(d)** PCA of gene expression by subsets of ILCs and NK cells. Numbers along axes indicate relative scaling of the principal variables. ImmGen nomenclature: spleen NK, NK.CD127⁻.Sp; liver NK, NK.CD49b⁺.Lv; siLP NK, NK.CD127⁻.SI; siEL ILC1, ILC1.IEL.SI; spleen ILC1, ILC1.CD127⁺.Sp; liver ILC1, ILC1.CD49b⁻.Lv; siLP ILC1, ILC1.CD127⁺.SI; siLP NKp46⁺RORγ^{lo} ILC3, ILC3.NKp46⁺.RORγ^{lo}.SI; siLP NKp46⁺RORγ^{hi} ILC3, ILC3.NKp46⁺.RORγ^{hi}.SI; siLP NKp46⁺CD4⁻ LTi-like ILC3, ILC3.NKp46⁺.4⁻.SI; siLP NKp46⁺CD4⁺ LTi-like ILC3, ILC3.NKp46⁺.4⁺.SI; and siLP ILC2, ILC2.SI. **(e)** Hierarchical clustering of subsets of ILCs and NK cells based on the 10% of genes with the greatest variability. ImmGen nomenclature as in **d**. Data are representative of two independent experiments (**b,c**) with *n* = 1–2 mice per tissue (**b**) or *n* = 2–4 mice per tissue (**c**) or are pooled from one to three experiments per sample with cells pooled from 3–5 mice each (**d,e**).

subsets (Fig. 2c). Liver and splenic ILC1 cells each expressed some characteristic transcripts with a change in expression of greater than twofold relative to their expression in all other subsets (Fig. 2d,e),

but we found no transcripts with relative expression greater than twofold higher in siLP ILC1 cells. This suggested there were few characteristic factors expressed by individual ILC1 subsets among

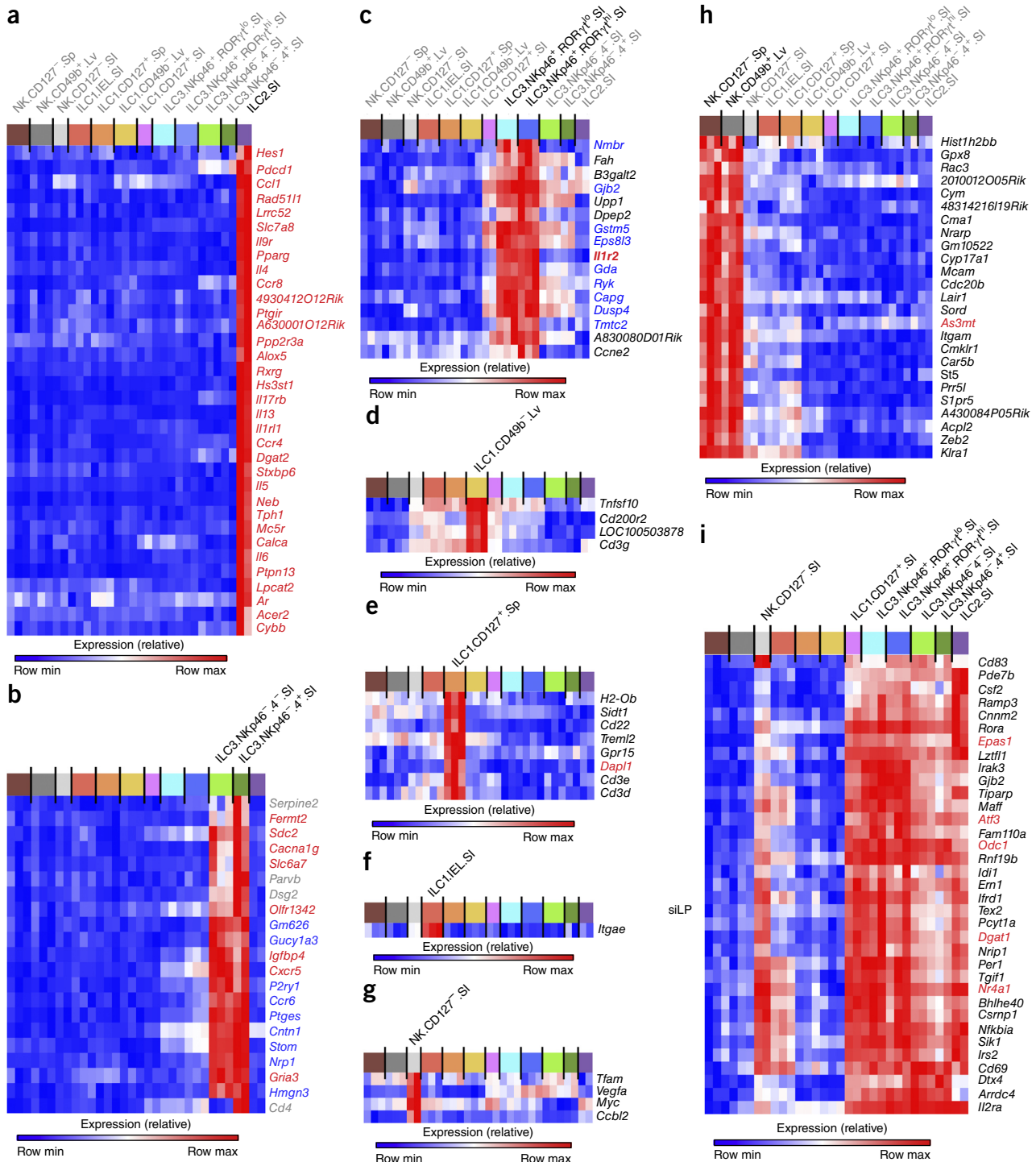


Figure 2 Characteristic transcripts of individual ILC subsets. Transcripts upregulated more than fourfold (red font) or twofold (black font) in individual subsets (a,d-g), two similar subsets (b,c,h) or all subsets from siLP (i), based on pairwise comparisons between the index subset(s) (in bold above heat map) and all other subsets (full gene lists, **Supplementary Table 1**). For similar ILC3 cell subsets consisting of NKp46⁻CD4⁺ LTI-like and CD4⁻ LTI-like cells (b) or NKp46⁺RORγ^{hi} and NKp46⁺RORγ^{lo} cells (c), some transcripts had expression that was more than fourfold higher (b) or twofold higher (c) in one subset but not the other; in this case, blue indicates shared transcripts expressed by both subsets. Gray (b, right margin) indicates four transcripts with twofold higher expression in CD4⁺ LTI-like ILC3 cells than in CD4⁻ LTI-like ILC3 cells and fourfold higher expression in CD4⁺ LTI-like ILC3 cells than in all other subsets. ImmGen nomenclature as in **Figure 1**. Data are pooled from one to three experiments per sample with cells pooled from three to five mice each; two to three replicate samples are shown per subset.

all ILCs and NK cells. Unexpectedly, the only transcript that showed expression greater than twofold higher in siIEL ILC1 cells compared to other ILC subsets was *Itgae*, which encodes CD103 (integrin $\alpha_E\beta_7$) (Fig. 2f). Human IEL ILC1 cells express CD103¹⁶, but it was not previously known to be expressed by mouse IEL ILC1 cells, which lack surface expression of CD103; this would suggest post-transcriptional regulation of CD103 in mouse IEL ILC1 cells. Focusing on NK cells, we identified four genes with expression that was twofold higher in siLP NK cells than in all ILCs (Fig. 2g). Splenic and liver NK cells expressed no characteristic transcripts. However, when we assessed both liver and spleen NK cells as a group, we found 25 transcripts with expression that was at least twofold higher than in all other subsets, although siLP NK cells, IEL ILC1 cells and splenic ILC1 cells also expressed these transcripts at lower levels (Fig. 2h). We concluded that within our dataset, the mRNA profiles of ILC2 cells and LTi-like ILC3 cells were unique, whereas the profiles of NKp46⁺ ILC3 cells, ILC1 cells and NK cells showed considerable overlap.

A transcriptional signature shared by all siLP subsets

We next sought to determine whether there were any transcripts that were expressed in all subsets from an individual tissue among siLP, liver and spleen. Given that siLP NK cells and ILC1 cells were found to cluster further from liver and splenic subsets by hierarchical clustering and PCA, we focused on the siLP (Fig. 1d,e). Pairwise comparisons of all subsets from the siLP with remaining subsets from liver and spleen revealed that all siLP subsets expressed a core 35-transcript signature (Fig. 2i), which included several transcription factor–encoding transcripts such as *Rora*, *Atf3*, *Nr4a1*, *Maff*, *Epas1*, *Bhlhe40* and *Per1*. Furthermore, all siLP-resident ILCs had high expression of transcript encoding the activation marker CD69 and varied expression of *Csf2*, which encodes the cytokine GM-CSF. Although the production of GM-CSF by ILC2 cells¹⁸, ILC3 cells^{24,25} and NK cells²⁶ is known, its production by ILC1 cells has not been reported, to our knowledge. Thus, ILC subsets in the siLP seemed to be more activated than ILCs in other tissues, probably because of their constant exposure to varied environmental signals from the microbiome and incoming nutrients, including well-documented transcriptional activators such as vitamin A^{21,27} and ligands of the transcription factor AhR (aryl hydrocarbon receptor)^{6,28}.

