Distinct developmental pathways generate functionally distinct populations of natural killer cells

Yi Ding, Marieke Lavaert, Simon Grassmann, Victor Band, Liang Chi, Arundhoti Das, Sumit Das, Christelle Harly, Susannah Shissler, Justin Malin, et al.

Different developmental pathways generate functionally distinct populations of natural killer cells

Yi Ding[1], Marieke Lavaert[1], Simon Grassmann[2], Victor Band[3], Liang Chi[3], Arundhoti Das[1], Sumit Das[1], Christelle Harly[4,5], Susannah C. Shissler[1], Justin Malin[1], Dingkang Peng[6],Yongge Zhao[1], Jinfang Zhu[6], Yasmine Belkaid[7], Joseph C. Sun[2], Avinash Bhandoola[1,*]

[1]T Cell Biology and Development Unit, Laboratory of Genome Integrity, Center for Cancer Research, National Cancer Institute, National Institute of Health, Bethesda, MD, USA

[2]Immunology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[3]Metaorganism Immunity Section, Laboratory of Host Immunity and Microbiome, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

[4]Nantes Université, Inserm UMR 1307, CNRS UMR 6075, Université d'Angers, CRCI2NA, Nantes, France

[5]LabEx IGO "Immunotherapy, Graft, Oncology", Nantes, France

[6]Molecular and Cellular Immunoregulation Section, Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

[7]Metaorganism Immunity Section, Laboratory of Host Immunity and Microbiome, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA; NIAID Microbiome Program, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA.

*Email: avinash.bhandoola@nih.gov

Natural killer (NK) cells function by eliminating virus-infected cells or tumor cells. Here, we identified a NK lineage-biased progenitor population, termed early NK progenitors (ENKP), which developed into NK cells independently of common precursors for ILCs (ILCPs). ENKP-derived NK cells (ENKP_ NK cells) and ILCP-derived NK cells (ILCP_ NK cells) were also transcriptionally different. We devised combinations of surface markers that identified highly enriched ENKP_NK and ILCP_NK cell populations in wild-

type mice. Furthermore, Ly49H$^+$ NK cells that responded to mouse cytomegalovirus (MCMV) infection primarily developed from ENKPs whereas ILCP_NK cells were better IFN-γ producers upon Salmonella and Herpes Simplex Virus (HSV) infections. Interestingly, human CD56^{dim} and CD56^{bright} NK cells were transcriptionally similar to ENKP_NK cells and ILCP_NK cells, respectively. Our findings establish the existence of two pathways of NK cell development that generate functionally distinct NK cell subsets in mice, and further suggest these pathways may be conserved in humans.

Innate lymphoid cells (ILCs) are transcriptionally and functionally like T cells but lack rearranged antigen receptors. ILCs are classified into helper ILCs which secrete similar cytokines as T helper cells and killer ILCs or NK cells which play critical roles in controlling viral infection and eradicating malignant cells by secreting perforin, granzymes and cytokines such as IFN-γ[1]. ILCs develop from ALPs (all lymphoid precursors). Downstream of ALPs, the earliest defined ILC-specified progenitor population is early innate lymphoid progenitors (EILPs), which can differentiate into both helper ILCs and NK cells and gives rise to common precursors for ILCs (ILCPs), a progenitor population for helper ILCs that still retains NK lineage potential[2-5]. Because transcription factor PLZF is highly expressed in ILCPs but not expressed in upstream progenitors or mature ILCs, it has been used for lineage tracing of ILCP-derived cells[3]. Previous work established that NK cells are labeled at a lower ratio than helper ILCs in PLZF-Cre, Rosa26-reporter (PLZF lineage-tracing) mice, which suggested an ILCP-independent pathway for NK development[3]. Previously defined refined NK progenitors (rNKPs) are also traced by PLZF at intermediate levels, suggesting they are a heterogeneous population and some of them may develop from ILCPs[6,7]. Thus, developmental pathways between ALPs and mature NK cells are poorly defined.

The functional heterogeneity of NK cells has long been studied in both mouse and human[8-12]. Around 50% of mature NK cells in C57BL/6 mice express NK activating receptor Ly49H that recognizes the mouse cytomegalovirus (MCMV) antigen m157 and these Ly49H$^+$ NK cells preferentially respond to MCMV infection[13,14]. However, it is not known whether functionally distinct NK subsets develop via different pathways. Here, we used single-cell RNA-sequencing (scRNA-seq) to assess potential NK progenitor populations within bone marrow (BM). We

identified an *Flt3*$^{+}$*Il2rb* (encoding CD122)$^{+}$ cluster that, by trajectory analysis, represented an ILCP-independent pathway of NK cell development. Consistently, we found that these Flt3^{+}CD122^{+} progenitors, here referred to as early NK progenitors (ENKPs), were not traced by PLZF, and were biased to NK cell fates with limited potential for differentiation into helper ILCs *in vitro* and *in vivo*. In addition, we showed ENKPs but not ILCPs efficiently gave rise to Ly49H^{+} NK cells that were largely untraced by PLZF, and ENKP-derived NK cells expanded dramatically upon MCMV infection. We also identified NK populations that are enriched for ENKP-derived or ILCP-derived NK cells in wild-type B6 mice by surface markers. Further, ILCP_NK cells were more robust IFN-γ producers during Salmonella infection and Herpes Simplex Virus (HSV) infection. Our data reveal the existence of parallel developmental pathways that give rise to functionally distinct NK subsets.

Characterization of early NK cell development in bone marrow

To identify candidate precursors of NK cells, we performed single-cell RNA sequencing (scRNA-seq) on cell populations (Lin^{-}CD27^{+}2B4^{+}CD122^{+}) sorted from bone marrow that we reasoned were likely to contain NK lineage progenitors, because CD27 and 2B4 are expressed on ALP, EILP, as well as ILCP populations[15,16]. We also sequenced previously described progenitors including a population containing ALP (Lin^{-}kit^{+}Tcf7^{YFP-}α4β7^{-}Flt3^{+}IL-7Rα^{+}), EILP (Lin^{-}kit^{+}Tcf7^{YFP+}α4β7^{+}Thy1.2^{-}), ILCP (Lin^{-}kit^{+}Tcf7^{YFP+}α4β7^{+}Flt3^{-}Thy1.2^{+}), ILC2P (Lin^{-}CD25^{+}Sca1^{+}) ; and separately, group 1 ILCs (Lin^{T-}NK1.1^{+}CD122^{+}) from bone marrow as potential upstream and downstream populations of NK progenitors (Fig. 1a, and Extended Data Fig. 1a,b)[15,16]. ALP, EILP, ILCP, ILC2P, NK and ILC1 clusters were defined based on the expression pattern of surface markers and transcription factors (Fig. 1b). The UMAP revealed that the ILCP cluster, which expressed high level of *Zbtb16* (encoding PLZF), overlapped with some of the cells sorted as candidate NK progenitors (Figure 1a,b and Extended Data Fig. 1b, c), suggesting overlap between rNKPs (defined as Lin^{-}2B4^{+}CD27^{+}Flt3^{-}CD122^{+} by flow cytometry[6]) and ILCPs. Consistently, some rNKPs express PLZF[7], indicating rNKPs overlap with ILCPs. A cluster was identified as *Il2rb* (encoding CD122)$^{+}$*Flt3*$^{-}$*Zbtb16*^{lo}, which likely is another subset of rNKPs, thus we named this cluster as rNKP' (Fig. 1a and 1b).

Interestingly, a cluster within Lin$^-$CD27$^+$2B4$^+$CD122$^+$ NK progenitors formed close to EILP and was marked by high expression of both *Il2rb* and *Flt3* (Fig. 1a, b and Extended data Fig. 1b). *Flt3* is expressed by upstream progenitors such as ALP and EILP during ILC development but downregulated in downstream progenitors such as ILCP[15], suggesting this *Il2rb*$^+$*Flt3*$^+$ NK progenitor cluster could be an early-stage developmental intermediate during NK cell development. Accordingly, we tentatively named this cluster "Early NK Progenitor" (or ENKP). Another recently defined NK/ILC1 progenitor population termed aceNKPs expresses NKG2A/C/E[17]; however the ENKP cluster did not express *Klrc1*, *Klrc2* and *Klrc3* (encoding NKG2A, NKG2C and NKG2E, respectively) (Extended Data Fig. 1d), suggesting ENKPs represent a novel population. We predicted the developmental relationships among clusters for the NK lineage using pseudotime reconstruction. Interestingly, a pathway including the ENKP but not ILCP cluster showed a sequence of predicted development as "ALP -> EILP -> ENKP -> rNKP' -> NK" (lineage 1), which suggested a potential ILCP-independent pathway of NK development. The known lineage for ILC2 development[15] was also detected as "ALP -> EILP -> ILCP -> ILC2P" (lineage 2) (Extended Data Fig. 1e). However, we did not detect an ILCP-dependent pathway of NK cell or ILC1 development, suggesting intermediate populations may be missing in our analysis, consistent with evidence indicating ILCPs leave the bone marrow to circulate and differentiate at extramedullary sites[18,19]. Surface markers and transcription factors critical for NK cell and helper ILC development were displayed as a function of pseudotime (Extended Data Fig. 1e). In lineage 1, *Flt3* was highly expressed in ALP and EILP clusters and still maintained in the ENKP cluster but downregulated in rNKP', whereas *Il2rb* was upregulated starting from ENKP cluster. Cells in the ENKP cluster expressed high levels of *Nfil3*, *Tcf7* and *Id2,* but low levels of *Eomes* and *Gata3* (Fig. 1b and Extended Data Fig. 1c). Also, *Zbtb16* was expressed at low levels in lineage 1. These data suggested the existence of an ILCP-independent pathway of NK cell development in bone marrow of adult mice, and identified ENKPs as an interesting candidate population that might represent an early precursor for NK cells.

Consistent with the deduced phenotype of the ENKP cluster identified by scRNA-seq, we identified a Lin$^-$NK1.1$^-$CD5$^-$CD3ε$^-$CD27$^+$2B4$^+$Flt3$^+$CD122$^+$ population in bone marrow by flow cytometry; thus, we named this population identified by flow "ENKP" as well. The frequency of ENKPs was approximately 0.005% (1 in 20,000 cells, similar to EILP[2]) among total live cells in

bone marrow (Fig. 1c). This population is different from rNKP, which was identified as Lin^- $CD27^+2B4^+Flt3^-CD122^+$ cells[6]; thus ENKP is a previously undefined population. Next, we assessed transcription factor requirements for ENKP development. ENKPs were greatly reduced in *Nfil3* deficient mice but remained intact in *Eomes*, *Tcf7* and *Gata3* deficient mice (Fig. 1d,e and Extended Data Fig. 2a,b). NK cell numbers were significantly reduced in *Nfil3*, *Eomes* and *Tcf7* deficient mice but remained intact in *Gata3* deficient mice, as expected (Fig. 1d,e and Extended Data Fig. 2a,b)[2,20-22]. Thus, among the transcription factors we assessed, only *Nfil3* is required for ENKP development. Because ILCPs fail to develop in mice lacking *Tcf7*[16] or *Gata3*[15,23], these results also indicate that ENKPs are not downstream of ILCPs. We further characterized ENKPs by surface markers compared with ALPs, EILPs and ILCPs. ENKPs expressed high levels of α4β7 and Kit but low levels of IL-7Rα; they did not express Thy1.2 as ILCP did, consistent with their representing an early progenitor population (Extended Data Fig. 2c). Approximately 70% of ENKPs had historical expression of *Il7r* (Extended Data Fig. 2d), which was similar to the ratios in ALPs and EILPs[15], indicating ENKPs develop from *Il7r*-expressing lymphoid progenitors, as predicted for NK cells[24].

ENKPs are NK lineage-biased progenitors and developmentally independent of ILCP

To assess the lineage potential of ENKPs *in vitro*, we seeded ENKPs or ILCPs on OP9 stromal cells with the cytokines stem cell factor (SCF), Flt3L, IL-7 and IL-2. Both ILCPs and ENKPs gave rise to $NK1.1^+$ cells, which include both NK cells and ILC1s. However, unlike ILCPs, which efficiently gave rise to ILC2/ILC3s ($ICOS^+$), ENKPs had very limited potential for ILC2/ILC3s (Fig. 2a), indicating they are an NK-biased progenitor population.

We next sought to determine the lineage potential of ENKPs at the clonal level *in vitro*. To distinguish NK from ILC1s *in vitro*, we used $Eomes^{GFP}$ reporter mice to identify NK cells, and CD200R to identify ILC1s as CD200R is highly expressed in ILC1s *in vivo*[25]. In pilot experiments, we cultured mature $Eomes^{GFP-}$ ILC1s and $Eomes^{GFP+}$ NK cells on OP9 culture for 7 days; ILC1s remained $CD200R^{hi}$ and $Eomes^-$ whereas NK cells were nearly all $Eomes^+$ and did not express CD200R (Extended Data Fig. 3a). Thus, Eomes and CD200R can be used to distinguish NK cells from ILC1s *in vitro*. Among the $ICOS^+$ population, we used α4β7 to further distinguish ILC3s and ILC2s (Fig. 2b). The majority of ENKP-derived clones (74%) contained

NK lineage cells, and 54% of ENKP clones were NK lineage-committed; whereas most ILCP-derived clones (92%) contained only helper ILC lineages (Fig. 2b and Extended Data Fig. 3b). Thus, compared with ILCPs, ENKPs are NK lineage-biased progenitors *in vitro*. Compared to ENKPs and ILCPs, mature NK cells generated negligible numbers of NK cells after *in vitro* culture (Extended Data Fig. 3c), indicating that the observed results cannot be attributed to potential NK cell contamination during sorting. Since ENKPs but not ILCPs express Flt3, we differentiated ENKPs *in vitro* with and without Flt3 ligand (Flt3L) to assess whether their ability to generate NK cells depends on Flt3. ENKPs cultured without Flt3L generated fewer NK cells than ENKPs cultured with Flt3L, although the results did not reach statistical significance (Extended Data Fig. 3d). The data indicate that the Flt3 signal may have an effect on ENKP expansion, but is not strictly required for the development of ENKP-derived NK cells.

To characterize the ILC lineage potential of ENKPs *in vivo*, we transferred ENKPs or ILCPs into NOD SCID gamma (NSG) immune-deficient recipient mice and assessed frequency of donor-derived NK cells and ILC1s within group 1 ILCs in liver. ILCPs gave rise to both NK cells and ILC1s, whereas ENKPs generated primarily NK cells but very few ILC1s (Fig. 2c,d). To compare the efficiency of ENKPs and ILCPs to generate NK cells *in vivo*, we co-transferred the same number of ALPs as competitors with ENKPs or ILCPs and found that ENKPs generated NK cells significantly more efficiently than ILCPs in both spleen and liver (Fig. 2e and Extended Data Fig. 3e). As expected, ALPs but not ILCPs or ENKPs gave rise to B cells (Fig. 2e; see Extended Data Fig. 3f-h for examples of sort purity). Thus, ENKPs are NK lineage-biased progenitors which can give rise to NK cells efficiently. Further, the development of both ENKP-derived NK cells (ENKP_NK cells) and ILCP-derived NK cells (ILCP_NK cells) but not ILC1s depended on Eomes, indicating ILCP_NK cells are distinct from ILC1s (Extended Data Fig. 3i).