Transcription factors, cytokines and chemokines of ILC subsets

To address broad patterns of gene expression among ILC subsets and classes, we investigated the expression of previously reported and other (unreported) transcription factors, chemokines, cytokines and other secreted factors. The transcription factors with the highest relative expression included the well-documented ILC-defining *Id2* (encoded by *Id2*), the ILC3 class-defining ROR γ t (encoded by *Rorc*), the ILC1- and NKp46⁺ ILC3-defining T-bet (encoded by *Tbx21*) and the NK cell-defining *Eomes* (encoded by *Eomes*) (Fig. 3a). ILC2 cells showed higher expression of the ILC2-defining transcription factors GATA-3 (encoded by *Gata3*) and ROR α (encoded by *Rora*), but these were also expressed by all ILCs (Fig. 3a), consistent with an early role in ILC development, at least for GATA-3 (ref. 19). *Nfil3*, which encodes the transcription factor NFIL3 (also known as E4BP4), is required for the development of NK cells^{29,30}, ILC2 cells and ILC3 cells^{31,32}, and had its highest expression in siLP subsets; NKp46⁺ ILC3 cells, ILC1 cells and NK cells showed the highest *Nfil3* transcript levels, followed by LTi-like ILC3 and ILC2 cells, whereas much lower *Nfil3* transcript levels were present in liver and spleen ILC1 and NK cells (Fig. 3a). Furthermore, two transcription factors identified in the siLP signature, ATF3 (encoded by *Atf3*) and Nur77 (encoded

by *Nr4a1*) (Fig. 2i), were expressed at levels similar to those of lineage-defining transcription factors (Fig. 3a). Collectively, these data suggested a substantial role for the intestinal microenvironment in the expression of certain transcription factors, which might subsequently have diverging roles among ILC classes. Analysis of chemokines and their receptors (Fig. 3b), as well as of cytokines and their receptors (Fig. 3c), revealed both shared and distinct expression patterns. Beyond the known signature cytokine and chemokine circuitries, we identified a candidate feed-forward loop for ILC2 cells, which expressed both the chemokine receptor CCR8 and its ligand CCL1 (Fig. 3b). We also identified ILC2 cell expression of *Bmp7* and *Bmp2*, the latter of which encodes a protein known to modulate intestinal peristalsis by binding the BMP receptor on enteric neurons³³. In addition, we found expression of *Il2* in several ILC populations (Fig. 3c), which suggested that ILCs might be able to activate T cells or other ILCs through signaling via its receptor, IL-2R.

Shared and distinct expression profiles among siLP subsets

We next focused our analysis of transcriptional profiles on the four major CD127⁺ ILC subsets within the siLP: ILC1 cells, ILC2 cells, NKp46⁺ ROR γ t^{hi} ILC3 cells and CD4⁻ LTi-like ILC3 cells (Fig. 3d). Comparison of siLP ILC subsets revealed overlapping patterns of gene expression that were not identified in individual subset signatures (Fig. 3d and Supplementary Table 2). For example, ILC2 cells and LTi-like ILC3 cells shared 17 transcripts, including *Arg1* and *Ret*. Arginase-1 (encoded by *Arg1*) marks fetal and adult ILCs and facilitates the identification of developing ILCs in the siLP³⁴. The receptor tyrosine kinase encoded by *Ret* is also expressed by fetal CD11c⁺ lymphoid tissue initiator cells and is required for the development of Peyer's patches³⁵, but it has not, to our knowledge, been previously reported to be expressed by fetal or adult ILCs, including LTi or LTi-like ILC3 cells. Together these results suggested that, at least in fetal mice, ILC2 cells and LTi-like ILC3 cells share a common progenitor³⁴, although their functional relevance in adult siLP ILCs remains to be investigated. ILC1 cells and ILC2 cells shared 19 transcripts, including *Ets2*. *Ets-1* and *Ets-2* interact with proteins of the Id family, and whereas *Ets-1* has been linked to the early development of NK cells³⁶, *Ets-2* has not. Thus, *Ets-2* might be relevant to the development or maintenance of ILC1 and ILC2 cells.

NKp46⁺ and LTi-like ILC3 cells have long been known to share many characteristics, owing to their mutual production of IL-22 and expression of ROR γ t. However, NKp46⁺ ILC3 cells and NKp46⁺ ILC1 cells shared higher relative expression of a greater number of transcripts (*Tbx21*, *Ifng* and *Il12rb*) than did NKp46⁺ ILC3 cells and LTi-like ILC3 cells (Fig. 3d and Supplementary Table 2). Although T-bet is required for the development of NKp46⁺ ILC3 cells^{8–10}, these cells produced little IFN- γ in response to IL-12 and IL-23 (data not shown). Because IFN- γ is well documented as being post-transcriptionally regulated³⁷, the presence of the *Ifng* transcript in NKp46⁺ ILC3 cells suggested that T-bet might be sufficient to induce transcription but that other factors are needed for protein production, such as bacterial infections *in vivo*⁴. In addition to their shared transcripts, NKp46⁺ ILC3 cells and NKp46⁺ ILC1 cells had significantly different expression (by greater than twofold) of 213 genes, with the genes upregulated in NKp46⁺ ILC3 cells including many genes that were also expressed at higher levels by LTi-like ILC3 cells (Fig. 3e). Thus, NKp46⁺ ILC3 cells had a transcriptional profile with characteristics intermediate between those of NKp46⁻ LTi-like ILC3 cells and NKp46⁺ ILC1 cells, which might enable functional plasticity. Their functional polarization toward ILC3 or ILC1 cells probably depends on the tissue microenvironment.

As discussed above, a pairwise comparison did not identify characteristic transcripts in siLP ILC1 cells compared to all profiled ILC subsets. However, in comparisons only to siLP ILC2 cells and ILC3 subsets, we found 75 transcripts expressed at least twofold higher in ILC1 cells (Fig. 3d). For example, ILC1 cells demonstrated a greater cytotoxic capacity, as indicated by their expression of *Gzma* and *Prf1* (Fig. 3d), which respectively encode granzyme A and perforin, although this might have been due to an imperfect distinction between siLP ILC1 cells and NK cells (discussed below). ILC1 cells

also had substantial expression of *Il21r*, which encodes a member of the common γ -chain cytokine family (Fig. 3d). When we included an additional comparison between ILC1 cells and NK cells in the siLP, we found that ILC1 cells expressed only four transcripts at least twofold higher than other siLP subsets: *Gpr55*, *Trat1*, *Mmp9* and *Cpne7* (Supplementary Table 2). Thus, although ILC1 cells from the small intestine were more like NK cells than were other ILC subsets from the small intestine, they showed no obvious characteristic markers when we included NK cells in our comparisons.