Next, we used PLZF lineage tracing mice to assess the developmental relationship between ILCP and ENKP. As PLZF is also expressed before specification of hematopoietic lineages[3], we eliminated the background labeling of PLZF and generated long-term bone marrow chimeras, by transferring unlabeled YFP$^-$ LK (Lin$^-$kit$^+$) cells which are enriched for hematopoietic stem cells and downstream progenitors from Zbtb16^{creGFP}; Rosa26-stop-YFP mice into lethally irradiated wild-type recipient mice, as previously described[3]. NK cells and ILC1s in liver were labeled at

similar ratios to published studies[3,7] (Fig 3a). However, ENKPs were not labeled by PLZF, indicating they did not develop from ILCPs (Fig 3a). Next, to compare the labeling of ENKP_NK cells and ILCP_NK cells in spleen by PLZF, we transferred YFP$^-$ ENKPs, ILCPs or ALPs from Zbtb16^{creGFP}; Rosa26-stop-YFP mice into NSG recipient mice. Unlike ILCP_ NK cells, which were highly labeled (~70%) by PLZF, ENKP_ NK cells were labeled at a very low ratio (~10%). ALP-derived NK cells were labeled at a ratio (~25%) similar with total NK cells in long-term chimeras (Fig 3a,b). Thus, ENKP_ NK cells and ILCP_ NK cells have different developmental pathways, and it is likely that NK cells are a mixture of ENKP-derived and ILCP-derived NK cells. In summary, these results showed ENKPs to be developmentally independent of ILCPs, and further indicate that there are at least two pathways for NK cell development.

ENKP-derived and ILCP-derived NK cells are transcriptionally distinct

To investigate whether ENKP_NK cells and ILCP_ NK cells are transcriptionally different, we adoptively transferred ENKPs, ILCPs, or ALPs as a control, into NSG recipient mice. We subsequently sorted progenitor-derived NK cells from spleen and performed multiplexed scRNA-seq for NK cells derived from ENKP (ENKP_NK cells), ILCP (ILCP_NK cells) and ALP (ALP_NK cells) (Fig. 4a). ENKP_NK cells and ILCP_NK cells expressed comparable levels of NK genes including *Eomes*, *Tbx21* (encoding T-bet) and *Klrk1* (encoding NKG2D) (Fig. 4c and Extended Data 4b). Genes expressed at higher levels in ENKP_NK cells included Ly49 receptors such as *Klra8* and *Klra4*, cytotoxic genes *Gzma*, *Gzmb*, and *Prf1* and transcription factors *Zeb2*, *Irf8*, and *Klf2*. In contrast, ILCP_NK cells expressed higher levels of recognition molecule *Cd226*, chemokine receptors including *Ccr2* and *Cxcr3*, transcription factor *Tcf7*, the cytokine gene *Ifng* and also *Cd160*, which is critical for IFN-γ production[26] (Fig. 4b, c and Extended Data Fig. 4a, b and c). Gene Ontology analysis showed ENKP_NK cells were enriched for cytotoxicity pathways (Extended Data Fig. 4d). ILCP_NK cells were characterized by biological processes including cell-cell adhesion, lymphocyte activation and regulation of IFN-γ production (Extended Data Fig. 4d). Heatmap clustering of ALP_NK cells showed two major column clusters which have genes expressed higher in ENKP_NK cells and ILCP_NK cells, respectively, with the "ENKP_NK" column cluster more abundant than "ILCP_NK" column cluster, further suggesting the NK population is a mixture of ENKP_NK cells and ILCP_NK cells, with ENKP_NK cells being more abundant (Extended Data Fig. 4e).

UMAP identified 2 clusters; NK cells from all three developmental origins fell into both clusters, with ENKP_NK cells enriched for cluster 0 whereas ILCP_NK cells enriched for cluster 1 (Extended Data Fig 4f). Genes differentially expressed between cluster 0 and cluster 1 were similar to genes differentially expressed between ENKP_NK cells and ILCP_NK cells (Extended Data Fig. 4a, g), which is consistent with the observation that ENKP_NK cells and ILCP_NK cells were enriched for cluster 0 and cluster 1, respectively (Extended Data Fig 4f). Although ENKP_NK cells and ILCP_NK cells both contributed to cluster 0 and cluster 1, differentially expressed genes, including *Klra8*, also could be detected within the same cluster when comparing NK cells derived from ENKPs and ILCPs (Extended Data Fig. 4h).

In addition, we utilized our bone marrow data (see Fig. 1a) and published liver data[27] to compare the transcriptomes of ENKP_NK cells, ILCP_NK cells and ILC1s. First, within bone marrow total NK cells and liver total NK cells, we identified ENKP_NK cells and ILCP_NK cells by expression levels of a list of genes which are differentially expressed between ENKP_NK cells and ILCP_NK cells (see Methods). Then we compared the transcriptome of both NK subsets and ILC1s in the same tissues, as transcriptome can be influenced by both tissue microenvironment[28] and batch effects. In both liver and bone marrow, ILCP_NK cells clustered together with ENKP_NK cells rather than with ILC1s (Extended Data Fig. 4i), indicating ILCP_NK cells are transcriptionally more similar to ENKP_NK cells than ILC1s. Interestingly, ILCP_NK cells expressed intermediate levels of genes which are differentially expressed between NK cells and ILC1s (see Methods), suggesting ILCP_NK cells may exhibit a hybrid phenotype, sharing some characteristics of both NK cells and ILC1s (Extended Data Fig. 4j).

In summary, these data indicate that ENKP_NK cells and ILCP_NK cells are transcriptionally different, with ENKP_NK cells expressing Ly49 receptors and higher levels of cytotoxic genes, and ILCP_NK cells expressing higher levels of genes related with IFN-γ production. The results suggested ENKP_NK cells and ILCP_NK cells may be functionally distinct.

Ly49H$^+$ NK cells primarily develop from ENKPs

Ly49H is a member of the Ly49 family of NK cell receptors and is expressed on approximately half of the NK cells in C57BL/6 mice[13,14,29]. Ly49H recognizes the viral protein m157 presented on cells infected by mouse cytomegalovirus (MCMV) and activates NK cells to eliminate virus-infected cells[14]. Our scRNA-seq analysis showed that *Klra8* (encoding Ly49H) expression in ENKP_NK cells was much higher than in ILCP_NK cells, suggesting Ly49H$^+$ NK cells may develop from ENKPs (Fig. 4b, c and Extended Data Fig. 4a-c). To test this hypothesis, we adoptively transferred ENKPs or ILCPs into NSG mice and assessed frequencies of Ly49H$^+$ NK cells developed from these progenitors. Indeed, ILCP_ NK cells were mostly Ly49H$^-$ whereas more than half of ENKP_NK cells expressed Ly49H (Fig. 4d). Adoptive transfer experiments using ALPs as competitors showed that ENKPs gave rise to Ly49H$^+$ NK cells much more efficiently than ILCPs (Extended Data Fig. 5a). Thus, the majority of Ly49H$^+$ NK cells develop from ENKPs but not ILCPs. Ly49H was not expressed on ENKPs (Extended Data Fig. 5b), indicating Ly49H is upregulated at more downstream stages during NK cell development. In addition, consistent with scRNA-seq data, ENKP_NK cells also had higher frequencies of cells expressing Ly49D (encoded by *Klra4*) but comparable frequencies of cells expressing Ly49E/F (encoded by *Klra5/Klra6*) (Extended Data Fig. 5c) and lower levels of TCF1-YFP (encoded by *Tcf7*^{YFP}) (Extended Data Fig. 5d-e).

When we examined the ratio of NK cells that were lineage traced with PLZF and the ratio of NK cells expressing Ly49H in different organs from PLZF-tracing long-term chimeras, there was an inverse correlation between PLZF labeling and Ly49H expression (Extended Data Fig. 5f-i). This is consistent with our finding that Ly49H$^+$ NK cells mostly derived from ENKPs and ENKP_NK cells are labelled by PLZF at lower ratios than ILCP_NK cells. Indeed, PLZF-tracing ratios were lower in Ly49H$^+$ NK cells as compared to Ly49H$^-$ NK cells (Extended Data Fig. 5f). Overall, PLZF-tracing ratios were lower in liver, spleen and lung (all ~18%) and higher in mesenteric lymph node (25%) and thymus (46%) (Extended Data Fig. 5f). NK cells in other tissues including salivary glands, uterus and bone marrow were also lowly traced by PLZF (Extended Data Fig. 5h). Because ENKP_NK cells, ILCP_NK cells and total NK cells were marked at different ratios in PLZF lineage-tracing mice (Fig. 3b, Extended Data Fig. 5f), and assuming that only ENKP_NK cells and ILCP_NK cells contributed to the total NK cell population, we were able to calculate the percentage of ENKP_NK cells and ILCP_NK cells (see

Methods). These calculations suggested more than 80% of NK cells in spleen, liver, lung, salivary glands, uterus and bone marrow develop from ENKPs. Consistently, the majority of ALP_NK cells recovered from spleen were transcriptionally similar to ENKP_NK cells (Extended Fig. 4e), and Ly49H^{+} NK cells were labeled by PLZF at ratios significantly lower than Ly49H^{-} NK cells in the tissues we assessed (Extended Data Fig 5f), further indicating that the majority of Ly49H^{+} NK cells develop from ENKPs.

Human NK subsets resemble mouse ENKP-derived and ILCP-derived NK cells

In humans, NK cells include two major subsets, termed CD56^{dim}CD16^{+} (CD56^{dim}) and CD56^{dim}CD16^{-} (CD56^{bright}) NK cells. Within the CD56^{dim} subset, human NK cells can be further subdivided into CD57^{-} and CD57^{+} populations, with the latter representing a more activated or mature population[30]. The CD56^{bright} NK cells are more efficient in producing cytokines including IFN-γ[31]. These observations are reminiscent of the mouse NK cell subsets elucidated in the present study.

To compare mouse NK cells with human NK cells, we analyzed previously published scRNA-seq data of human blood NK cells which includes both CITE-seq and scRNA-seq data[32,33]. This allowed us to distinguish CD56^{bright}, CD56^{dim}CD57^{-} and CD56^{dim}CD57^{+} clusters by surface protein expression of CD56 and CD57 (Extended Data Fig. 6a-b). These clusters were present in both human cytomegalovirus (HCMV) seropositive and HCMV seronegative individuals, with an enrichement of CD56^{dim}CD57^{+} NK cells in HCMV^{+} samples (Fig. 4e). Both CD56^{dim}CD57^{-} and CD56^{dim}CD57^{+} NK cells expressed higher levels of genes upregulated in ENKP_NK cells, whereas CD56^{bright} NK cells expressed higher levels of genes with higher expression in ILCP_NK cells (Fig. 4f). Human killer immunoglobin-like receptors (KIRs) are functionally similar to mouse Ly49 receptors[34]. NK cells expressing KIRs including *KIR2DL1*, *KIR2DL2* and *KIR2DL3* expanded upon HCMV infection[35]; these KIRs were mostly expressed in CD56^{dim} NK cells but not CD56^{bright} NK cells (Extended Data Fig. 6c), consistent with previous studies[36]. Importantly, CD56^{dim}CD57^{-} and CD56^{dim}CD57^{+} NK cells had higher ENKP_NK scores whereas CD56^{bright} NK cells had higher ILCP_NK scores (see Methods for scores), in both the HCMV^{+} and HCMV^{-} donors (Fig. 4g), suggesting HCMV status did not affect the transcriptional similarity between human NK subsets and mouse NK subsets. Furthermore, we also assessed the

correlation between each of the mouse NK clusters identified by scRNA-seq (Extended Data Fig. 4f) with each of the human NK clusters by transcriptional similarity. Mouse NK cluster 0 which was enriched in ENKP_ NK cells, was more correlated with human $CD56^{dim}CD57^{-}$ and $CD56^{dim}CD57^{+}$ NK cells whereas mouse NK cluster 1 which was enriched in ILCP_ NK cells, was more correlated with human $CD56^{bright}$ NK cells (Extended Data Fig. 6d). Consistently, the total ENKP_ NK cell population was more correlated with human $CD56^{dim}CD57^{-}$ and $CD56^{dim}CD57$ NK cells whereas total ILCP_ NK population was more correlated with human $CD56^{bright}$ NK cells (Extended Data Fig. 6e). Together, these results indicate that developemental heterogeneity in NK cells may be evolutionarily conserved between human and mouse.

ENKP-derived NK cells respond more robustly to MCMV infection than ILCP-derived NK cells

Since ENKPs and ILCPs have different lineage potentials for $Ly49H^{+}$ NK cells, we challenged ENKP or ILCP-reconstituted NSG mice with MCMV to compare the function of ENKP_NK cells and ILCP_NK cells. NSG mice were reconstituted with the same number of ENKPs or ILCPs. The mice were then challenged with MCMV and $Ly49H^{+}$ NK cells were subsequently assessed in spleen and blood (Fig 5a). Upon MCMV infection, the numbers of ILCP_NK cells in spleen remained low (Fig 5b-d). However, the numbers of ENKP_NK cells were approximately 150-fold higher than ILCP_NK cells in spleen with MCMV infection (Fig. 5d) whereas those were only 4-fold higher without MCMV infection (Fig. 5c). In addition, ENKP_NK cells consisted primarily of $Ly49H^{+}$ NK cells following MCMV infection (Fig. 5b and d). These data indicated ENKP-derived NK cells expanded dramatically. Consistently, $Ly49H^{+}$ ENKP_NK cells in blood also expanded robustly upon MCMV infection (Extended Data Fig. 7a-c). To additionally compare ENKP_NK cells and ILCP_NK cells, we co-transferred ENKPs and ILCPs in the same mice. Because ILCPs generated 4-fold fewer NK cells than ENKPs (Fig. 5c), we co-transferred of ENKPs and ILCPs at a 1:4 ratio to normalize the number of NK cells from each source prior to MCMV infection. Following infection, ENKP_ NK cells represented the overwhelming majority compared to ILCP_NK cells in the same environment (Extended Data Fig. 7d). In addition, in long-term PLZF lineage-tracing chimeras, the ratios of ILCP_NK-enriched PLZF-traced NK cells were significantly decreased after MCMV infection, suggesting PLZF-untraced (YFP^{-}) NK cells expanded to a greater degree than PLZF-traced (YFP^{+}) NK cells

(Fig. 5e-f), and further demonstrating that ENKP_NK cells expanded more than ILCP_NK cells during MCMV infection. In summary, ENKP_NK cells expanded more robustly than ILCP_NK cells upon MCMV infection.

ENKP_NK cells and ILCP_NK cells distinguished by surface markers

We identified surface markers differentially expressed between ENKP_NK cells and ILCP_NK cells by scRNA-seq, and verified some of these markers by flow cytometry, including Ly49H, Ly49D, CD226, Thy1.2 and CXCR3 (Fig. 4d, Extended Data Fig. 5c and 8a). Combinations of these markers allowed us to identify an ENKP_NK-enriched cell population ($Ly49H^+$ or $Ly49D^+$), which was lowly traced by PLZF (~10%); and an ILCP_NK-enriched cell population ($Ly49H^-Ly49D^-CD226^+Thy1.2^+$ or $CXCR3^+$), which was highly traced (~50%), similar to the tracing ratios in ILC1s in this set of experiments (~55%) (Fig. 6a). The ENKP_NK-enriched cell population expressed lower levels of TCF1-YFP than the ILCP_NK-enriched cell population (Extended Data Fig. 8b), consistent with the results by adoptive transfer (Extended Data Fig. 5d). Furthermore, the PLZF labeling ratios of the NK subsets identified by these surface markers remained unchanged after MCMV infection (Extended data Fig. 8c), indicating these markers identify populations that do not substantially interconvert. The expansion of the ENKP_NK-enriched cell population in long-term chimeras and wild-type B6 mice during MCMV infection was similar to the population expansion of $Ly49H^+$ NK cells in wild-type mice (Extended Data Fig. 8d-e), as reported[37].

Consistent with the adoptive transfer analysis (Extended Data Fig. 3i), both ENKP_NK-enriched and ILCP_NK-enriched cell populations but not ILC1s were greatly reduced in *Eomes*-deficient mice (Extended Data Fig. 8f). However, these NK subsets remained intact in *Zbtb16*-deficient mice (Extended Data Fig. 8g), indicating the developmental divergence between ENKP_NK cells and ILCP_NK cells does not depend on PLZF.

Downregulation of CD27 and upregulation of CD11b and KLRG1 is correlated with NK cell maturation[38,39]. However, when $CD27^+CD11b^-$ NK cells were adoptively transferred to new hosts, ~30% of them remained $CD11b^-$, suggesting the $CD27^+CD11b^-$ NK subset does not only contain immature NK cells[38]. Interestingly, the $CD27^+CD11b^-$ subset was enriched within ILCP_

NK cells compared to ENKP_NK cells by adoptive transfer of ILCPs or ENKPs (Extended Data Fig. 9a). Consistently, the $CD27^{+}CD11b^{-}$ NK subset was traced by PLZF at a higher ratio than $CD27^{-}CD11b^{+}$ NK subset (Extended Data Fig. 9b). Similarly, ENKP_NK cells expressed higher levels of KLRG1 (Extended Data Fig. 9c) and $KLRG1^{-}CD11b^{-}$ "immature" NK cells were traced by PLZF at a higher ratio than $KLRG1^{+}CD11b^{+}$ NK cells (Extended Data Fig. 9d). These data indicate that previously defined immature NK populations are likely a mixture of immature NK cells and mature ILCP_NK cells. We also assessed recently described $Ly49H^{+}CD27^{+}CD62L^{-}$ tissue-resident NK cells that are phenotypically similar to ILCP_NK cells[40]. Although ILCP_NK cells expressed lower levels of CD62L than ENKP_NK cells (Extended Data Fig. 9c), $Ly49H^{+}CD27^{+}CD62L^{-}$ NK cells were lowly traced by PLZF, indicating they are not derived from ILCPs but are likely a subset of ENKP_NK cells (Extended Data Fig. 9e).

ILCP_NK cells are better cytokine producers

To assess cytokine production and cytotoxicity by ENKP_NK cells versus ILCP_NK cells, we first separately sorted ILCP_NK-enriched and ENKP_NK-enriched cell populations, and stimulated them *in vitro* to assess their expression of IFN-γ, TNF-α and CD107a (Extended Data Fig. 10a). The ILCP_NK-enriched cell population contained higher frequencies of cells expressing IFN-γ and TNF-α but comparable frequencies of CD107a-expressing cells when compared to the ENKP_NK-enriched cell population (Extended Data Fig. 10b). In addition, we also assessed *in vitro* killing ability for both ENKP_NK-enriched and ILCP_NK-enriched cell populations by co-culturing them with Yac-1 target cells. We found both NK subsets killed Yac-1 cells comparably (Extended Data Fig. 10c). *In vitro* differentiated ILCP_NK cells also had higher frequencies of cells expressing IFN-γ upon stimulation as compared to ENKP_NK cells (Extended Data Fig. 10d). These data indicate that ILCP_NK cells have better cytokine producing ability than ENKP_NK cells.

Function of ILCP_NK cells and ENKP_NK cells upon Salmonella infection and Herpes Simplex Virus (HSV) infection

NK cells have been suggested to contribute to immunity against Salmonella infection[41,42]. We assessed total cell number and IFN-γ expression in ILCP_NK-enriched and ENKP_NK-enriched cell populations in cecum lamina propria upon Salmonella infection through oral gavage (Fig.

6b). Although both NK subsets increased in number after Salmonella infection (Fig. 6c and Extended Data Fig. 10e), the ILCP_NK-enriched cell population expressed higher levels of IFN-γ than ENKP_NK-enriched subset, indicating that ILCP_NK cells contribute to the immune response to Salmonella infection and have better cytokine producing ability than ENKP_NK cells (Fig. 6d,e and Extended Data Fig. 10f). Interestingly, the ILC1 population did not expand and expressed significantly lower levels of IFN-γ than the ILCP_NK-enriched cell population (Fig. 6d,e and Extended Data Fig. 10f). NK cells also can help control Herpes Simplex Virus (HSV) infection[43]. We assessed total cell number and IFN-γ expression in ILCP_NK-enriched and ENKP_NK-enriched cell populations in ear skin upon HSV infection by intradermal injection (Fig. 6e). Similar to Salmonella infection, both NK subsets expanded during HSV infection (Fig. 6f and Extended Data Fig. 10g) and the ILCP_NK-enriched cell population contained higher frequencies of cells expressing IFN-γ (Fig. 6g-i and Extended Data Fig. 10h), indicating that ILCP_NK cells contribute to the immune response to HSV infection and have better cytokine producing ability than ENKP_NK cells. Although the ILC1 population also expanded after HSV infection, the expansion of the ILCP_NK-enriched cell population was more robust (34-fold vs 3-fold on day 3, 40-fold vs 11-fold on day 7) (Fig. 6g). Thus these two additional models of infection each showed that both ILCP_NK cells and ENKP_NK cells contribute to immune responses, but ILCP_NK cells are more robust cytokine producers during these infections.

Discussion

The early development of NK cells and the relationship between functional heterogeneity of NK cells and developmental differences are both poorly understood. Here we established the existence of two separate pathways of NK cell development in mice. We identified ENKPs as a novel early NK progenitor which was the primary source of Ly49H^{+} NK cells, and established that ENKP_NK cells expanded dramatically upon MCMV infection. In contrast, ILCP_NK cells were heavily lineage-traced by PLZF, were largely depleted for expression of Ly49H, and failed to robustly respond to MCMV infection. ILCP_NK cells contained higher frequencies of cells expressing IFN-γ *in vitro* and upon infections with Salmonella or HSV, consistent with previous work establishing that Ly49H^{lo} NK cells possessed greater potential for IFN-γ production than Ly49H^{hi} NK cells[44]. Interestingly, mouse ENKP_NK and ILCP_NK cell populations were

transcriptionally and functionally similar to human CD56^{dim} and CD56^{bright} NK cells, respectively. However, it is likely that differences of IFN-γ expression between ILCP_NK cells and ENKP_NK cells are context-dependent as human CD56^{bright} NK cells produce more IFN-γ in response to cytokines[45] but less IFN-γ in response to target cells or engagement of activating receptors compared to CD56^{dim} NK cells[46].

Previous studies established that ILCPs give rise to NK cells *in vivo*[3,4], and ILCP_NK cells are phenotypically similar to previously described ILC1-like NK cells[47,48], although the developmental relationships between ILC1-like NK cells and ILCP_NK cells remain to be determined. ILCP_NK cells shared some phenotypes of ILC1s as well. However, our study showed that ILCP_NK cells also express the canonical NK cell transcription factor Eomes and both ILCP_NK cells and ENKP_NK cells but not ILC1s depend on *Eomes* for development, indicating they are distinct from previously described ILC1s[49,50]. We also compared transcriptomes of ILCP_NK cells, ENKP_NK cells and ILC1s, and found that ILCP_NK cells were transcriptionally more similar to ENKP_NK cells than to ILC1s. After Salmonella infection, both ILCP_NK-enriched and ENKP_NK-enriched cell populations expanded, but the ILC1 population did not. Thus, ILCP_NK cells are transcriptionally and functionally more similar to ENKP_NK cells than to ILC1s.

ILCPs gave rise to NK cells less efficiently than ENKPs, which could be explained by the lower frequency of precursors that have NK lineage potential within the ILCP population revealed by our clonal assay (Fig. 2b). Thus, the clonal burst of NK progeny may be equivalent or even higher from NK-lineage competent cells within ILCPs compared to ENKPs.

Mouse Ly49H^{+} NK cells have been shown to possess adaptive immune features after exposure to MCMV infection[13,30]. Similarly, human CD56^{dim}NKG2C^{+} NK cells expand robustly in HCMV-seropositive individuals[51], where HCMV-induced “memory” NK cells are enriched within the CD56^{dim} NK population. These similarities between mouse and human suggest that conserved developmental pathways may underlie the functional heterogeneity in both species. Human NKG2C and mouse Ly49H are similar in both function and adaptor DAP12 usage[13,51], and

HCMV and MCMV have similar genomic structures[52,53], suggesting co-evolution of CMV and NK cells expressing CMV-specific receptors[54].

Whether human CD56^{bright} NK cells are precursors of CD56^{dim} NK cells, or whether these populations also reflect developmental heterogeneity, is not clearly known[55,56]. However, transcriptional signatures derived from developmentally distinct mouse NK lineages were differentially expressed in human blood NK populations, suggesting that developmental heterogeneity also contributes to the heterogeneous human NK subsets. Indeed, barcoding-based lineage tracing revealed that CD56^{-}CD16^{+} and CD56^{+}CD16^{-} NK cells in non-human primates are developmentally different[57], indicating different development pathways of NK cells may exist and be conserved in multiple vertebrate species. In future, similar barcoding approaches will provide additional insight into the developmental pathways that generate NK cells in mice.

We described surface markers that discriminate NK populations enriched for ENKP_NK cells and ILCP_NK cells in wild-type B6 mice, which will facilitate future studies of the potential distinct functions of ENKP_ NK cells and ILCP_NK cells in other models. Recruitment of conventional type 1 dendritic cells (cDC1) in tumor microenvironment depends on NK cells that produce the chemoattractants XCL1 and CCL5[58]. Interestingly, *Xcl1* was highly expressed in ILCP_NK cells but not ENKP_NK cells whereas *Ccl5* was expressed more highly in ENKP_NK cells (Extended Data Fig. 4d), suggesting that ILCP_NK cells could be an important source of XCL1 production in some tumor microenvironments, and ILCP_NK cells and ENKP_NK cells may collaborate in eliminating tumors. Further, *XCL1* is also expressed more highly in human CD56^{bright} NK cells whereas *CCL5* is expressed more highly in CD56^{dim} NK cells (Fig. 4e and g); thus this potential collaboration between NK subsets might be conserved.

Motifs for ZEB2 and TCF7 were enriched in CD56^{dim}CD57^{+} NK cells and CD56^{bright} NK cells, respectively[10], implicating these transcription factors in regulating gene expression in specific NK subsets. *Zeb2* and *Tcf7* are also relatively highly expressed in ENKP_NK cells and ILCP_NK cells, respectively (Fig. 4c), thus gene regulatory programs in NK cells are predicted to be conserved between human and mouse. ILCP_NK cells expressed higher levels of *Cd69*

(Fig. 4b) and GO enrichment analysis also revealed ILCP_NK cells are involved in biological processes such as cell-cell adhesion, lymphocyte activation (Extended Data Fig. 4e), which suggests their tissue-resident features, similar to ILC1. We speculate that some specialization for tissue-residency is established developmentally, which is one direction for future work.

Our study implicates different developmental pathways as underlying functional heterogeneity of NK cells and identifies precursors for conventional NK cells expressing Ly49H in mice. We speculate ILCP_NK cells evolved initially, and viruses and perhaps specifically CMV drove the existence of a separate population of NK cells with adaptive features. Thus, although the activating receptors for response to CMV are different in humans and mice, the underlying logic of developmental pathways that generate NK cell subsets specialized for the same threat may be conserved. Further, if these developmental pathways were already present in the common ancestor of humans and mice, as seems likely, then we predict such developmental pathways are present in many related vertebrate species.

Methods

Mice

$Tcf7^{YFP}$, $Tcf7^{GFP}$, $Zbtb16^{GFPcre}$, $Tcf7^{-/-}$, $Nfil3^{-/-}$, $Il7r^{cre}$, $Gata3^{flox}$, $Eomes^{flox}$, $Eomes^{GFP}$ mice were described previously[2,3,16,23,24,59-62]. B6-Ly5.2 (CD45.1) and B6.129X1-$Gt(ROSA)26Sor^{tm1(EYFP)Cos}$/J (strain #: 006148) mice were obtained from Jackson Laboratory. NOD-scid $IL2Rgamma^{null}$ (NSG) mice were obtained from National Cancer Institute. All mice except NSG mice are maintained on C57BL/6 background. Mice were 6-18 weeks of age and of either sex. Animal procedures were approved by relevant National Institutes of Health Animal Care and Use Committees.

Antibodies and flow cytometry

Bone marrow, spleen and liver samples were incubated with a mix of purified rat, mouse and hamster immunoglobulin G (IgG) before the addition of specific antibodies, and then stained and analyzed in FACS buffer (PBS containing 2 mM EDTA and 0.5% FBS). The following anti-mouse antibodies were used for immunostaining: 2B4 (eBio244F4, PE-Cy7, Thermo Fisher Scientific, 25-2441-82), α4β7 (DATK32, BV711, BD biosciences, 740701; DATK32, PerCP-

Cy5.5, Thermo Fisher Scientific, 46-5887-82), B220 (RA3-6B2, biotin, Thermo Fisher Scientific, 13-0452-86), CD11c (N418, biotin, Thermo Fisher Scientific, 13-0114-85), CD122 (TM-b1, PE, Thermo Fisher Scientific, 12-1222-82), CD19 (eBio1D3, biotin, Thermo Fisher Scientific, 13-0193-85; 6D5, BV605, BioLegend, 115540), CD200R (OX110, APC, Thermo Fisher Scientific, 17-5201-82), CD25 (BC96, PE, Thermo Fisher Scientific, 12-0259-80), CD27 (LG.7F9, PerCP-Cy5.5, Thermo Fisher Scientific, 46-0271-82), CD3 (17A2, AF700, Thermo Fisher Scientific, 56-0032-82), CD3e (eBio500A2, biotin, Thermo Fisher Scientific, 13-0033-86; 145-2C11, BV510, BioLegend, 100353; 145-2C11, SB600, Thermo Fisher Scientific, 63-0031-82), CD4 (GK1.5, biotin, Thermo Fisher Scientific, 13-0041-82; RM4-5, SB600, Thermo Fisher Scientific, 63-0042-82), CD45.1 (A20, eF450, Thermo Fisher Scientific, 48-0453-82; A20, PE-eF610, Thermo Fisher Scientific, 61-0453-82), CD45.2 (104, eF450, Thermo Fisher Scientific, 48-0454-82; 104, PerCP-Cy5.5, BioLegend, 109828), CD49a (HMα1, APC, BioLegend, 142606), CD49b (DX5, AF700, Thermo Fisher Scientific, 56-5971-82; DX5, APC-Cy7, Thermo Fisher Scientific, 47-5971-82; DX5, PE-eF610, Thermo Fisher Scientific, 61-5971-82), CD5 (53-7.3, BV605, BD biosciences, 563194), CD8α (53-6.7, biotin, Thermo Fisher Scientific, 13-0081-85; 53-6.7, SB600, Thermo Fisher Scientific, 63-0081-82), CD8β (CT-CD8b, biotin, Thermo Fisher Scientific, MA5-17845; eBioH35-17.2, SB600, Thermo Fisher Scientific, 63-0083-82), Flt3 (A2F10.1, PE-CF594, BD biosciences, 562537), Gr-1 (RB6-8C5, biotin, Thermo Fisher Scientific, 13-5931-86), ICOS (7E.17G9, PE, Thermo Fisher Scientific, 12-9942-82), IL-7Rα (A7R34, APC, Thermo Fisher Scientific, 17-1271-82), kit (2B8, BV785, BioLegend, 105841; 2B8, PE-Cy7, Thermo Fisher Scientific, 25-1171-82), Ly49H (3D10, PE, BioLegend, 144706; 3D10, FITC, Thermo Fisher Scientific, 11-5886-82), Ly49D (4E5, BV 711, BD biosciences, 742559), Mac-1 (M1/70, biotin, Thermo Fisher Scientific, 13-0112-86), NK1.1 (PK136, APC-Cy7, Thermo Fisher Scientific, 47-5941-82; PK136, biotin, Thermo Fisher Scientific, 13-5941-85; PK136, BV785, BioLegend, 108749), NKp46 (29A1.4, PE-Cy7, Thermo Fisher Scientific, 25-3351-82), streptavidin (eF450, Thermo Fisher Scientific, 48-4317-82; APC-Cy7, Thermo Fisher Scientific, 47-4317-82), Sca1 (D7, AF700, Thermo Fisher Scientific, 56-5981-82), TCRβ (H57-597, biotin, Thermo Fisher Scientific, 13-5961-85; H57-597, SB600, Thermo Fisher Scientific, 63-5961-82), TCRγδ (eBioGL3, biotin, Thermo Fisher Scientific, 13-5711-85; eBioGL3, SB600, Thermo Fisher Scientific, 63-5711-82), Ter119 (TER-119, biotin, Thermo Fisher Scientific, 13-5921-85), Thy1.2 (53-2.1, AF700, BioLegend, 140324; 53-2.1,

BV605, BioLegend, 140318), CD226 (TX42.1, BV650, BioLegend, 133621), CXCR3 (CXCR3-173, APC, Thermo Fisher Scientific, 17-1831-82). The lineage 'cocktail' (Lin) is a mix of the following biotin antibodies: anti-Gr-1, NK1.1, CD19, Ter119, CD11c, CD3, CD4, CD8α, CD8β, TCRβ, TCRγδ, Ly-6D, B220, Mac-1. T lineage 'cocktail' (Lin^T) contains anti-CD3, CD4, CD8α, CD8β, TCRβ, TCRγδ. Non-T/NK lineage 'cocktail' contains anti-Gr-1, CD19, Ter119, CD11c, Ly-6D, B220.