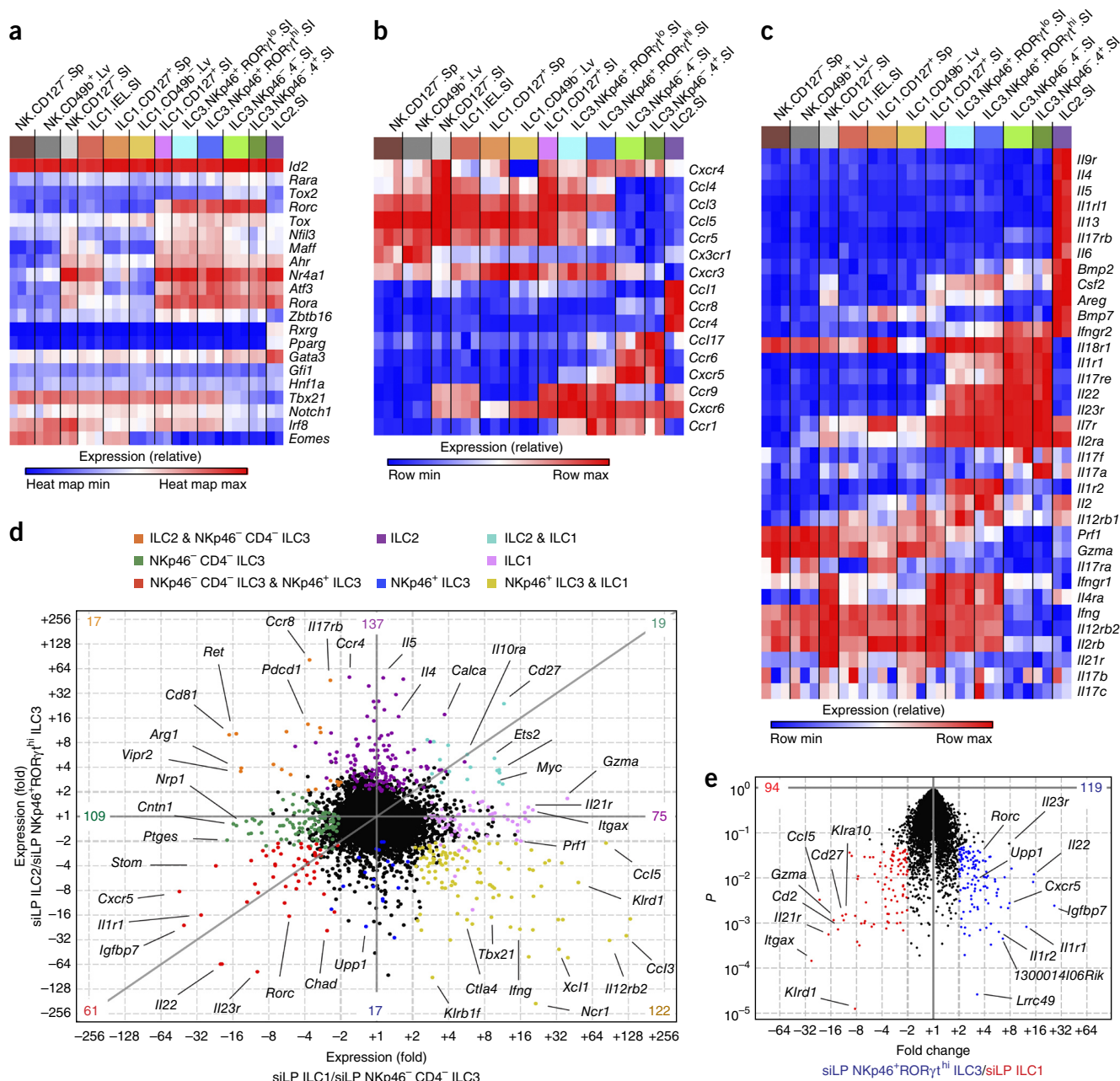


Figure 3 Spectrum of distinct and shared transcriptional profiles among ILC subsets. **(a)** Heat map of the expression of selected transcription factors, normalized by expression within the entire heat map. **(b,c)** Heat maps of the expression of selected chemokine receptors and ligands **(b)** and of cytokine receptors and ligands **(c)**, normalized by expression within each row. **(d)** Transcripts upregulated by a single subset or two subsets among selected siLP ILCs (colors in plot match key; full gene lists, Supplementary Table 2). Numbers in plot indicate transcripts upregulated by at least twofold in that subset (colors match key). **(e)** Comparison of gene expression in RORγt^{hi} NKp46⁺ ILC3 cells (*n* = 3 replicates) versus that in ILC1 cells (*n* = 2 replicates) by volcano plot; numbers in corners indicate transcripts significantly upregulated by at least twofold. *P* ≤ 0.05 (*t*-test). ImmGen nomenclature as in Figure 1. Data are pooled from one to three **(a-c)** or one to two **(d,e)** experiments per sample with cells pooled from three to five mice each; two to three replicate samples are shown per subset **(a-c)**.

Defining novel transcripts within ILC3 cells

Pairwise comparison of all ILC3 subsets versus other profiled cells revealed a 42-transcript ILC3 cell signature (Fig. 4a). This signature included well-studied transcripts such as *Il23r*, *Rorc* and *Il22*, as well as several molecules previously unreported to be expressed by ILC3 cells, such as *Pram1*, a target of retinoic acid and activator of the kinase Jnk. This result supported the role for retinoic acid in the development of ILC3 cells^{21,27}. Notably, *Il17* was not among the transcripts with expression more than twofold higher in ILC3 cells than in other ILCs, nor was it expressed uniquely by any individual ILC3 cell subset (Fig. 2b,c). This result suggested that IL-17 was not a major product of ILC3 cells in the small intestine, at least in young adult mice at steady state.

Comparison of the major adult LTI-like population, CD4⁻ LTI-like ILC3 cells, with NKp46⁺RORγ^{hi} ILC3 cells revealed a total of 508 genes with a significant difference in expression of greater than twofold (Fig. 4b and Supplementary Table 3). Among the transcripts

most highly and most significantly upregulated by CD4⁻ LTI-like ILC3 cells, we found *Nrp1*. Also identified as part of our LTI-like ILC3 cell signature (Fig. 2b), *Nrp1* is a coreceptor for several ligands, including the immunoregulatory factor VEGF, transforming growth factor-β1 and semaphorins, and may have a function in negatively regulating the immune response, in part through enhanced survival of regulatory T cells³⁸. It is also one of a few markers that distinguish natural regulatory T cells from peripherally generated mucosa-derived induced regulatory T cells^{39,40}. However, to our knowledge, until now *Nrp1* has never been reported to be expressed by LTI-like ILC3 cells. We confirmed by flow cytometry that *Nrp1* was expressed in greater amounts in LTI-like ILC3 cells than in other siLP subsets (Fig. 4c). We also found higher expression of CD25 protein in LTI-like ILC3 cells than in other siLP subsets (Fig. 4c). Gating on *Nrp1*⁺CD25⁺ cells among CD3⁻CD19⁻ siLP lymphocytes resulted in a cell population with considerable enrichment for LTI-like ILC3 cells (Fig. 4d) and production of IL-22 in naive IL-22 reporter mice⁴¹ (Fig. 4e). Among

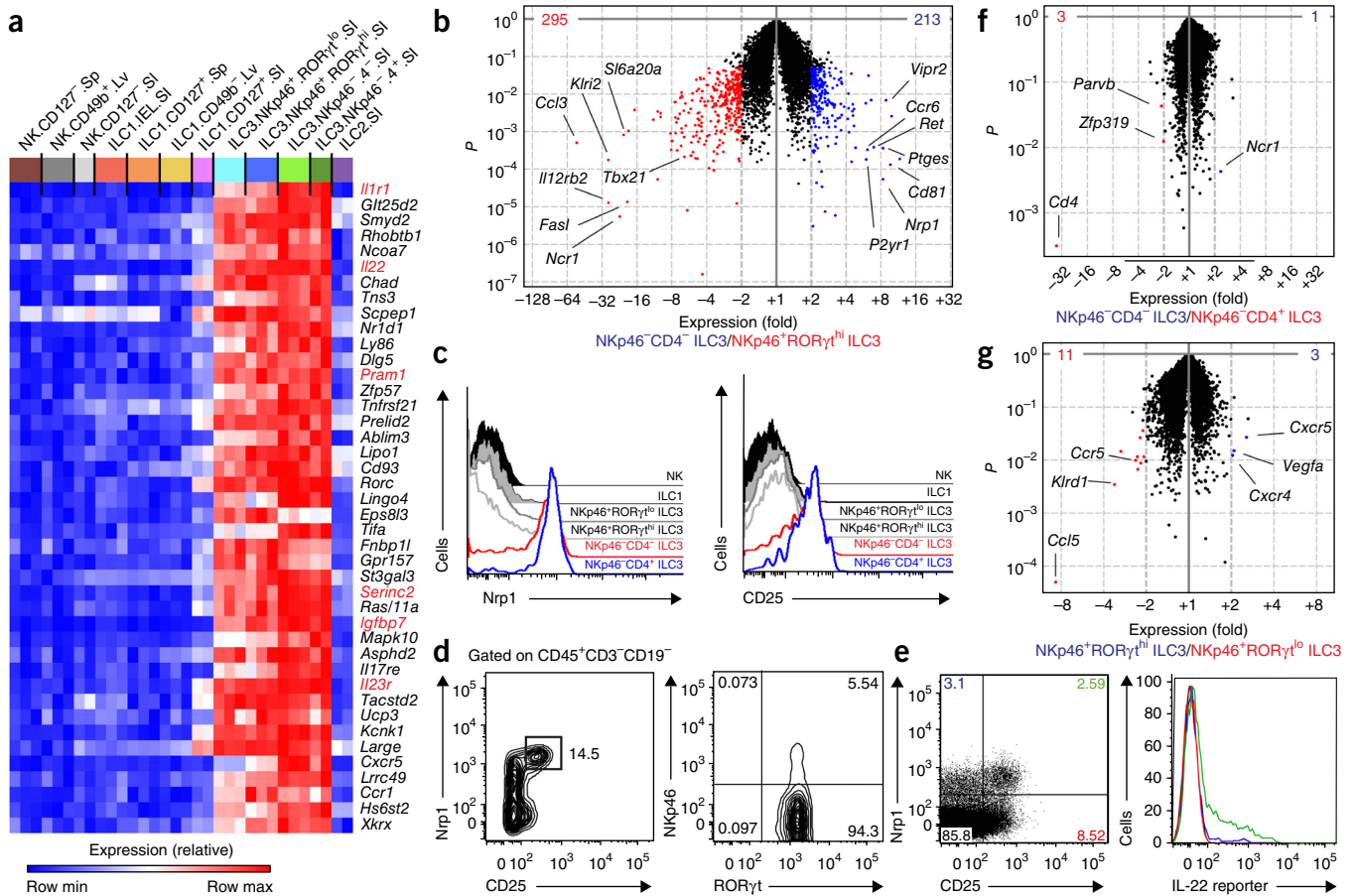


Figure 4 ILC3-specific transcriptional programs and cell-surface markers. (a) Heat map of mRNA transcripts upregulated more than fourfold (red font) or twofold (black font) in RORγ⁺ ILC3 cell subsets relative to their expression in all other subsets. ImmGen nomenclature as in Figure 1. (b) Comparison of gene expression in NKp46⁺RORγ^{hi} ILC3 cells with that in CD4⁻ LTI-like ILC3 cells (*n* = 3 replicates each), by volcano plot; numbers in corners indicate transcripts significantly upregulated (*P* ≤ 0.05 (*t*-test)) by at least twofold (colors in plot match those below plot; full gene lists, Supplementary Table 3). (c) Expression of *Nrp1* and CD25 by siLP NK cell, ILC1 cell, NKp46⁺RORγ^{hi} ILC3 cell, NKp46⁺RORγ^{hi} ILC3 cell, NKp46⁻CD4⁻ LTI-like ILC3 cell and NKp46⁻CD4⁺ LTI-like ILC3 cell subsets from RORγ^{EGFP} reporter mice. (d) Percentage of siLP cells from C57BL/6 wild-type mice electronically gated on CD45⁺ and CD3⁻CD19⁻ cells that expressed both CD25 and *Nrp1* (left), and extent of LTI-like ILC3 phenotype among these cells (from left outlined area) based on surface NKp46 and intracellular RORγ^t stainings (right). Number adjacent to outlined area (left) indicates percent *Nrp1*⁺CD25⁺ cells; numbers in quadrants (right) indicate percent cells in each throughout. (e) Production of the IL-22 reporter (right) by cells obtained from naive IL-22 reporter mice and shown (left) with gating on CD45⁺, CD3⁻CD19⁻ and NKp46⁻ cells (colors of lines (left) match those in quadrants (right)). (f,g) Comparison of gene expression in CD4⁺ LTI-like cells (*n* = 2 replicates) with that in CD4⁻ LTI-like ILC3 cells (*n* = 3 replicates) (f) and in NKp46⁺RORγ^{hi} ILC3 cells with that in NKp46⁺RORγ^{lo} ILC3 cells (*n* = 3 replicates each) (g) by volcano plot (additional information, Supplementary Table 3). *P* ≤ 0.05 (*t*-test). Data in c–e are representative of three experiments.

© 2015 Nature America, Inc. All rights reserved. npg

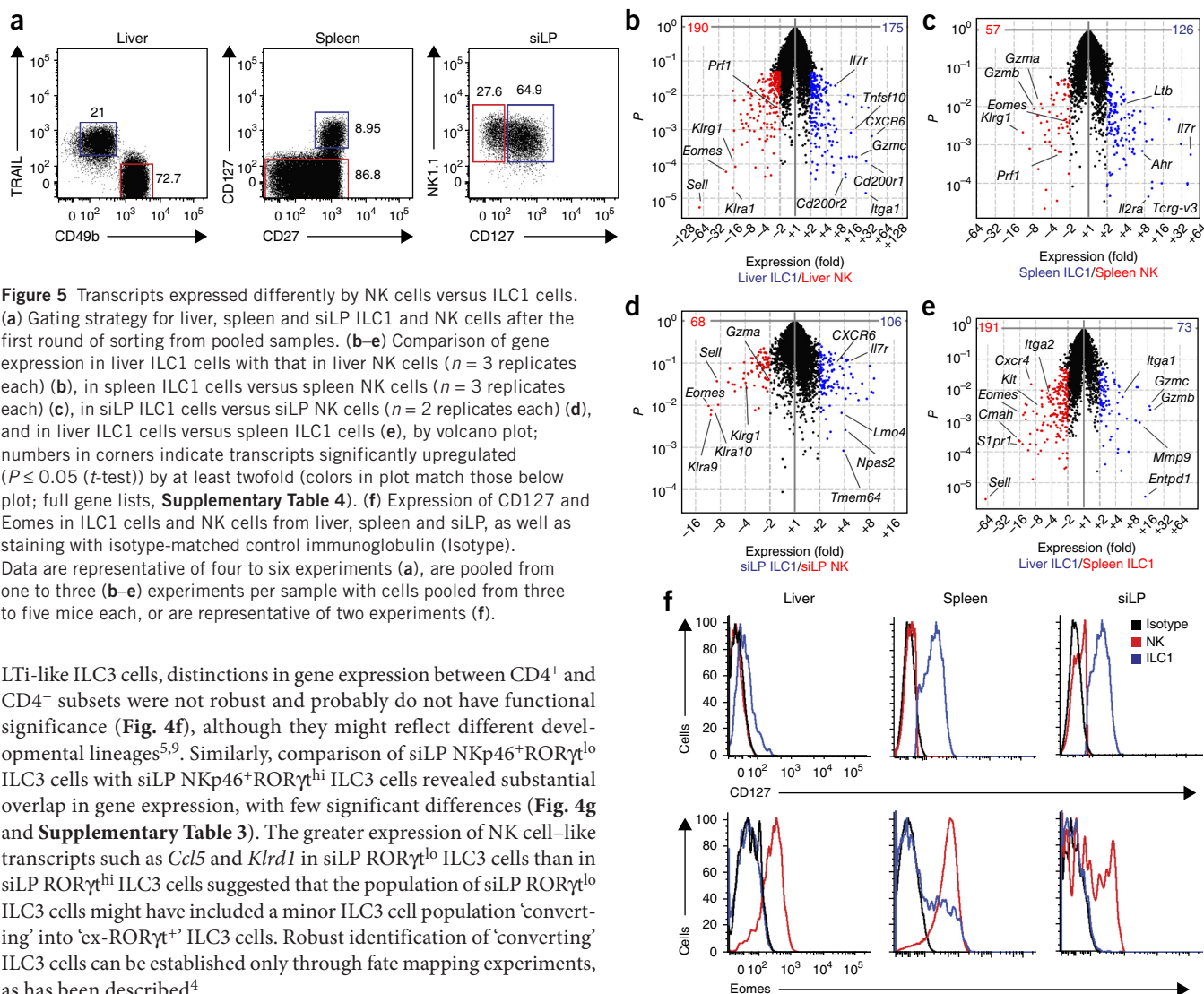


Figure 5 Transcripts expressed differently by NK cells versus ILC1 cells. (a) Gating strategy for liver, spleen and siLP ILC1 and NK cells after the first round of sorting from pooled samples. (b–e) Comparison of gene expression in liver ILC1 cells with that in liver NK cells ($n = 3$ replicates each) (b), in spleen ILC1 cells versus spleen NK cells ($n = 3$ replicates each) (c), in siLP ILC1 cells versus siLP NK cells ($n = 2$ replicates each) (d), and in liver ILC1 cells versus spleen ILC1 cells (e), by volcano plot; numbers in corners indicate transcripts significantly upregulated ($P \leq 0.05$ (t -test)) by at least twofold (colors in plot match those below plot; full gene lists, **Supplementary Table 4**). (f) Expression of CD127 and Eomes in ILC1 cells and NK cells from liver, spleen and siLP, as well as staining with isotype-matched control immunoglobulin (Isotype). Data are representative of four to six experiments (a), are pooled from one to three (b–e) experiments per sample with cells pooled from three to five mice each, or are representative of two experiments (f).

LTi-like ILC3 cells, distinctions in gene expression between CD4⁺ and CD4⁻ subsets were not robust and probably do not have functional significance (Fig. 4f), although they might reflect different developmental lineages^{5,9}. Similarly, comparison of siLP NKp46⁺RORγ^{lo} ILC3 cells with siLP NKp46⁺RORγ^{hi} ILC3 cells revealed substantial overlap in gene expression, with few significant differences (Fig. 4g and **Supplementary Table 3**). The greater expression of NK cell-like transcripts such as *Ccl5* and *Klra1* in siLP RORγ^{lo} ILC3 cells than in siLP RORγ^{hi} ILC3 cells suggested that the population of siLP RORγ^{lo} ILC3 cells might have included a minor ILC3 cell population ‘converting’ into ‘ex-RORγ⁺’ ILC3 cells. Robust identification of ‘converting’ ILC3 cells can be established only through fate mapping experiments, as has been described⁴.

Transcriptional differences between ILC1 cells and NK cells

One subject that has been particularly controversial is the difference between ILC1 cells and NK cells, in part because of a lack of markers characteristically and consistently expressed in NK cells and ILC1 cells in various organs. We used different sorting strategies in each tissue to discriminate between ILC1 cells and NK cells, consistent with what has been reported before^{2,42}. We achieved the best separation of subsets in the liver and the worst such separation in the siLP (Fig. 5a). Comparisons of NK cells and ILC1 cells from liver, spleen and siLP reflected the degree of separation between populations during sorting, with the greatest number of significantly differently expressed genes in the liver and the least in the siLP. Nonetheless, ILC1 replicates clustered together (Fig. 1e) and were transcriptionally distinct from NK cells in liver, spleen and siLP (Fig. 5b–d and **Supplementary Table 4**). As expected, *Eomes* expression was significantly different, with a difference in expression of greater than twofold in NK cells relative to its expression in ILC1 cells in all tissues analyzed (Fig. 5b–d and **Supplementary Table 4**). The amount of transcripts encoding proteins of the cytotoxic machinery was generally greater in NK cells than in ILC1 cells (Fig. 5b–d). However, in the siLP, the number of perforin-encoding transcripts was only 1.7-fold greater in NK cells than in ILC1 cells, and in the liver, only granzyme K was expressed at

higher levels by NK cells than by ILC1 cells (**Supplementary Table 4**). In fact, granzyme A and granzyme C had higher expression in liver ILC1 cells than in liver NK cells (Fig. 5b and **Supplementary Table 4**), consistent with the reported cytolytic activity of liver ILC1 cells⁴³. Liver and spleen ILC1 cells selectively expressed transcripts encoding known ILC1 cell-surface markers. In liver ILC1 cells, these transcripts included *Itga1* and *Tnfrsf10* (Fig. 5b), which encode CD49a and TRAIL, markers that have been used for separating liver ILC1 cells from NK cells⁴⁰. Splenic ILC1 cells had high expression of *IL2ra* and *IL7r* (Fig. 5c), which encode cytokine receptors that collectively enable this population to escape regulation by regulatory T cells¹⁵. The transcripts with the most significant higher expression in siLP ILC1 cells than in siLP NK cells included only previously uncharacterized transcripts such as *Tmem64*, *Npas2* and *Lmo4*; however, these transcripts were expressed in other ILCs of the small intestine (Fig. 5d) and therefore probably would not be useful markers. Comparison of splenic and liver ILC1 cell subsets revealed that splenic ILC1 cells expressed markers of immature and mature NK cells, including CXCR4, c-Kit and *Eomes*^{13,42} (Fig. 5e). These differences in transcription could be explained by the finding that our sorting strategy based on CD127 and CD27 included an *Eomes*⁺ subset among the splenic ILC1 cells (Fig. 5f). We also noted that the siLP NK cell subset identified by