BM progenitor definition and isolation

BM progenitors were defined and isolated by flow cytometric sort as follows. ALPs ($Lin^-Kit^+\alpha4\beta7^-Tcf7^-Flt3^+IL\text{-}7R\alpha^+$), ILCPs ($Lin^-Kit^+\alpha4\beta7^+Tcf7^{YFP+}Flt3^-Thy1.2^+$ or $Lin^-Kit^+\alpha4\beta7^+PD\text{-}1^+Flt3^-Thy1.2^+$) and ENKPs ($Lin^-CD27^+2B4^+Flt3^+CD122^+$) were isolated from $Tcf7^{YFP}$ mice. For PLZF tracing (Fig 3b), ILCPs were sorted as $Lin^-Kit^+\alpha4\beta7^+Tcf7^{GFP+}Flt3^-Thy1.2^+$. Before sorting of BM progenitors, bone marrow samples were stained with biotin lineage 'cocktail' (Lin), and then depleted by streptavidin beads (Biolegend, 480016) according to the manufacturer's instructions.

Cell culture

BM progenitors were sorted by flow cytometric cell sorting. BM progenitors were cultured on irradiated OP9 stromal cells in α-MEM supplemented with 20% fetal bovine serum, penicillin and streptomycin. Bulk cultures were supplemented with 30 ng/ml SCF, 30 ng/ml IL-7, 30 ng/ml Flt3L and 30 ng/ml IL-2, and progenitor differentiation was examined at 7d after seeding. Single cell cultures were supplemented with 30 ng/ml SCF, 30 ng/ml IL-7, 30 ng/ml Flt3L and 10 ng/ml IL-2, and progenitor differentiation was examined at 10d after seeding. For single cell assay, positive well was defined as > 10 total ILCs, and positive lineage was defined as frequency > 10% and number > = 3. For *in vitro* stimulation, NK cells sorted from bulk culture were incubated with IL-12 (20 ng/ml), IL-18 (10 ng/ml) or PMA (50 ng/ml), Ionomycin (500 ng/ml) for 1 hr, then Brefeldin A was added to a final concentration of 5 ug/ml for another 2 hrs' incubation. Cytokines were purchased from PeproTech.

Adoptive transfer

For ENKP and ILCP transfer, around 200-500 progenitors (CD45.2) were injected into NSG recipient mice (CD45.1) if not mentioned specifically. For competitive assay, around 300-500 test progenitors (CD45.2) were mixed with same numbers of competitor progenitors (CD45.1) and injected into NSG recipient mice (CD45.1). Reconstitution of lymphocytes was examined at 6-weeks post-transfer. For PLZF lineage tracing chimeras, $3X10^5$ YFP^- LK (Lin^-Kit^+) progenitors (CD45.2) obtained by cell sorting were injected into lethally irradiated (850 rads) wild-type B6-Ly5.2 recipient mice (CD45.1), or injected into sublethally irradiated (250 rads) NSG mice. Reconstitution of lymphocytes and progenitors was examined at 12-weeks post-transfer.

Calculation of ENKP_NK cell and ILCP_NK cell contributions to total NK population

To calculate contribution percentage of ENKP_NK cells and ILCP_NK cells, the following equation set was used: 10%*X + 70%*Y = a%; X + Y = 100%. 10% was the observed PLZF tracing ratio of ENKP_NK cells, 70% was the observed PLZF tracing ratio of ILCP_NK cells, a% was the observed PLZF tracing ratio of total NK cells in different tissues (for example, in liver a% was 18%, Extended Fig. 5h), X was the calculated percentage of ENKP_NK cells and Y was the calculated percentage of ILCP_NK cells.

scRNA-seq and analysis for mouse samples

Cells from bone marrow were isolated by sorting from $Tcf7^{YFP}$ mice (Fig. 1). NK cells from spleen were sorted from NSG mice adoptively transferred with ALPs, ENKPs or ILCPs (Fig 4). scRNA-seq libraries were prepared using 10X Genomics Single Cell 3' Reagent Kits v3.1, according to the manufacturer's instructions. The obtained libraries were sequenced with a NextSeq system. Base calling was performed using RTA 3.9.2 (bone marrow samples) or RTA 3.9.25 (spleen sample), and alignment was performed using cellranger v6.1.1 (STAR 2.7.2a). For spleen sample, hashtaging antibodies used for the analysis were mHTO01, mHTO02, and mHTO03 TotalSeqA bought from BioLegend, and demultiplexing was performed using cellranger v6.1.1 (Bcl2fastq 2.20.0). Cells with low UMI counts were determined by the 10x Genomics Cell Ranger Algorithm. Single-cell analysis was performed using Seurat (version 4.1.0). Cells with >6% mitochondrial gene content and >8000 genes (bone marrow samples) or >4000 genes (spleen sample) were excluded from analysis. Mitochondrial content and cell-

cycle related genes were regressed out. T cell contaminants were removed from bone marrow samples. Cells were analyzed mainly by Seurat's functions[63]. Cells were clustered using Seurat's "FindClusters" function. Pseudotime analysis was performed using "Slingshot" package (version 2.2.1)[64]. Differentially expressed genes (DEGs) were identified using Seurat's "FindMarkers" function considering those genes expressed in at least 10% of cells in the clusters with a minimum log2 fold change threshold of 0.25. Violin plots were made using Seurat's normalized data. Average gene expression from cells of the same developmental origin was generated using the Seurat's "AverangeExpression" function. Gene ontology pathway analysis of DEGs was carried out with the "enrichGO" function. Volcano plots were generated using "ggplot2" package (version 3.4.0).

To compare transcriptomes of mouse NK subsets and ILC1s, liver ILC1 and NK cell data were downloaded from the gene expression omnibus (GSE163452[27]) and loaded into Seurat. In addition, in house generated bone marrow derived NK and ILC1 scRNAseq (see Figure 1a) was loaded into Seurat. For published liver data, the hashtags used to label the publicly available data were demultiplexed using the HTODemux function, following CLR normalization. Cells from Hobit-KO mice were discarded, and cells were flagged as low quality and removed if they had <= 1000 or >= 8000 reads, and > 6% mitochondrial reads. All datasets were processed according to a standard Seurat pipeline. The FindNeighbors and FindClusters functions were used to cluster the cells and identify NK cells or ILC1s. Gene set scores were calculated using ENKP_NK gene list (*Zeb2, Irf8, Klf2, Gzmb, Gzma, Prf1, Itgam, Ccl5, Klra3, Klra4, Klra7, Klra8, Klra9*) and ILCP_NK gene list (*Tcf7, Cd160, Cd226, Cd7, Cd27, Ccr2, Cxcr3, Xcl1, Ltb, Ifng, Cd69*). These scores were fed into the kmeans function, allowing cells to be clustered using these gene scores. This resulted in the identification of 2 NK cell clusters, allowing them to be annotated as ENKP_NK or ILCP_NK. For each population of NK cells and ILC1s the average gene expression was calculated using the AverageExpression function, and Euclidean distance was calculated using the dist function. The averaged data was used to examine expression of genes previously determined to be differentially expressed between NK cells and ILC1s[17] that also were significantly different between NK cells and ILC1s in both liver and bone marrow, with a fold change >= 0.25 and FDR of <= 0.05.

scRNA-seq analysis of human NK cell

The human NK scRNAseq data was obtained from GSE130430[32] and GSE197037[33] was analysed using a standard Seurat analysis. Batch correction was performed using the RunHarmony function from the "harmony" package to correct batch effects. The CITEseq data was normalized by applying a centered log ratio normalization. Clusters identified using the RNA modality were annotated based on the protein expression of CD56 and CD57. The annotation was extended using the glmnet and cv.glmnet function to all datasets. Cross species comparisons were done by identifying homologs between mouse and human using the homologene package; and by turning mouse genes to uppercase. Differentially expressed gene (DEG)s were identified by applying the "FindMarkers" function from Seurat within the populations or samples of one species. The DEG shared between species were retained and for each population or sample the counts of individual cells were summed using the "AggregateExpression" function and log transformed, generating pseudobulk data. The genes of the pseudobulk data were centered by dividing by their average gene expression. Spearman correlation was done on these matrices using the "cor" function, allowing correlation of the populations between different species. Scores were calculated by averageing the expression of gene sets identified in the mouse scRNA-seq data (ENKP_NK score: ZEB2, IRF8, KLF2, GZMB, GZMA, PRF1, ITGAM and CCL5; ILCP_NK score: TCF7, CD160, CD226, CD7, CCR2, CXCR3, XCL1, LTB, IFNG and CD69) and substracting their average by the average gene expression of 100 randomly sampled genes. Plots were generated using either the "Seurat" or "ggplot2" packages.

MCMV infection

MCMV (Smith strain) was prepared as described previously[44] and virus titer was 2200 PFU/ml. MCMV stock was 1:400 diluted in 1XPBS and 200 µl of MCMV diluent was intraperitoneally injected into each mouse. Blood and spleen samples were analyzed by flow cytometry 7 days after infection.

Salmonella infection and cecal immune cell isolation

Salmonella Typhimurium strain IR715, a nalidixic acid resistant strain, was used for all experiments. Bacteria was cultured in Luria broth (LB) at 37°C shaking for 18 hours. Cells were washed in PBS and adjusted to final concentration of $5x10^8$ CFU/mL in PBS. C57BL6 mice, 8–

10-week-old, were gavaged with 20mg streptomycin, and fasted overnight. 24 hours later mice were gavaged with 10^8 CFU S. Typhimurium in 200 uL PBS. Infection was confirmed by measuring CFU at 24 hours by homogenizing fecal pellets in PBS and plating on LB agar with 50μg/mL nalidixic acid. Cecal contents was also plated at takedown to quantify S. Typhimurium CFU. Mice were sacrificed at 1, 3 or 5 days post infection to analyze gut immune cells in the cecum. Immune cells from the cecum of S. Typhimurium infected mice were isolated as previously described with some modifications[65]. Briefly, ceca were extracted and contents removed into sterile PBS for CFU quantification. Ceca were then cut open longitudinally, washed and incubated in complete medium (RPMI 1640 plus 3% fetal bovine serum, 1 mM pyruvate, 0.1 mM nonessential amino acids, 2 mM glutamine, 20 mM HEPES, 100 U/ml penicillin, 0.1 mg/ml streptomycin, and 50 μM β-mercaptoethanol) with 5mM EDTA and 1mM dithiothreitol for 15 minutes at 37°C stirring. Ceca were shaken 3 times in 10mL RPMI + 2mM EDTA to remove epithelial cells, then cut into small pieces for digestion in RPMI media containing 0.5 mg/ml Collagenase Type 4 (Worthington Biochemical), 8% FBS and 50 mM $CaCl_2$. After a 30-minute digestion at room temperature with stirring, cells were passed through a 70μm filter and washed with 10mL complete media. Finally, cells were pelleted, resuspended in 37.5% Percoll, spun at 600g for 5 minutes, and cell pellets were resuspended in complete media for flow cytometry analysis.

Herpes simplex virus-1 (HSV-1) infection and ear digestion

Female mice were injected intradermally with 15uL of HSV-1 ($1x10^6$ PFU in sterile tissue-culture-grade PBS) in the ear pinnae using a 29G needle. Control mice were injected with same amount of sterile tissue-culture-grade PBS. Immune responses were analyzed at day 3 or 7 post infection. At day 3 or 7 post infection, mice were euthanized and ears were cut and the ventral and dorsal sheets of ears were separated and digested in the media with DNase and Liberase (RPMI 1640 media with 2 mM L-glutamine, 1 mM sodium pyruvate and nonessential amino acids, 55 mM β-mercaptoethanol, 20 mM HEPES, 100 U/ml penicillin, 100 mg/ml streptomycin, 0.5 mg/ml DNase and 0.025 mg/ml Liberase TL purified enzyme blend) at 37°C for 1 hour and 45min. Digested ears were homogenized and filtered through 70-μm cell strainer.

***In vitro* cytotoxicity assay**

Freshly sorted spleen NK cells were co-cultured with Yac-1 mouse lymphoma cells at 4:1, 2:1 and 1:1 ratios in IMDM media (supplemented with 20% fetal bovine serum, nonessential amino acids, L-glutamine, 2-mercaptoethanol, penicillin and streptomycin) for 4 hours. Yac-1 cells were traced by CFSE to distinguish from NK cells. Specific lysis was calculated by the formula: (a - b) / a × 100%; a is the number of Yac-1 cells without NK cell co-culture; b is the number of Yac-1 cells with NK cell co-culture.

Statistical analysis

Statistical significance was performed with GraphPad Prism. Differences between groups were determined by a two-tailed unpaired or paired Student's t test and were considered to be statistically significant when P values were less than 0.05. P value < 0.05 (*), P value < 0.01(**), P value < 0.001 (***).

Data availability

Single cell RNA sequencing data will be deposited on GEO and made available.

Acknowledgements

We thank V. Lazarevic and C. Spinner for help acquiring mouse strains, and V. Shapiro for critical comments on the manuscript. We are grateful for support by the staff of the NCI CCR Single Cell Analysis Facility for sequencing bone marrow and spleen samples and support by the staff of CCR/NCI flow cytometry core for cell sorting. This work is supported by the Division of Intramural Research of the NCI, NIAID and NHLBI, ANR JCJC AAPG2022: ANR-22-CE15-0040-01 (C.H.).

Author contributions

Conceptualization: A.B., Y.D.; Methodology: Y.D., M.L., S.G., V.B., L.C., S.C.S, J.M.; Investigation: Y.D., M.L., S.G., V.B., L.C., A.D., S.D., C.H., S.C.S., J.M., D.P., Y.Z., J.Z., Y.B.,J.C.S.; Visualization: Y.D., M.L.; Supervision: A.B.; Writing: Y.D., M.L., V.B., L.C., A.B.

Competing interests

Authors declare that they have no competing interests.