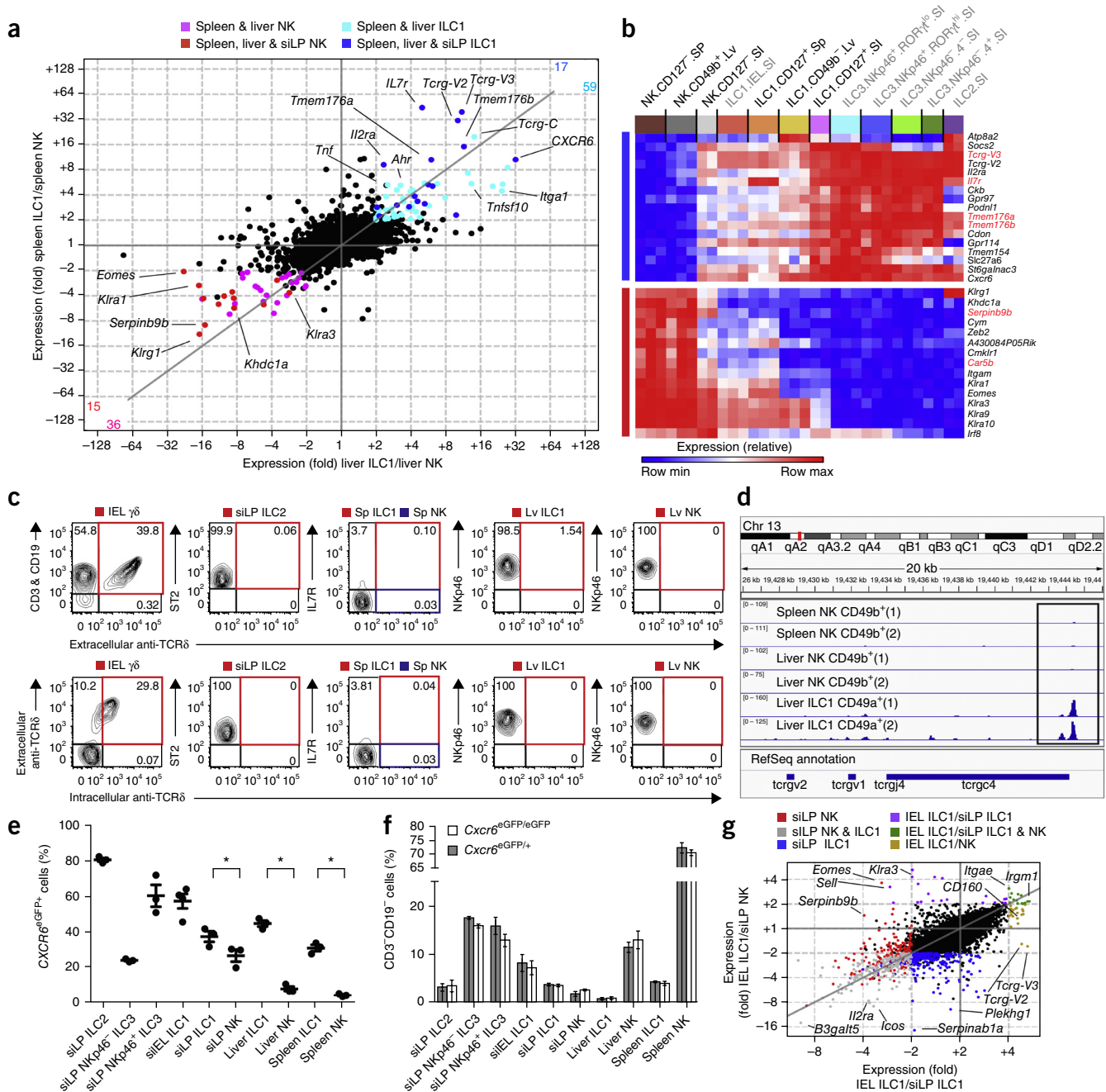


Figure 6 Generation of a core ILC signature distinct from that of NK cells. **(a)** Transcripts expressed differently by ILC1 cells relative to their expression by NK cells from various tissues (colors in plot indicate transcripts upregulated by two or three subsets and match those in key); numbers in plot indicate transcripts with a significant ($P \leq 0.05$ (t -test)) difference in expression of at least twofold in liver and spleen ILC1 and NK cells ($n = 3$ replicates each) or with an additional filter for difference in expression of at least twofold in siLP ILC1 cells versus NK cells ($n = 2$ replicates each). **(b)** Gene signatures generated according to data in **a** (colors on left match those in **a**); red font indicates transcripts upregulated by at least fourfold in all three comparisons of ILC1 cells versus NK cells (full gene lists, **Supplementary Tables 5 and 6**). **(c)** Extracellular and intracellular staining of TCR δ on (or in) ILCs and NK cells (electronically gated as in **Fig. 1c**, except splenocytes were positively selected with beads coated with anti-CD49b) and IEL $\gamma\delta$ T cells (positive control), assessed by flow cytometry; for intracellular staining, phycoerythrin-conjugated monoclonal antibody to TCR δ was used to stain IELs extracellularly before intracellular staining with fluorescein isothiocyanate-conjugated anti-TCR δ . Numbers in outlined areas (gates) indicate percent TCR δ^+ cells (outline colors (red and blue) match those in keys above plots), except IEL $\gamma\delta$ cells, which are shown as a percentage of CD45 $^+$ cells (additional information, **Supplementary Fig. 2**). Sp, spleen; Lv, liver. **(d)** 3' end targeted RNA-sequencing data from a publicly available data set (GEO accession code **GSE52043**), showing the locus encoding TCR γ (chr13:19,426,000–19,446,000). **(e)** Frequency of CXCR6 $^+$ cells in each ILC subset from *Cxcr6^{eGFP/+}* reporter mice. Each symbol represents an individual mouse (three or four per genotype per tissue); small horizontal lines indicate the mean (\pm sem). $*P \leq 0.05$ (one-way analysis of variance (ANOVA) with multiple comparisons). **(f)** Frequency of ILCs in *Cxcr6^{eGFP/+}* and *Cxcr6^{eGFP/eGFP}* mice. All comparisons are not significant (two-way ANOVA with multiple comparisons). **(g)** Gene expression in siLP ILC1 cells, siLP NK cells and siEL ILC1 cells; colors indicate transcripts upregulated by one or two subsets (match key above plot). Data are representative of one to three experiments per sample with cells pooled from three to five mice each (**a, b, g**; $n = 2$ –3 replicates each), two independent experiments (**c, d**) or three independent experiments (**e, f**; error bars, mean \pm sem of three or four mice per genotype per tissue).

cell-surface markers contained a large population of *Eomes*⁻ cells (Fig. 5f). Thus, NK cells and ILC1 cells could not be discriminated on the basis of CD127 and/or CD27 in the spleen and siLP. The use of *Eomes*^{eGFP} reporter mice might be useful in future attempts to better discriminate NK cells from ILC1 cells, as has been described^{4,42}.

A core, NK cell–distinct ILC signature

If ILC1 cells are different from NK cells as a class, we reasoned that there should be transcripts common among ILC1 cells from all tissues that differ from those of NK cell subsets, and vice versa (Fig. 6a). Spleen and liver ILC1 cells shared expression of genes previously reported to be expressed by ILC1 cells,^{8,13,15,42,43} such as *Tnfrsf10*, *Tnfrsf11* and *Il2* (Fig. 6a and Supplementary Table 5), although *Il2* was filtered out by our methods because of variability between replicates (Fig. 3c). Transcripts with higher expression in NK cell subsets included *Eomes*, *Itgam* (which encodes CD11b (integrin α_M)) and members of the *Klra* family of genes (which encode receptors of the Ly49 family) (Supplementary Table 5). This NK cell–specific signature was consistent with that identified in a more limited data set comparing NK cells with ILC1 cells and ex-ROR γ ⁺ ILC3 cells from the siLP⁴. Visualization of genes expressed differently in the liver, spleen and siLP for the entire data set revealed that with few exceptions, genes upregulated in ILC1 cells relative to their expression in NK cells had even higher expression in many other ILC subsets (Fig. 6b). Genes upregulated in all NK cells were consistently not expressed in other ILCs (Fig. 6b), except for low transcript levels in ILC1 cells. The transcripts with the highest expression in ILCs relative to their expression in NK cells (Fig. 6a) were *Tcrp-V3*, *Tmem176a*, *Tmem176b*, *Il7r* and *Cxcr6* (Fig. 6b and Supplementary Table 6). The high expression of several *TCRg* transcripts by all ILC subsets compared to NK cells (Fig. 6a,b) was unexpected, as ILCs by definition do not express recombined antigen receptors. We first used flow cytometry to confirm that no T cell antigen receptor (TCR) δ -chains were expressed in any of our cell types, either extracellularly or intracellularly (Fig. 6c and Supplementary Fig. 2a); this suggested that they lacked functional TCR $\gamma\delta$ expression. Furthermore, from a published RNA-sequencing data set that included liver ILC1 cells, liver NK cells and spleen NK cells from mice deficient in recombination-activating gene 1 (*Rag1*^{-/-} mice)⁴³, we found high levels of the 3' end of the locus encoding TCR γ , annotated as the '*Tcrp C4*' transcript, present in liver ILC1 cells but not in NK cells from liver or spleen (Fig. 6d). By PCR, we confirmed that the transcript encoding TCR γ -V3 was a germline transcript (data not shown). As cytokines IL-7 and IL-15, which activate the transcription factor STAT5, are well known to mediate germline expression of the locus encoding TCR γ ^{44,45}, we concluded that ILCs had open chromatin at the locus encoding TCR γ and germline transcription, probably due to signaling through IL-7. We also assessed the ability to use CXCR6 as an ILC marker through the use of a *Cxcr6*^{eGFP} reporter mouse. Although we found significant differences between ILC1 and NK cell populations in their frequency of CXCR6⁺ cells (Fig. 6e and Supplementary Fig. 2b), not all ILCs were labeled (Fig. 6e and Supplementary Fig. 2b). Additionally, at steady state, CXCR6 was not required for the development or tissue homing of ILCs, as we found no difference between *Cxcr6*^{eGFP/+} mice and *Cxcr6*^{eGFP/eGFP} mice in their frequency of ILCs (Fig. 6f).