References

1 Vivier, E. *et al.* Innate Lymphoid Cells: 10 Years On. *Cell* **174**, 1054-1066 (2018). https://doi.org:10.1016/j.cell.2018.07.017

2 Yang, Q. *et al.* TCF-1 upregulation identifies early innate lymphoid progenitors in the bone marrow. *Nat Immunol* **16**, 1044-1050 (2015). https://doi.org:10.1038/ni.3248

3 Constantinides, M. G., McDonald, B. D., Verhoef, P. A. & Bendelac, A. A committed precursor to innate lymphoid cells. *Nature* **508**, 397-401 (2014). https://doi.org:10.1038/nature13047

4 Xu, W. *et al.* An Id2(RFP)-Reporter Mouse Redefines Innate Lymphoid Cell Precursor Potentials. *Immunity* **50**, 1054-1068 e1053 (2019). https://doi.org:10.1016/j.immuni.2019.02.022

5 Klose, C. S. N. *et al.* Differentiation of type 1 ILCs from a common progenitor to all helper-like innate lymphoid cell lineages. *Cell* **157**, 340-356 (2014). https://doi.org:10.1016/j.cell.2014.03.030

6 Fathman, J. W., Bhattacharya, D., Inlay, M. A., Seita, J., Karsunky, H. & Weissman, I. L. Identification of the earliest natural killer cell-committed progenitor in murine bone marrow. *Blood* **118**, 5439-5447 (2011). https://doi.org:10.1182/blood-2011-04-348912

7 Constantinides, M. G. *et al.* PLZF expression maps the early stages of ILC1 lineage development. *Proc Natl Acad Sci U S A* **112**, 5123-5128 (2015). https://doi.org:10.1073/pnas.1423244112

8 Park, E. *et al.* Toxoplasma gondii infection drives conversion of NK cells into ILC1-like cells. *Elife* **8** (2019). https://doi.org:10.7554/eLife.47605

9 McFarland, A. P. *et al.* Multi-tissue single-cell analysis deconstructs the complex programs of mouse natural killer and type 1 innate lymphoid cells in tissues and circulation. *Immunity* **54**, 1320-1337 e1324 (2021). https://doi.org:10.1016/j.immuni.2021.03.024

10 Collins, P. L. *et al.* Gene Regulatory Programs Conferring Phenotypic Identities to Human NK Cells. *Cell* **176**, 348-360 e312 (2019). https://doi.org:10.1016/j.cell.2018.11.045

11 Dogra, P. *et al.* Tissue Determinants of Human NK Cell Development, Function, and Residence. *Cell* **180**, 749-763 e713 (2020). https://doi.org:10.1016/j.cell.2020.01.022

12 Karo, J. M., Schatz, D. G. & Sun, J. C. The RAG recombinase dictates functional heterogeneity and cellular fitness in natural killer cells. *Cell* **159**, 94-107 (2014). https://doi.org:10.1016/j.cell.2014.08.026

13 Sun, J. C., Beilke, J. N. & Lanier, L. L. Adaptive immune features of natural killer cells. *Nature* **457**, 557-561 (2009). https://doi.org:10.1038/nature07665

14 Smith, H. R. *et al.* Recognition of a virus-encoded ligand by a natural killer cell activation receptor. *Proc Natl Acad Sci U S A* **99**, 8826-8831 (2002). https://doi.org:10.1073/pnas.092258599

15 Harly, C., Cam, M., Kaye, J. & Bhandoola, A. Development and differentiation of early innate lymphoid progenitors. *J Exp Med* **215**, 249-262 (2018). https://doi.org:10.1084/jem.20170832

16 Harly, C. *et al.* The transcription factor TCF-1 enforces commitment to the innate lymphoid cell lineage. *Nat Immunol* **20**, 1150-1160 (2019). https://doi.org:10.1038/s41590-019-0445-7

17 Rodriguez-Rodriguez, N. *et al.* Identification of aceNKPs, a committed common progenitor population of the ILC1 and NK cell continuum. *Proc Natl Acad Sci U S A* **119**, e2203454119 (2022). https://doi.org:10.1073/pnas.2203454119

18 Zeis, P. *et al.* In Situ Maturation and Tissue Adaptation of Type 2 Innate Lymphoid Cell Progenitors. *Immunity* **53**, 775-792 e779 (2020). https://doi.org:10.1016/j.immuni.2020.09.002

19 Lim, A. I. *et al.* Systemic Human ILC Precursors Provide a Substrate for Tissue ILC Differentiation. *Cell* **168**, 1086-1100 e1010 (2017). https://doi.org:10.1016/j.cell.2017.02.021

20 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). https://doi.org:10.1038/ni.1787

21 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). https://doi.org:10.1016/j.immuni.2011.11.016

22 Yagi, R. *et al.* The transcription factor GATA3 is critical for the development of all IL-7Ralpha-expressing innate lymphoid cells. *Immunity* **40**, 378-388 (2014). https://doi.org:10.1016/j.immuni.2014.01.012

23 Zhong, C. *et al.* Differential Expression of the Transcription Factor GATA3 Specifies Lineage and Functions of Innate Lymphoid Cells. *Immunity* **52**, 83-95 e84 (2020). https://doi.org:10.1016/j.immuni.2019.12.001

24 Schlenner, S. M. *et al.* Fate mapping reveals separate origins of T cells and myeloid lineages in the thymus. *Immunity* **32**, 426-436 (2010). https://doi.org:10.1016/j.immuni.2010.03.005

25 Weizman, O. E. *et al.* ILC1 Confer Early Host Protection at Initial Sites of Viral Infection. *Cell* **171**, 795-808 e712 (2017). https://doi.org:10.1016/j.cell.2017.09.052

26 Tu, T. C. *et al.* CD160 is essential for NK-mediated IFN-gamma production. *J Exp Med* **212**, 415-429 (2015). https://doi.org:10.1084/jem.20131601

27 Friedrich, C. *et al.* Effector differentiation downstream of lineage commitment in ILC1s is driven by Hobit across tissues. *Nat Immunol* **22**, 1256-1267 (2021). https://doi.org:10.1038/s41590-021-01013-0

28 Ricardo-Gonzalez, R. R. *et al.* Tissue signals imprint ILC2 identity with anticipatory function. *Nat Immunol* **19**, 1093-1099 (2018). https://doi.org:10.1038/s41590-018-0201-4

29 Smith, K. M., Wu, J., Bakker, A. B., Phillips, J. H. & Lanier, L. L. Ly-49D and Ly-49H associate with mouse DAP12 and form activating receptors. *J Immunol* **161**, 7-10 (1998).

30 O'Sullivan, T. E., Sun, J. C. & Lanier, L. L. Natural Killer Cell Memory. *Immunity* **43**, 634-645 (2015). https://doi.org:10.1016/j.immuni.2015.09.013

31 Fehniger, T. A. *et al.* Differential cytokine and chemokine gene expression by human NK cells following activation with IL-18 or IL-15 in combination with IL-12: implications for the innate immune response. *J Immunol* **162**, 4511-4520 (1999).

32 Yang, C. *et al.* Heterogeneity of human bone marrow and blood natural killer cells defined by single-cell transcriptome. *Nat Commun* **10**, 3931 (2019). https://doi.org:10.1038/s41467-019-11947-7

33 Ruckert, T., Lareau, C. A., Mashreghi, M. F., Ludwig, L. S. & Romagnani, C. Clonal expansion and epigenetic inheritance of long-lasting NK cell memory. *Nat Immunol* **23**, 1551-1563 (2022). https://doi.org:10.1038/s41590-022-01327-7

34 Kelley, J., Walter, L. & Trowsdale, J. Comparative genomics of natural killer cell receptor gene clusters. *PLoS Genet* **1**, 129-139 (2005). https://doi.org:10.1371/journal.pgen.0010027

35 Beziat, V. *et al.* NK cell responses to cytomegalovirus infection lead to stable imprints in the human KIR repertoire and involve activating KIRs. *Blood* **121**, 2678-2688 (2013). https://doi.org:10.1182/blood-2012-10-459545

36 Jacobs, R. *et al.* CD56bright cells differ in their KIR repertoire and cytotoxic features from CD56dim NK cells. *Eur J Immunol* **31**, 3121-3127 (2001). https://doi.org:10.1002/1521-4141(2001010)31:10<3121::aid-immu3121>3.0.co;2-4

37 Dokun, A. O., Kim, S., Smith, H. R., Kang, H. S., Chu, D. T. & Yokoyama, W. M. Specific and nonspecific NK cell activation during virus infection. *Nat Immunol* **2**, 951-956 (2001). https://doi.org:10.1038/ni714

38 Chiossone, L., Chaix, J., Fuseri, N., Roth, C., Vivier, E. & Walzer, T. Maturation of mouse NK cells is a 4-stage developmental program. *Blood* **113**, 5488-5496 (2009). https://doi.org:10.1182/blood-2008-10-187179

39 Huntington, N. D. *et al.* NK cell maturation and peripheral homeostasis is associated with KLRG1 up-regulation. *J Immunol* **178**, 4764-4770 (2007). https://doi.org:10.4049/jimmunol.178.8.4764

40 Flommersfeld, S. *et al.* Fate mapping of single NK cells identifies a type 1 innate lymphoid-like lineage that bridges innate and adaptive recognition of viral infection. *Immunity* **54**, 2288-2304 e2287 (2021). https://doi.org:10.1016/j.immuni.2021.08.002

41 Harrington, L., Srikanth, C. V., Antony, R., Shi, H. N. & Cherayil, B. J. A role for natural killer cells in intestinal inflammation caused by infection with Salmonella enterica serovar Typhimurium. *FEMS Immunol Med Microbiol* **51**, 372-380 (2007). https://doi.org:10.1111/j.1574-695X.2007.00313.x

42 Kupz, A. *et al.* Contribution of Thy1+ NK cells to protective IFN-gamma production during Salmonella typhimurium infections. *Proc Natl Acad Sci U S A* **110**, 2252-2257 (2013). https://doi.org:10.1073/pnas.1222047110

43 Habu, S., Akamatsu, K., Tamaoki, N. & Okumura, K. In vivo significance of NK cell on resistance against virus (HSV-1) infections in mice. *J Immunol* **133**, 2743-2747 (1984).

44 Adams, N. M. *et al.* Cytomegalovirus Infection Drives Avidity Selection of Natural Killer Cells. *Immunity* **50**, 1381-1390 e1385 (2019). https://doi.org:10.1016/j.immuni.2019.04.009

45 Cooper, M. A. *et al.* Human natural killer cells: a unique innate immunoregulatory role for the CD56(bright) subset. *Blood* **97**, 3146-3151 (2001). https://doi.org:10.1182/blood.v97.10.3146

46 Fauriat, C., Long, E. O., Ljunggren, H. G. & Bryceson, Y. T. Regulation of human NK-cell cytokine and chemokine production by target cell recognition. *Blood* **115**, 2167-2176 (2010). https://doi.org:10.1182/blood-2009-08-238469

47 Lopes, N., Galluso, J., Escaliere, B., Carpentier, S., Kerdiles, Y. M. & Vivier, E. Tissue-specific transcriptional profiles and heterogeneity of natural killer cells and group 1 innate lymphoid cells. *Cell Rep Med* **3**, 100812 (2022). https://doi.org:10.1016/j.xcrm.2022.100812

48 Yomogida, K. *et al.* Hobit confers tissue-dependent programs to type 1 innate lymphoid cells. *Proc Natl Acad Sci U S A* **118** (2021). https://doi.org:10.1073/pnas.2117965118

49 Nixon, B. G. *et al.* Cytotoxic granzyme C-expressing ILC1s contribute to antitumor immunity and neonatal autoimmunity. *Sci Immunol* **7**, eabi8642 (2022). https://doi.org:10.1126/sciimmunol.abi8642

50 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). https://doi.org:10.7554/eLife.01659

51 Lopez-Verges, S. *et al.* Expansion of a unique CD57(+)NKG2Chi natural killer cell subset during acute human cytomegalovirus infection. *Proc Natl Acad Sci U S A* **108**, 14725-14732 (2011). https://doi.org:10.1073/pnas.1110900108

52 Chee, M. S. *et al.* Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr Top Microbiol Immunol* **154**, 125-169 (1990). https://doi.org:10.1007/978-3-642-74980-3_6

53 Rawlinson, W. D., Farrell, H. E. & Barrell, B. G. Analysis of the complete DNA sequence of murine cytomegalovirus. *J Virol* **70**, 8833-8849 (1996). https://doi.org:10.1128/JVI.70.12.8833-8849.1996

54 Sun, J. C. & Lanier, L. L. The Natural Selection of Herpesviruses and Virus-Specific NK Cell Receptors. *Viruses* **1**, 362 (2009). https://doi.org:10.3390/v1030362

55 Poli, A., Michel, T., Theresine, M., Andres, E., Hentges, F. & Zimmer, J. CD56bright natural killer (NK) cells: an important NK cell subset. *Immunology* **126**, 458-465 (2009). https://doi.org:10.1111/j.1365-2567.2008.03027.x

56 Melsen, J. E., Lugthart, G., Lankester, A. C. & Schilham, M. W. Human Circulating and Tissue-Resident CD56(bright) Natural Killer Cell Populations. *Front Immunol* **7**, 262 (2016). https://doi.org:10.3389/fimmu.2016.00262

57 Wu, C. *et al.* Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells. *Cell Stem Cell* **14**, 486-499 (2014). https://doi.org:10.1016/j.stem.2014.01.020

58 Bottcher, J. P. *et al.* NK Cells Stimulate Recruitment of cDC1 into the Tumor Microenvironment Promoting Cancer Immune Control. *Cell* **172**, 1022-1037 e1014 (2018). https://doi.org:10.1016/j.cell.2018.01.004

59 Wang, R. *et al.* T cell factor 1 regulates thymocyte survival via a RORgammat-dependent pathway. *J Immunol* **187**, 5964-5973 (2011). https://doi.org:10.4049/jimmunol.1101205

60 Zhu, Y. *et al.* T-bet and eomesodermin are required for T cell-mediated antitumor immune responses. *J Immunol* **185**, 3174-3183 (2010). https://doi.org:10.4049/jimmunol.1000749

61 Daussy, 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). https://doi.org:10.1084/jem.20131560

62 Kashiwada, M. *et al.* IL-4-induced transcription factor NFIL3/E4BP4 controls IgE class switching. *Proc Natl Acad Sci U S A* **107**, 821-826 (2010). https://doi.org:10.1073/pnas.0909235107

63 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021). https://doi.org:10.1016/j.cell.2021.04.048

64 Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018). https://doi.org:10.1186/s12864-018-4772-0

65 Stacy, A. *et al.* Infection trains the host for microbiota-enhanced resistance to pathogens. *Cell* **184**, 615-627 e617 (2021). https://doi.org:10.1016/j.cell.2020.12.011

Figure 1 Characterization of early NK cell development in bone marrow

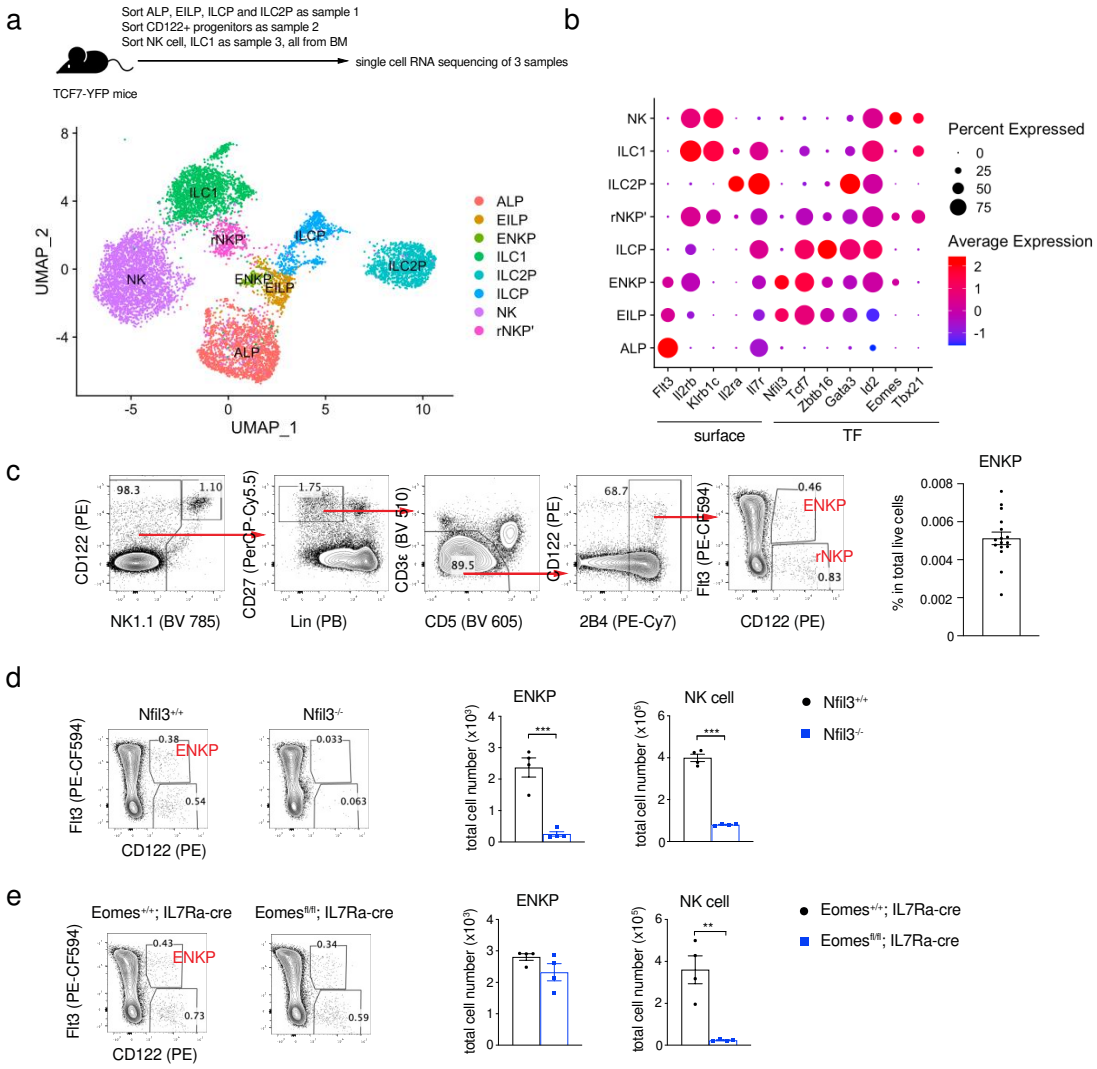


Fig.1 Characterization of early NK cell development in bone marrow. **a** - **b**, scRNA-seq of bone marrow cells**:** Experimental schematic and UMAP (**a**), and dotplots (**b**) of expression of surface genes and transcription factors grouped by clusters in **a**. **c,** Gating strategy and frequency of ENKPs combined from 3 independent experiments. **d,** Representative flow plots of ENKPs and quantifications of ENKPs and NK cells in bone marrow of *Nfil3*-deficient mice, n = 4. Combined data from 2 independent experiments. **e,** Representative flow plots of ENKPs and quantifications of ENKPs and NK cells in bone marrow of *Eomes*-deficient mice, n = 4. Combined data from 2 independent experiments. For **c**, **d** and **e**, data are mean ± s.e.m, and for **d** and **e**, statistical analysis was performed using two-tailed unpaired Student's t-tests.

Figure 2 ENKPs are early NK-lineage biased progenitors

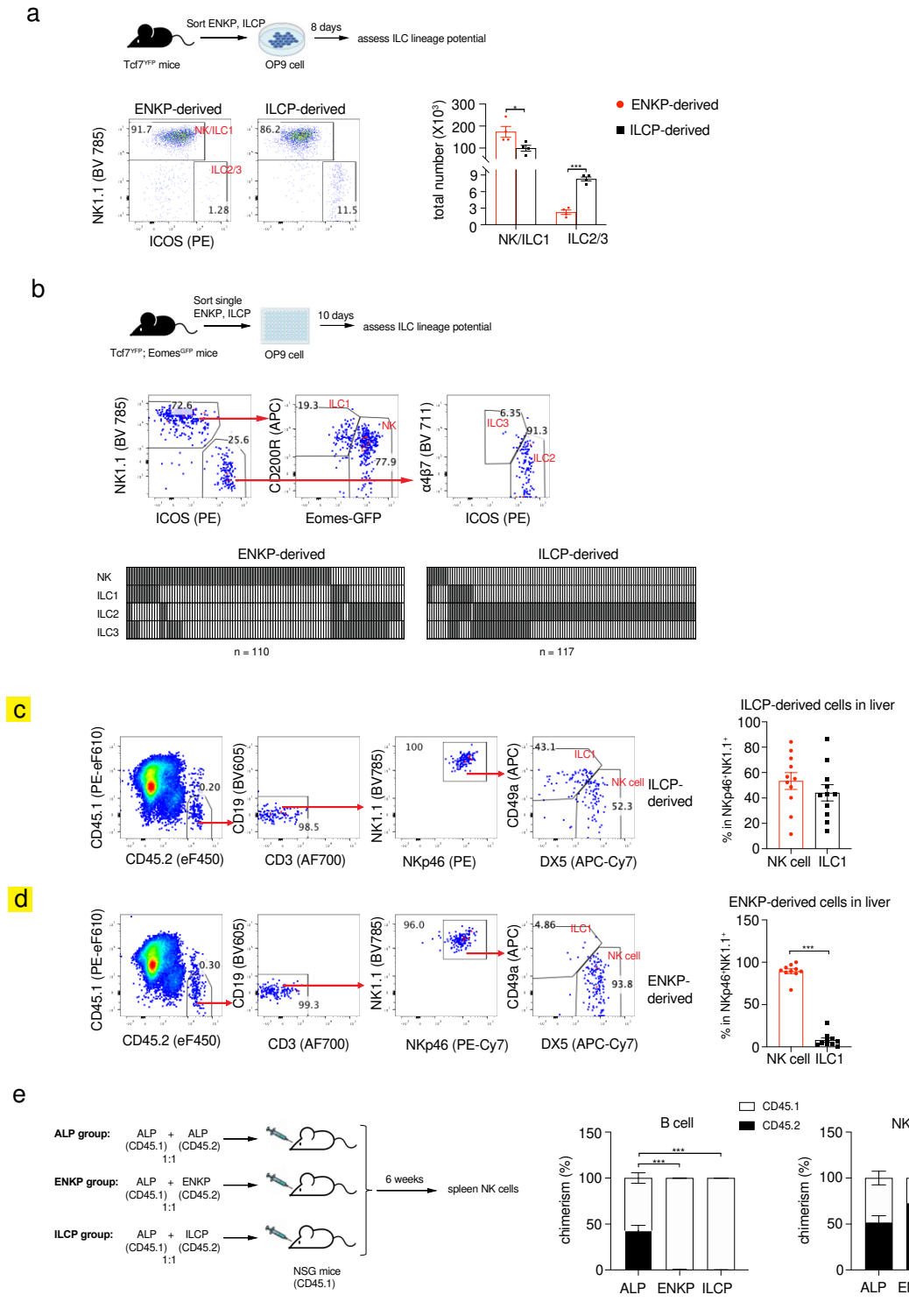


Fig. 2 ENKPs are early NK-lineage biased progenitors. a, Strategy of in vitro bulk assay for ILC lineage potential, representative flow plots and quantification of in vitro differentiated ILCs, n = 4. Representative data from

3 independent experiments. **b,** Strategy of in vitro clonal assay for ILC lineage potential, gating strategy of ILCs, and result of individual positive clones combined from 2 independent experiments. Each column represents one well, with detected ILC lineages in grey. **c-d,** Flow plots and quantification of group 1 ILCs in liver from ILCP (n = 10) (**c**) or ENKP-transferred (n = 10) (**d**) NSG mice combined from 3 independent experiments. **e,** Strategy of competitive assay and quantification of chimerism in spleen combined from 2 independent experiments. ALP group, n = 8; ENKP group, n = 9; ILCP group, n = 8. For **a, c-e**, data are mean ± s.e.m, and statistical analysis was performed using two-tailed unpaired Student's t-tests.

Figure 3 ENKPs develop independent of ILCPs

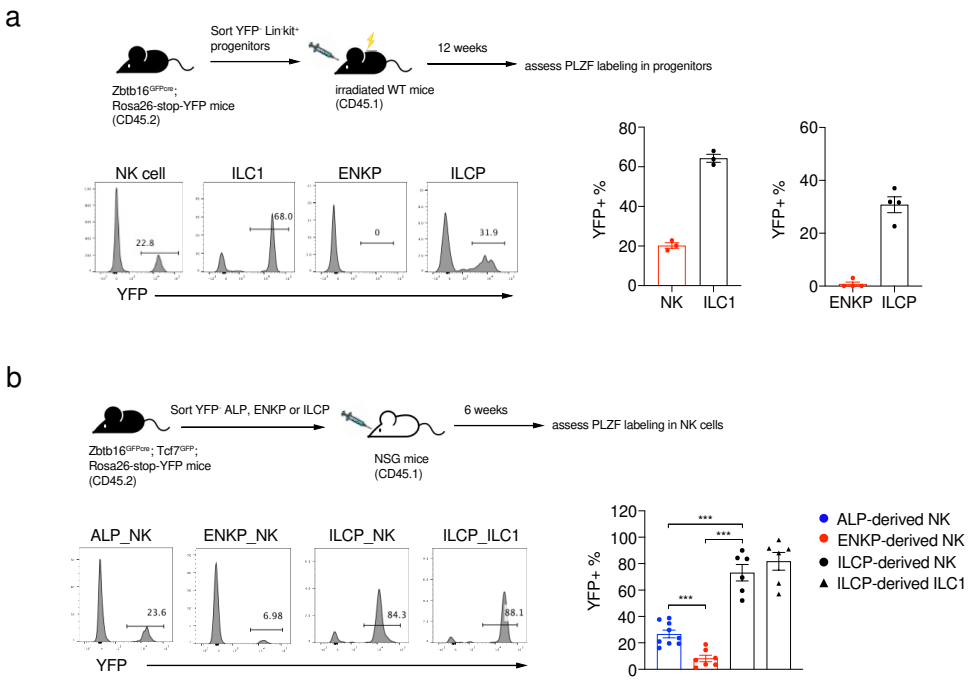


Fig. 3 ENKPs develop independent of ILCPs. a, Strategy of generating long-term chimeras for PLZF lineage tracing, representative histograms and quantification of group 1 ILCs (n = 3) in liver and PLZF-labeling ratios of progenitors (n = 4) in bone marrow combined from 2 independent experiments. **b,** Strategy of assessing PLZF lineage tracing in progenitor-derived NK cells, representative histograms and quantification of PLZF-labeling ratios of progenitor-derived NK cells in spleen and ILC1s in liver combined from 3 independent experiments. ALP-derived NK, n = 9; ENKP-derived NK, n = 7; ILCP-derived NK, n = 6; ILCP-derived ILC1, n = 6. Data are mean ± s.e.m, and statistical analysis was performed using two-tailed unpaired Student's t-test.

Figure 4 ENKP and ILCP give rise to distinct NK cells

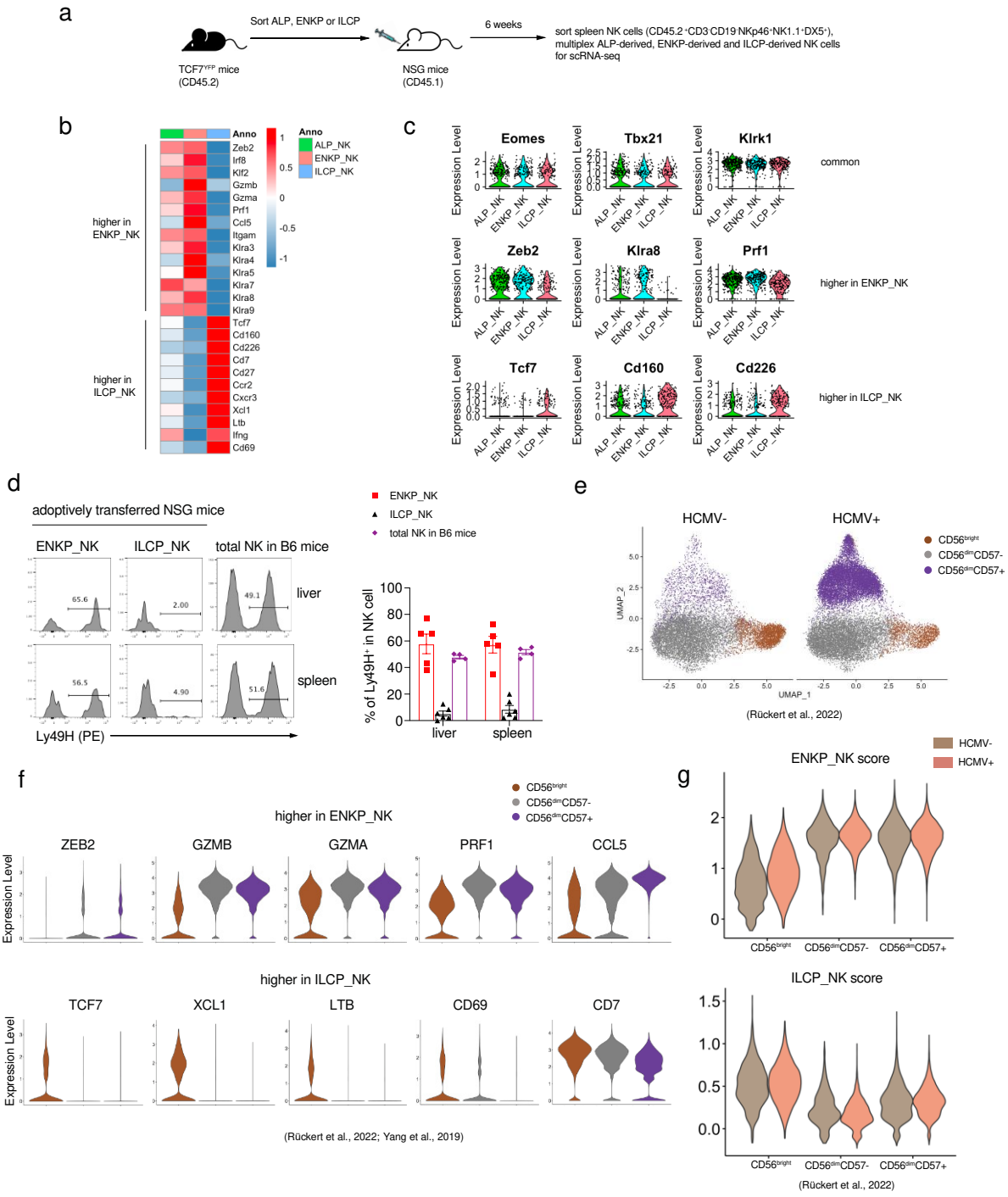


Fig. 4 ENKP and ILCP give rise to distinct NK cells. a, Strategy of scRNA-seq of NK cells derived from different progenitors. **b,** Heatmap of average expression levels of representative genes differentially expressed in ENKP_NK and ILCP_NK from scRNA-seq. Each column is one developmental origin. Data are shown as z-score, with high expression in red and low expression in blue. **c,** Violin plots grouped by developmental origins show representative

genes expressed similarly (upper) or differently (middle, higher in ENKP_NK; lower, higher in ILCP_NK) between ENKP_NK and ILCP_NK. **d,** Representative histograms, and quantification of $Ly49H^+$ ratios in NK cells derived from ENKPs or ILCPs combined from 2 independent experiments. “Total NK” group is total NK cells from wild-type C57BL/6 mice. ENKP_NK, n = 5; ILCP_NK, n = 6; total NK, n = 4. Data are mean ± s.e.m. **e**, UMAP of human blood NK cells split by HCMV status analyzed by scRNA-seq. Annotation defined by surface protein expression measured by CITE-seq. **f**, Violin plots of human blood NK cells visualizing homologous genes identified to be upregulated in either mouse ENKP_NK or ILCP_ NK cells, grouped according to the classification in Extended Data Fig. 6b. **g**, Violin plots of gene set scores calculated in the human single cell blood NK data using the human homologues of the genes upregulated in either ENKP or ILCP mouse NKs Plots split by HCMV status. For **d**, data are mean ± s.e.m. For **e**, **g**, data are from dataset of *Rückert et al., 2022*. For **f**, data are integrated from datasets of *Rückert et al., 2022* and *Yang et al., 2019*.