Finally, we sought to determine whether intestinal intraepithelial ILC1 cells should be classified as an ILC1 or NK cell population, as this population has unique transcription factors and cell-surface markers that prevent it from fitting clearly into an ILC1 or NK cell designation¹⁶. Comparisons of IEL ILC1 cells, siLP ILC1 cells and siLP NK cells revealed that IEL ILC1 cells expressed transcripts characteristic

of both NK cell and ILC1 populations (Fig. 6g). Whereas the NK cell signature transcripts *Eomes* and *Klra3* (Fig. 6b) had higher expression by IEL ILC1 cells than siLP ILC1 cells, the ILC1 signature transcripts *Tcrp-V2* and *Tcrp-V3* (Fig. 6b) were more abundant in IEL ILC1 cells than in siLP NK cells (Fig. 6g). It remains unclear whether IEL ILC1 cells are a single, unique subset with distinct developmental and functional characteristics, or whether siLP NK cells and ILC1 cells both traffic to the epithelium, where they become phenotypically indistinguishable, possibly in response to tissue factors in the epithelium.

DISCUSSION

Here we have provided the first comprehensive transcriptional analysis of the spectrum of ILC subsets reported in the siLP, liver and spleen, to our knowledge. The transcriptional programs we found should allow better definition of the individual ILC classes, as well as of ILC subsets within a class. ILC2 was the most homogeneous and distinguishable ILC class and expressed the greatest number of characteristic genes, with many transcripts expressed more than fourfold higher than in any other subset. These genes encoded well-documented factors as well as previously unrecognized genes, including the nuclear receptors RXR γ and PPAR γ and several other molecules involved in lipid metabolism. Within the ILC3 class, LTi-like ILC3 cells expressed several characteristic genes at higher levels than in other subsets, including genes previously unknown to be expressed in cells of the immune system, including *Cntn1*, *Slc6a7* and *Cacna1g*. These cells were distinct from NKp46⁺ ILC3 cells and were effectively marked by the previously unreported LTi-like factor *Nrp1*, as well as by CD25. The definition of the ILC1 class was the most problematic, because in comparisons of ILC1 cells with all other ILCs and NK cells, we found no markers specific for siLP ILC1 cells and few for liver and spleen ILC1 cells. The transcript encoding CD103 was an unexpected characteristic transcript for siLEL ILC1 cells, given that it was not present as CD103 protein in this subset in mouse studies and has been noted only in human cells¹⁶. Among siLP subsets, we also found a previously unrecognized 'intestinal' signature composed of activation markers such as CD69 and many transcription factors, including those encoded by *Rora*, *Atf3*, *Nr4a1* and *Maff*, probably reflective of continuous environmental exposure.

Beyond the unique factors, ILC classes also shared many transcripts. For example, siLP NKp46⁺ROR γ ⁺ ILC3 cells shared many transcripts with siLP ILC1 cells, including *Ifng* and *Il12rb2*. These data, which were consistent with published reports^{4,14}, provide a basis for the proposal of functional plasticity of siLP NKp46⁺ROR γ ⁺ ILC3 cells, which become similar to ILC1 cells in certain conditions yet to be defined. Furthermore, ILC1 cells and ILC2 cells shared expression of the transcription factor–encoding transcript *Ets-2*, which would suggest previously unknown transcriptional pathways shared by ILC1 cells and ILC2 cells. Notably, ILC2 cells and LTi-like ILC3 cells expressed genes that would suggest a function in neural and glial crosstalk. ILC2 cells expressed *Bmp2*, which encodes a molecule that has been found to modulate enteric motility in response to the microbiota³³; given the importance of motility in clearing helminth infections, this observation might indicate a previously unknown mechanism of innate defense. We discovered that both ILC2 cells and LTi-like ILC3 cells also expressed *Ret*, a proto-oncogene that encodes a receptor for the gliaderived neurotrophic family of molecules, which are known to drive the development of Peyer's patches³⁷ but currently remain unstudied in ILCs. Thus, in the siLP, ILCs may engage in crosstalk with neurons and glia in the steady state and during an immune response.

We generated a core ILC signature that included 17 genes, with the highest expression of germline transcripts encoding TCR γ , as well as *Cxcr6*, *Tmem176a*, *Tmem176b* and *Il7r*. We were surprised to find that

the gene with the highest expression by all ILCs relative to its expression in NK cells was the *TCRγ-V3* germline transcript, which has not been previously reported in ILCs, to our knowledge, but has been reported in a putative ILC3 cell line⁴⁶. We postulate that this might reflect signaling by IL-7R, which is expressed by all ILCs but not by mature NK cells⁴⁴. In our study, we confirmed that all ILCs had higher expression of *Cxcr6* than did NK cells. Two published studies have also investigated CXCR6 expression in ILCs. The first study demonstrated that a CXCR6⁺ early progenitor gives rise to both NK cells and ILCs but not T cells⁴⁷. The second study found that CXCR6 deficiency ‘preferentially’ affects the frequency and function of Nkp46⁺ ILC3 cells by preventing appropriate interactions with CD11b⁺ intestinal dendritic cells⁴⁸. We found that *CXCR6* did not mark the entire population, nor were any ILC frequencies affected by its loss; however, it is possible that loss of CXCR6 affects ILC function, and this should be tested in all ILC classes. Notably, *Tmem176b* has been reported to be a marker of innate lymphocytes as one of only three transcripts shared by NK cells, NKT cells and γδ T cells⁴⁹, although in our data set we found that ILC1, ILC2 and ILC3 cells had significantly higher expression of *Tmem176b* than did NK cells. These data suggest that the core ILC signature identified here may also extend to other innate and/or tissue-resident lymphoid cells.

IFN-γ-producing ILC1 cells and NK cells can develop from different progenitors and are, respectively, independent of and dependent on Eomes^{4,11}. However, in tissues they show overlapping phenotypes and functional programs. NK cells are well known to have cytolytic ability, but liver ILC1 cells express granzyme A and granzyme C and have also been shown to be cytolytic^{42,43,50}. In our study, liver ILC1 cells were clearly separated by TRAIL and CD49a, but cell-surface markers of CD127 with and without CD27 in the spleen and siLP, respectively, were insufficient to discriminate Eomes⁻ ILC1 cells from Eomes⁺ NK cells. Thus, transcriptional data generated using Eomes reporter mice might be useful for comparison to our data set^{4,42}. Moreover, the induction of cytokines such as IL-15 and/or IL-2 in certain pathologic conditions might further increase the phenotypic and functional similarity of ILC1 cells and NK cells. Thus, it remains unclear whether ILC1 cells and NK cells are truly distinct lineages or a spectrum of cells within a single lineage that includes ILC1 cells, immature NK cells and mature NK cells.

Our data offer the most complete transcriptional profile of ILCs and NK cells so far, to our knowledge, and provide a comprehensive view of the relationships among ILC subsets at steady state. Our findings should help define new avenues of research and should aid in the production of new tools for studying ILCs, especially with the identification of several molecular targets with high expression by all ILCs. Finally, they should also be a valuable resource for the scientific community, with access to our data set and comparisons to other published data sets generated under the same rigorous conditions, provided by the ImmGen Project.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. GEO: microarray data, [GSE37448](#).

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

ACKNOWLEDGMENTS

We thank our colleagues in the ImmGen consortium, especially C. Benoist and L. Lanier, for input and discussion; the core ImmGen team, K. Rothamel and

A. Rhodes, for contributions and technical assistance; M. Artyomov, G. Krishnan and J. Siegel for computational assistance; D. Sojka for discussion; E. Lantelme and D. Brinja for sorting assistance; P. Wang for microscopy assistance; and eBioscience and Affymetrix for support of the ImmGen Project. Supported by the US National Institutes of Health (R24AI072073 to the ImmGen Consortium; 1U01AI095542, R01DE021255 and R21CA16719 to the Colonna laboratory; MSTP T32 GM07200 to M.L.R.; and Infectious Disease Training Grant T32 AI 7172-34 to V.S.C.).