Figure 5 ENKP-derived NK cells respond to MCMV infection

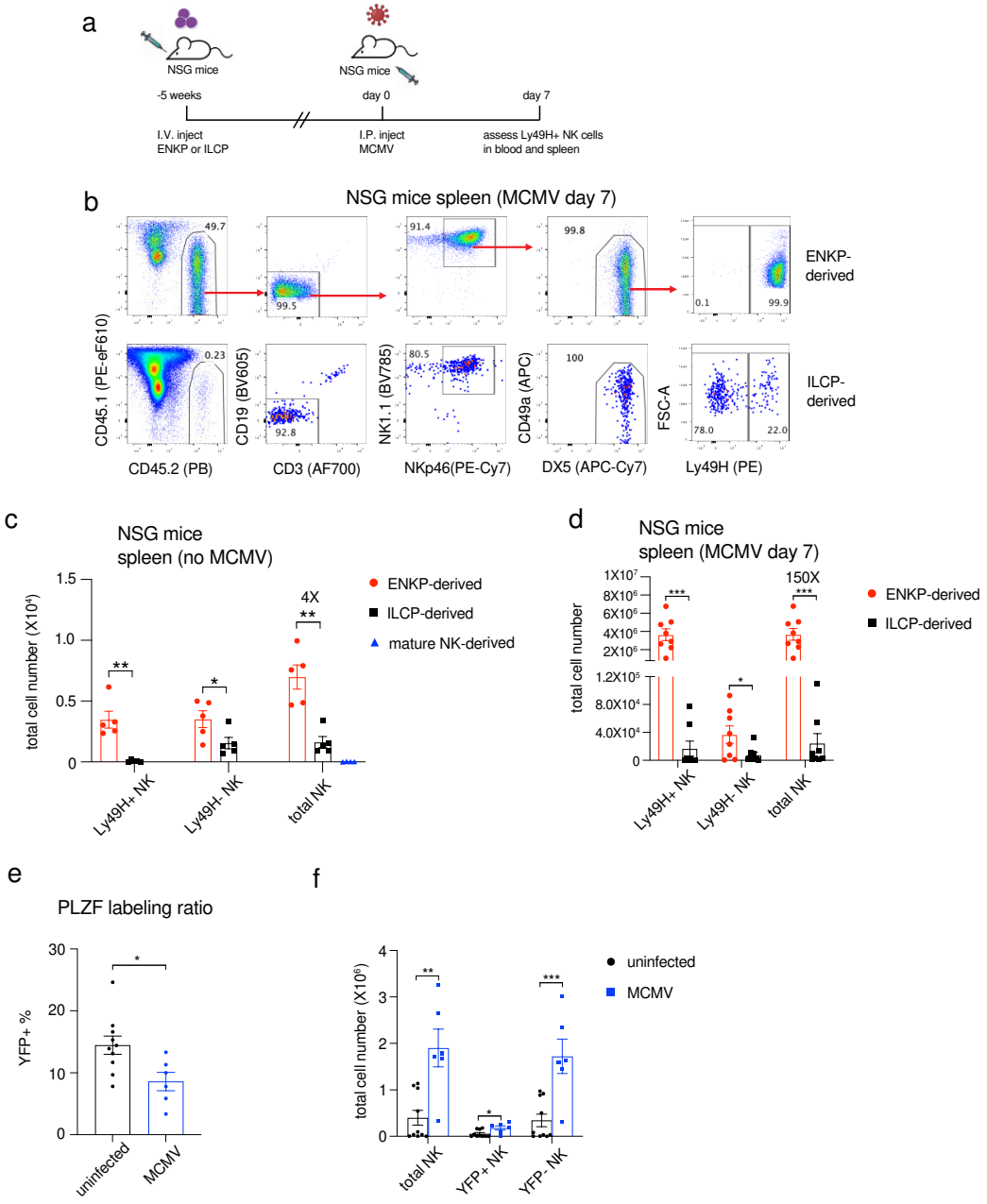


Fig. 5 ENKP-derived NK cells respond to MCMV infection. **a,** Strategy of assessing ENKP_NK cells and ILCP_NK cells for response to MCMV. **b, d**, Analysis of ENKP_NK cells and ILCP_NK cells from spleen on day 7 after MCMV infection. Representative flow plots (**b**), quantification of total cell numbers (n = 8) (**d**) of ENKP_NK cells and ILCP_NK cells combined from 2 independent experiments. 300 ENKPs or ILCPs were transferred into each NSG recipient mouse. **c,** Quantification of total numbers of ENKP_NK cells and ILCP_NK cells from NSG mice without MCMV infection, n = 5. Total numbers were normalized to 300 ENKPs, ILCPs or mature NK cells. Combined data from 2 independent experiments. **e**, Quantification of total NK cells analyzed in PLZF tracing long-

term chimeras on day 7 after MCMV infection combined from 2 independent experiments. control, n = 9; MCMV, n = 6. **f**, Quantification of NK subsets analyzed in PLZF tracing long-term chimeras on day 7 after MCMV infection combined from 2 independent experiments. control, n = 10; MCMV, n = 6. For **c** - **f**, data are mean ± s.e.m, and statistical analysis was performed using two-tailed unpaired Student's t-tests.

Figure 6　Characterization of functions of NK subsets

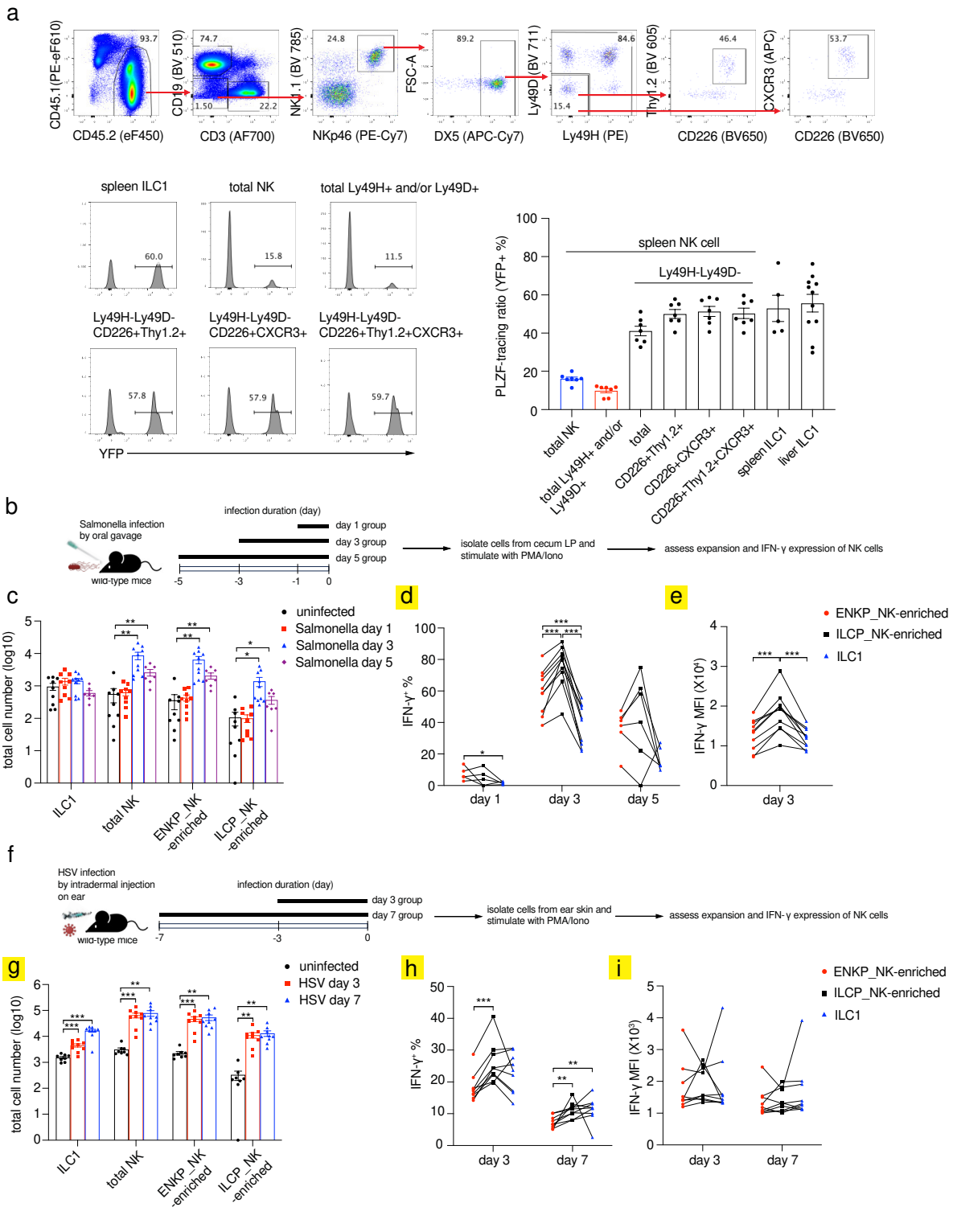
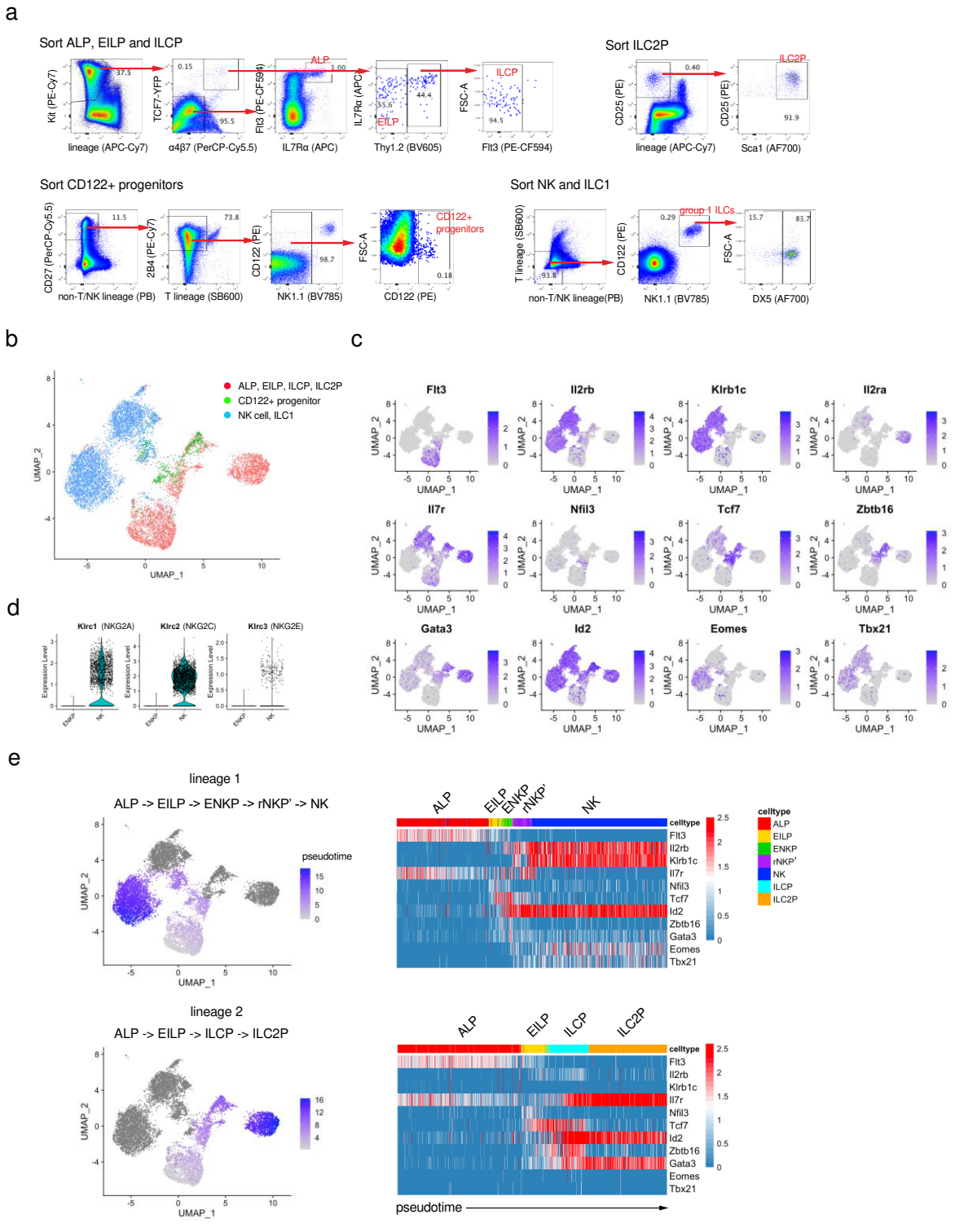


Fig. 6 Characterization of functions of NK subsets. a, Representative flow plots and histograms, and quantification of PLZF-tracing ratios in spleen NK subsets analyzed from irradiated NSG mice transferred with Lin^- Kit^+ progenitors (NSG PLZF-tracing long-term chimera) combined from 3 independent experiments. spleen NK cells, n = 7; spleen ILC1, n = 5; liver ILC1, n = 11. **b**, Strategy of assessing NK subsets for response to Salmonella infection. **c**, Quantification of number of NK cells and ILC1s in cecum LP after Salmonella infection combined from 2 independent experiments. control, n = 9; day 1, n = 9; day 3, n = 10; day 5, n = 7. **d - e**, Quantification of IFN-γ^+ ratios (**d**) and IFN-γ MFI (**e**) in NK subsets and ILC1s after Salmonella infection combined from 2 independent experiments. day 1, n = 4; day 3, n = 9; day 5, n = 6. **f**, Strategy of assessing NK subsets for response to HSV infection. **g**, Quantification of number of NK cells in ear skin after HSV infection combined from 2 independent experiments. control, n = 8; day 3, n = 9; day 5, n = 9. **h - i**, Quantification of IFN-γ^+ ratios (**h**) and IFN-γ MFI (**i**) in NK subsets and ILC1s after HSV infection combined from 2 independent experiments. day 3, n = 9; day 5, n = 9. For **c - e**, **g - i,** ENKP_NK-enriched population, $Ly49H^+$ and/or $Ly49D^+$; ILCP_NK-enriched population, $Ly49H^-$ $Ly49D^-CD226^+Thy1.2^+$. For **a, c - e**, **g - i,** data are mean ± s.e.m; for **c, g,** statistical analysis was performed using two-tailed unpaired Student's t-tests; for **d, e**, **h, i**, statistical analysis was performed using two-tailed paired Student's t-tests.

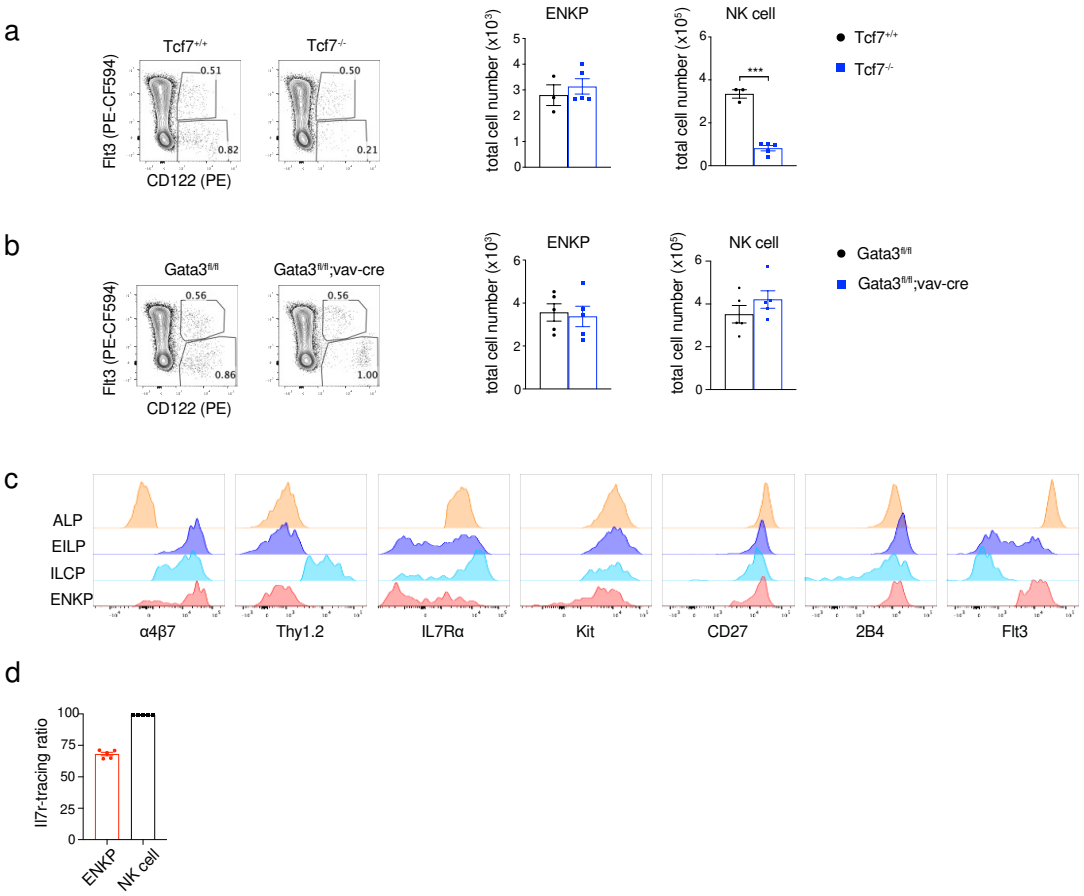
Extended Data Fig.1　Characterization of early NK cell development in bone marrow