AUTHOR CONTRIBUTIONS

M.L.R. analyzed data; A.F., M.L.R., J.S.L. and Y.W. sorted cell subsets; M.L.R., A.F. and V.S.C. performed follow-up experiments and analyzed data; S.G. maintained mice; S.K.D. provided critical reagents; M.L.R., A.F., S.G. and M.C. designed studies; M.L.R. and M.C. wrote the paper; and the ImmGen Consortium contributed to the experimental design and data collection.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

- Heng, T.S. & Painter, M.W. The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).
- Diefenbach, A., Colonna, M. & Koyasu, S. Development, differentiation, and diversity of innate lymphoid cells. *Immunity* **41**, 354–365 (2014).
- McKenzie, A.N., Spits, H. & Eberl, G. Innate lymphoid cells in inflammation and immunity. *Immunity* **41**, 366–374 (2014).
- Klose, C.S. *et al.* Differentiation of type 1 ILCs from a common progenitor to all helper-like innate lymphoid cell lineages. *Cell* **157**, 340–356 (2014).
- Sawa, S. *et al.* Lineage relationship analysis of RORγt⁺ innate lymphoid cells. *Science* **330**, 665–669 (2010).
- Lee, J.S. *et al.* AHR drives the development of gut ILC22 cells and postnatal lymphoid tissues via pathways dependent on and independent of Notch. *Nat. Immunol.* **13**, 144–151 (2011).
- Klose, C.S. *et al.* A T-bet gradient controls the fate and function of CCR6-Rorγt⁺ innate lymphoid cells. *Nature* **494**, 262–265 (2013).
- Sciumè, G. *et al.* Distinct requirements for T-bet in gut innate lymphoid cells. *J. Exp. Med.* **209**, 2331–2338 (2012).
- Rankin, L.C. *et al.* The transcription factor T-bet is essential for the development of Nkp46⁺ innate lymphocytes via the Notch pathway. *Nat. Immunol.* **14**, 389–395 (2013).
- Buonocore, S. *et al.* Innate lymphoid cells drive interleukin-23-dependent innate intestinal pathology. *Nature* **464**, 1371–1375 (2010).
- Constantinides, M.G., McDonald, B.D., Verhoef, P.A. & Bendelac, A. A committed precursor to innate lymphoid cells. *Nature* **508**, 397–401 (2014).
- Yu, J., Freud, A.G. & Caligiuri, M.A. Location and cellular stages of natural killer cell development. *Trends Immunol.* **34**, 573–582 (2013).
- Gordon, S.M. *et al.* The transcription factors T-bet and Eomes control key checkpoints of natural killer cell maturation. *Immunity* **36**, 55–67 (2012).
- Reynders, A. *et al.* Identity, regulation and *in vivo* function of gut Nkp46⁺RORγt⁺ and Nkp46⁺RORγt⁻ lymphoid cells. *EMBO J.* **30**, 2934–2947 (2011).
- Gasteiger, G., Hemmers, S., Bos, B.D., Sun, J.C. & Rudensky, A.Y. IL-2-dependent adaptive control of NK cell homeostasis. *J. Exp. Med.* **210**, 1179–1187 (2013).
- Fuchs, A. *et al.* Intraepithelial type 1 innate lymphoid cells are a unique subset of IL-12- and IL-15-responsive IFN-γ-producing cells. *Immunity* **38**, 769–781 (2013).
- Turner, J.E. *et al.* IL-9-mediated survival of type 2 innate lymphoid cells promotes damage control in helminth-induced lung inflammation. *J. Exp. Med.* **210**, 2951–2965 (2013).
- Mjösberg, J. *et al.* The transcription factor GATA3 is essential for the function of human type 2 innate lymphoid cells. *Immunity* **37**, 649–659 (2012).
- Yagi, R. *et al.* The transcription factor GATA3 is critical for the development of all IL-7Rα-expressing innate lymphoid cells. *Immunity* **40**, 378–388 (2014).
- Evans, R.M. & Mangelsdorf, D.J. Nuclear receptors, RXR, and the Big Bang. *Cell* **157**, 255–266 (2014).
- Spencer, S.P. *et al.* Adaptation of innate lymphoid cells to a micronutrient deficiency promotes type 2 barrier immunity. *Science* **343**, 432–437 (2014).
- Molofsky, A.B. *et al.* Innate lymphoid type 2 cells sustain visceral adipose tissue eosinophils and alternatively activated macrophages. *J. Exp. Med.* **210**, 535–549 (2013).
- Mohebiany, A.N., Harroch, S. & Bouyain, S. in *Cell Adhesion Molecules: Implications in Neurological Diseases, Advances in Neurobiology* Vol. 8 (eds. Berezin, V. & Walmod, P.S.) Chapter 8 (Springer Science+Business Media, 2014).
- Cella, M. *et al.* A human natural killer cell subset provides an innate source of IL-22 for mucosal immunity. *Nature* **457**, 722–725 (2009).
- Mortha, A. *et al.* Microbiota-dependent crosstalk between macrophages and ILC3 promotes intestinal tolerance. *Science* **343**, 1249–1253 (2014).
- Levitt, L.J. *et al.* Production of granulocyte/macrophage-colony stimulating factor by human natural killer cells. Modulation by the p75 subunit of the interleukin 2 and by the CD2 receptor. *J. Clin. Invest.* **88**, 67–75 (1991).

© 2015 Nature America, Inc. All rights reserved.



27. van de Pavert, S.A. *et al.* Maternal retinoids control type 3 innate lymphoid cells and set the offspring immunity. *Nature* **508**, 123–127 (2014).
28. Kiss, E.A. *et al.* Natural aryl hydrocarbon receptor ligands control organogenesis of intestinal lymphoid follicles. *Science* **334**, 1561–1565 (2011).
29. Gascoyne, D.M. *et al.* The basic leucine zipper transcription factor E4BP4 is essential for natural killer cell development. *Nat. Immunol.* **10**, 1118–1124 (2009).
30. Kamizono, S. *et al.* Nfil3/E4bp4 is required for the development and maturation of NK cells *in vivo*. *J. Exp. Med.* **206**, 2977–2986 (2009).
31. Seillet, C. *et al.* Nfil3 is required for the development of all innate lymphoid cell subsets. *J. Exp. Med.* **211**, 1733–1740 (2014).
32. Geiger, T.L. *et al.* Nfil3 is crucial for development of innate lymphoid cells and host protection against intestinal pathogens. *J. Exp. Med.* **211**, 1723–1731 (2014).
33. Muller, P.A. *et al.* Crosstalk between muscularis macrophages and enteric neurons regulates gastrointestinal motility. *Cell* **158**, 300–313 (2014).
34. Bando, J.K., Liang, H.E. & Locksley, R.M. Identification and distribution of developing innate lymphoid cells in the fetal mouse intestine. *Nat. Immunol.* <http://www.nature.com/nri/journal/vaop/ncurrent/full/nri.3057.html> (2014).
35. Veiga-Fernandes, H. *et al.* Tyrosine receptor RET is a key regulator of Peyer's patch organogenesis. *Nature* **446**, 547–551 (2007).
36. Ramirez, K. *et al.* Gene deregulation and chronic activation in natural killer cells deficient in the transcription factor ETS1. *Immunity* **36**, 921–932 (2012).
37. Carpenter, S., Ricci, E.P., Mercier, B.C., Moore, M.J. & Fitzgerald, K.A. Post-transcriptional regulation of gene expression in innate immunity. *Nat. Rev. Immunol.* **14**, 361–376 (2014).
38. Kumanogoh, A. & Kikutani, H. Immunological functions of the neuropillins and plexins as receptors for semaphorins. *Nat. Rev. Immunol.* **13**, 802–814 (2013).
39. Yadav, M. *et al.* Neuropilin-1 distinguishes natural and inducible regulatory T cells among regulatory T cell subsets *in vivo*. *J. Exp. Med.* **209**, 1713–1722 (2012).
40. Weiss, J.M. *et al.* Neuropilin 1 is expressed on thymus-derived natural regulatory T cells, but not mucosa-generated induced Foxp3+ T reg cells. *J. Exp. Med.* **209**, 1723–1742 (2012).
41. Shen, W. *et al.* Adaptive immunity to murine skin commensals. *Proc. Natl. Acad. Sci. USA* **111**, E2977–E2986 (2014).
42. Daussey, C. *et al.* T-bet and Eomes instruct the development of two distinct natural killer cell lineages in the liver and in the bone marrow. *J. Exp. Med.* **211**, 563–577 (2014).
43. Sojka, D.K. *et al.* Tissue-resident natural killer (NK) cells are cell lineages distinct from thymic and conventional splenic NK cells. *Elife* **3**, e01659 (2014).
44. Ye, S.K. *et al.* Induction of germline transcription in the Tcr γ locus by Stat5: implications for accessibility control by the IL-7 receptor. *Immunity* **11**, 213–223 (1999).
45. Zhao, H., Nguyen, H. & Kang, J. Interleukin 15 controls the generation of restricted T cell receptor repertoire of intraepithelial lymphocytes. *Nat. Immunol.* **6**, 1263–1271 (2005).
46. Allan, D.S. *et al.* An *in vitro* model of innate lymphoid cell function and differentiation. *Mucosal Immunol.* <http://www.nature.com/mi/journal/vaop/ncurrent/full/mi201471a.html> (2014).
47. Yu, X. *et al.* The basic leucine zipper transcription factor NFIL3 directs the development of a common innate lymphoid cell progenitor. *Elife* **10**, e04406 (2014).
48. Satoh-Takayama, N. *et al.* The chemokine receptor CXCR6 controls the functional topography of interleukin-22 producing intestinal innate lymphoid cells. *Immunity* **41**, 776–788 (2014).
49. Bezman, N.A. *et al.* ImmGen report: molecular definition of natural killer cell identity and activation. *Nat. Immunol.* **13**, 1000–1009 (2012).
50. Peng, H. *et al.* Liver-resident NK cells confer adaptive immunity in skin-contact inflammation. *J. Clin. Invest.* **123**, 1444–1456 (2013).

The complete list of authors is as follows:

Laura Shaw⁶, Bingfei Yu⁶, Ananda Goldrath⁶, Sara Mostafavi⁷, Aviv Regev⁸, Edy Y Kim⁹, Dan F Dwyer⁹, Michael B Brenner⁹, K Frank Austen⁹, Andrew Rhoads¹⁰, Devapregasan Moodley¹⁰, Hideyuki Yoshida¹⁰, Diane Mathis¹⁰, Christophe Benoist¹⁰, Tsukasa Nabekura¹¹, Viola Lam¹¹, Lewis L Lanier¹¹, Brian Brown¹², Miriam Merad¹², Viviana Cremasco¹³, Shannon Turley¹³, Paul Monach¹⁴, Michael L Dustin¹⁵, Yuesheng Li¹⁶, Susan A Shinton¹⁶, Richard R Hardy¹⁶, Tal Shay¹⁷, Yilin Qi¹⁸, Katelyn Sylvia¹⁸, Joonsoo Kang¹⁸, Keke Fairfax¹, Gwendalyn J Randolph¹, Michelle L Robinette¹, Anja Fuchs² & Marco Colonna¹

⁶Division of Biological Sciences, University of California San Diego, La Jolla, California, USA. ⁷Computer Science Department, Stanford University, Stanford, California, USA. ⁸Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. ⁹Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Boston, Massachusetts, USA. ¹⁰Division of Immunology, Department of Microbiology & Immunobiology, Harvard Medical School, Boston, Massachusetts, USA. ¹¹Department of Microbiology & Immunology, University of California San Francisco, San Francisco, California, USA. ¹²Icahn Medical Institute, Mount Sinai Hospital, New York, New York, USA. ¹³Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts, USA and Department of Cancer Immunology, Genentech, San Francisco, California, USA. ¹⁴Department of Medicine, Boston University, Boston, Massachusetts, USA. ¹⁵Skirball Institute of Biomolecular Medicine, New York University School of Medicine, New York, New York, USA. ¹⁶Fox Chase Cancer Center, Philadelphia, Pennsylvania, USA. ¹⁷Department of Life Sciences, Ben-Gurion University of the Negev, Be'er Sheva, Israel. ¹⁸Department of Pathology, University of Massachusetts Medical School, Worcester, Massachusetts, USA.

ONLINE METHODS

Mice. Consistent with ImmGen Project standards, 6- to 8-week-old male C57BL/6J mice and male B6.129P2(Cg)-*Rorc*^{tm2Litt}/J mice (with the sequence encoding enhanced green fluorescent protein (EGFP) inserted into the locus encoding ROR γ t) obtained from Jackson Laboratories and maintained at the Washington University School of Medicine (WUSM) under specific pathogen-free conditions were used for cell-sorting and validation experiments. B6.129P2-*Cxcr6*^{tm1Litt}/J mice with the sequence encoding EGFP inserted into the locus encoding CXCR6, obtained from Jackson Laboratories, were used for validation experiments in which mice were littermates or age-matched, gender-matched and co-housed when possible. Il-22 tdTomato reporter BAC transgenic mice were used for validation experiments⁴¹. The WUSM Animal Studies Committee approved all experiments.

Antibodies and flow cytometry. Anti-CD3e (145-2C11), anti-CD19 (eBio1D3), anti-CD27 (LG.7F9), anti-CD49b (HMa2, DX5), anti- $\gamma\delta$ (UC7-13D5; eBioGL3), anti-CD127 (A7R34; eBioSB/199), anti-CD49a (HMa1), anti-Nkp46 (29A1.4), anti-CD304 (3DS304M), anti-NK1.1 (PK136), anti-Sca-1 (D7), anti-ST2 (RMST2-2), anti-TRAIL (eBioN2B2), anti-CD4 (GK1.5), anti-KLRG1 (2F1), anti-ROR γ t (AFKJS-9), anti-Eomes (Dan11mag), SA ν -PE/Cy7, and isotype-matched control monoclonal antibodies were obtained from eBioscience. Anti-CD25 (PCB1) and 7-AAD were from BD Biosciences. Anti-CD45 (30-F11) was from Miltenyi Biotec. LIVE/DEAD Fixable Aqua and SA ν -APC were from Life Technologies. Monoclonal anti-Nkp46 (Cs96; rat immunoglobulin G) was produced and biotinylated in the Colonna lab at WUSM. Fc receptors were blocked before surface staining with supernatant from hybridoma cells producing monoclonal antibody to CD32 (HB-197; ATCC). For intracellular staining, the FXP3 staining kit (eBioscience) was used. Data were acquired on a BD FACSCanto II and analyzed with FlowJo software (Treestar).

Cell identification, isolation and microscopy. All cells were stained and sorted according to the published standard operations protocol on the ImmGen website (<http://www.immgen.org/Protocols/ImmGen%20Cell%20prep%20and%20sorting%20SOP.pdf>). Cells were isolated from 3 to 15 mice per sample. For isolation of siLP and siEL, small intestine beginning half a centimeter after the pylorus and ending half a centimeter before the cecum with Peyer's patches removed was dissociated using the Miltenyi Lamina Propria Dissociation kit, which includes two dithiothreitol/EDTA washes and an enzymatic digestion step of 30 min. For liver and spleen isolation, mice were perfused with 20 mL of cold PBS, and tissue was dissociated through a 70 μ m cell strainer. Lymphocytes were enriched at the interface between a gradient of 40% and 70% Percoll in HBSS. siELs were further pre-enriched by negative selection with anti-CD8 MicroBeads (Miltenyi). Liver and spleen were pre-enriched by negative selection with anti-CD19 and anti-CD4 MicroBeads (Miltenyi). Cells were double-sorted directly into TRIzol with a Becton-Dickinson FACSARIA II. The data browser of the ImmGen Project website shows flow cytometry plots from gating strategies and purity from the first round of sorting. Cytospins were prepared according to the same ImmGen standards, except from only one to three mice; splenic samples were positively pre-enriched with DX5-coated beads (Miltenyi). Sorted cells were spun onto slides, left to dry overnight, and

then fixed and stained with Diff-Quik. Pictures were taken at $\times 60$ with oil immersion using a Nikon Eclipse E800 and Leica Application Suite software.

Microarray and data analysis. Two to three replicates of RNA were obtained from each sample that passed quality control. RNA amplification and hybridization to the Affymetrix Mouse Gene 1.0 ST array were carried out by ImmGen with a standardized TRIzol extraction protocol (<https://www.immgen.org/Protocols/Total%20RNA%20Extraction%20with%20Trizol.pdf>). Data generation and quality-control documentation also followed the ImmGen protocol; these methods can be found online (https://www.immgen.org/Protocols/ImmGen%20QC%20Documentation_ALL-DataGeneration_0612.pdf), along with quality-control data, replicate information, and batch information from each sample. Data were analyzed with GenePattern software⁵¹ (Broad Institute). Raw data were normalized with RMA. Differences in gene expression were identified with the Multiplot Studio function of GenePattern, from a filtered subset of genes with coefficients of variation less than 0.1 in all samples and expression of at least 120 relative units in one subset by the class mean functions, a value that corresponds to 95% confidence of true expression across the ImmGen data set. For gene-expression signatures of individual subsets, pairwise comparisons were made between subsets, filtering for a change in expression of twofold or fourfold. For comparisons of two to four samples, probe sets were considered to have differences of expression for transcripts that expressed >120 relative units, with a change in expression of greater than twofold and *P* values of <0.05 (Student's *t*-test), except between siLP ILC1 and siLP NK cells, where the *P* value was not considered. Volcano plots and plots comparing change in expression (fold) versus change in expression (fold) were produced in Multiplot, and the degree of overlapping genes between subsets was calculated in MATLAB. Heat maps were generated with Gene-E (<http://www.broadinstitute.org/cancer/software/GENE-E/>). Data were log₂-transformed and visualized by 'relative' expression per row or 'global' expression in the heat map, as indicated, and rows were clustered with the Hierarchical Clustering function with the Pearson correlation as a metric. Where more than one probe set with the same gene annotation was found, the probe set with highest average expression was used. For PCA, the top 10% of the most variable probe sets was calculated with the PopulationDistances PCA program (S. Davis, Harvard Medical School). This program identifies differently expressed genes through ANOVA using the geometric standard deviation of populations to weight genes that vary in multiple populations. The data set for the top 10% of genes with the most variability was log₂-transformed in MATLAB and used to generate a PCA with the functions *pca* and *scatter3*. Hierarchical clustering of sample replicates was carried out in Gene-E from the same data set, with the Pearson correlation used as a metric. For RNA sequencing analysis, data from GEO accession code [GSE52043](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52043) were visualized in Integrative Genomics Viewer and aligned to the mm9 National Center for Biotechnology Information assembly of the mouse genome, as described⁴³.

Statistics. Prism (GraphPad Software) was used for statistical analysis of flow cytometry data. Data were tested using either one-way ANOVA or two-way ANOVA, as indicated.

51. Reich, M. *et al.* GenePattern 2.0. *Nat. Genet.* **38**, 500–501 (2006).