Extended Data Fig. 1 Characterization of early NK cell development in bone marrow. a, Gating strategy for sorting ALP, EILP, ILCP, ILC2P, CD122^{+} progenitor, NK cell and ILC1 from bone marrow for scRNA-seq. **b,**

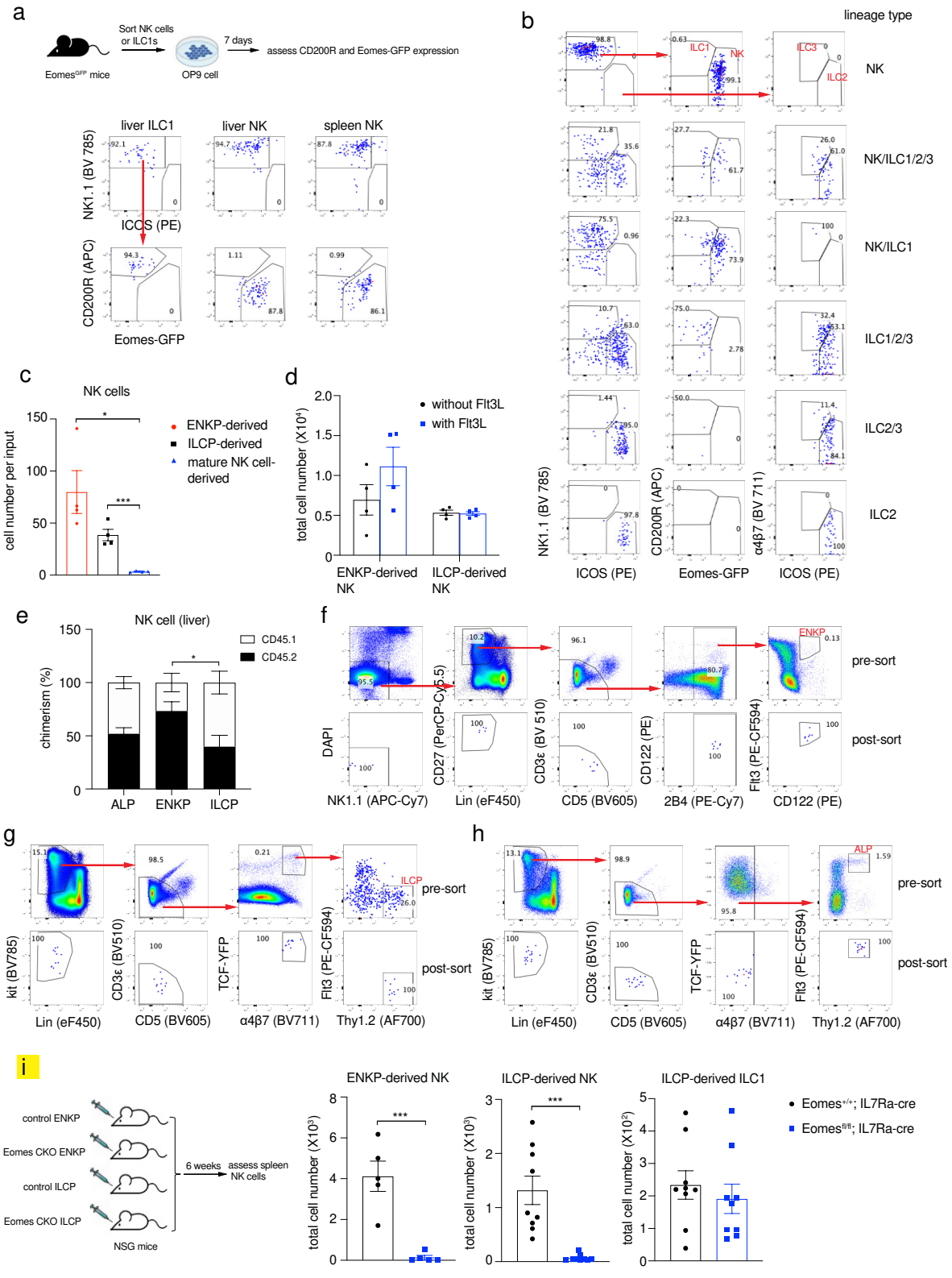
UMAP of scRNA-seq samples grouped by original idents. **c**, Feature plots of genes critical for identifying UMAP clusters. **d**, Violin plots of ENKP and NK cluster. **e**, Slingshot pseudotime and heatmap of gene expression ordered by pseudotime.

Extended Data Fig.2 Characterization of ENKP



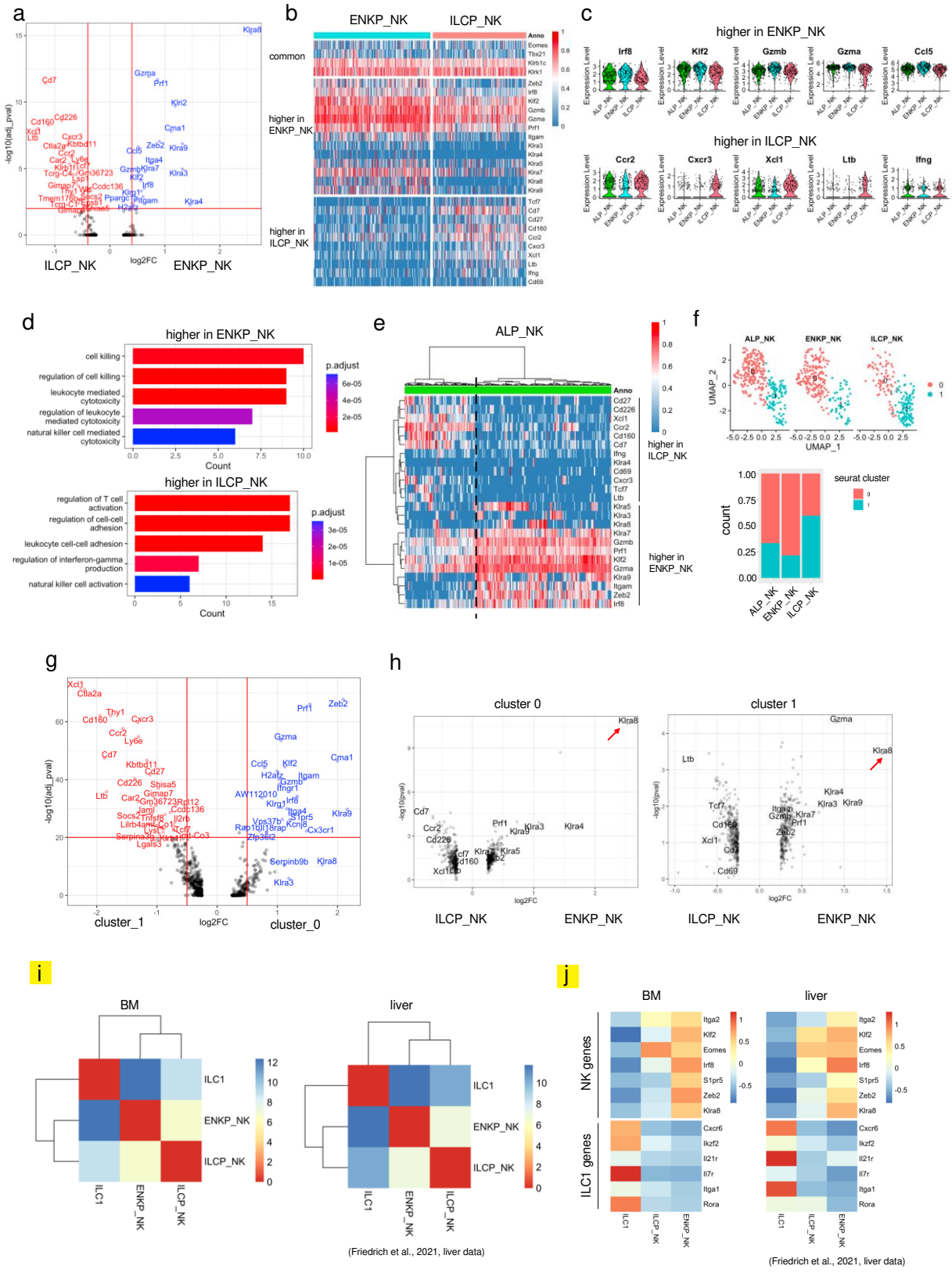
Extended Data Fig. 2 Characterization of ENKPs. a, Representative flow plots of ENKPs and quantifications of ENKPs and NK cells in bone marrow of *Tcf7*-deficient mice (n = 3) and control mice (n = 5). Combined data from 2 independent experiments. **b,** Representative flow plots of ENKPs and quantifications of ENKPs and NK cells in bone marrow of *Gata3*-deficient mice, n = 5. Combined data from 2 independent experiments. **c,** Representative histograms of surface marker expression by different progenitors from 2 independent experiments. **d**, Quantifications of *Il7r*-lineage tracing ratios in ENKPs and NK cells in bone marrow, n = 5. Representative data from 3 independent experiments. For **a**, **b** and **d**, data are mean ± s.e.m. For **a** and **b,** statistical analysis was performed using two-tailed unpaired Student's t-tests.

Extended Data Fig. 3 Assessment of lineage potential of ENKP



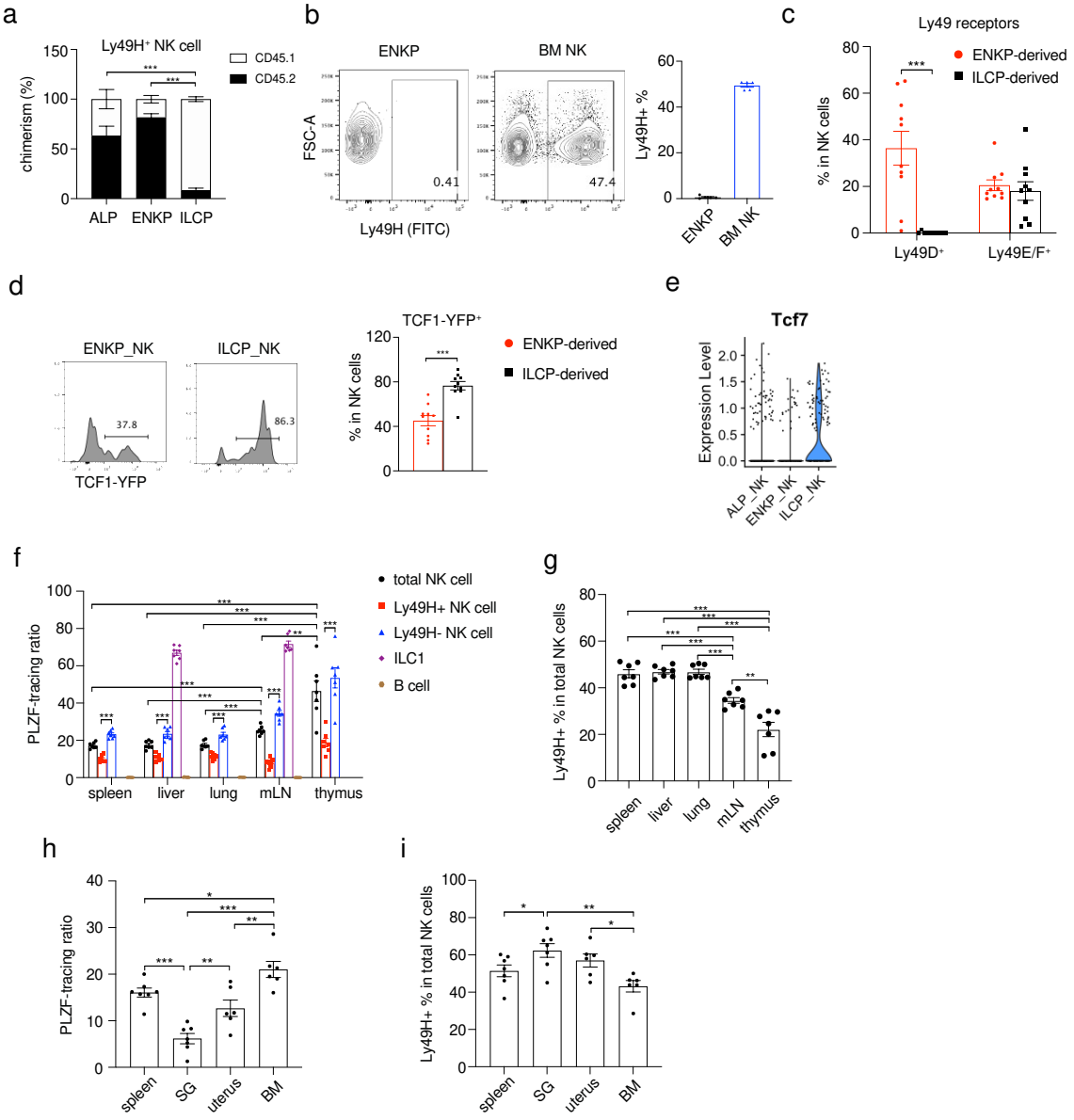
Extended Data Fig. 3 Assessment of lineage potential of ENKPs. a, Strategy and flow plots of in vitro cultured NK cells ($CD3^-CD19^-NKp46^+NK1.1^+CD49a^-DX5^+Eomes^{GFP+}$) and ILC1s ($CD3^-CD19^-NKp46^+NK1.1^+CD49a^+DX5^-Eomes^{GFP-}$) sorted from liver of $Eomes^{GFP}$ mice. **b**, Representative flow plots of in vitro clonal assay. **c**, Quantification of in vitro differentiated NK cells, comparing progenitors and mature NK cells, n = 4. NK cells were defined as $NK1.1^+Eomes^{GFP+}CD220R^-$. Representative data from 2 independent experiments. **d**, Quantification of in vitro differentiated NK cells, comparing with and without Flt3L, n = 4. NK cells were defined as $NK1.1^+Eomes^{GFP+}CD220R^-$. Representative data from 2 independent experiments. **e**, Quantification of chimerism after competitive assay in liver combined from 2 independent experiments. ALP group, n = 5; ENKP group, n = 5; ILCP group, n = 4. **f - h**, Representative flow plots of post-sort purity check for ENKP (**f**), ILCP (**g**) and ALP (**h**). **i**, Strategy and quantification of total spleen NK cells and ILC1s derived from control or Eomes-deficient progenitors combined from 2 independent experiments. ENKP-derived NK, n = 5; ILCP-derived NK, n = 9. For **c**, **d**, **e**, **i**, data are mean ± s.e.m., and statistical analysis was performed using two-tailed unpaired Student's t-tests.

Extended Data Fig. 4　ENKP_NK and ILCP_NK are transcriptionally different



Extended Data Fig. 4 ENKP_NK and ILCP_NK are transcriptionally different. a, Volcano plot of differentially expressed genes between ENKP_NK cells and ILCP_NK cells. **b,** Heatmap of genes equally expressed (upper) and differentially expressed between ENKP_NK and ILCP_NK (middle and lower). Each column represents one cell. **c,** Violin plots of genes expressed higher in ENKP_NK (upper) and higher in ILCP_NK (lower) grouped by developmental origins. **d,** GO enrichment analysis using genes expressed higher in ENKP_NK cells (upper) or ILCP_NK cells (lower). **e,** Hierarchically clustered heatmap of gene expression levels of genes differentially expressed between ENKP_NK cells and ILCP_NK cells in ALP_NK cells. Each column represents one cell. Dashed line indicates separation between two major column clusters. **f,** UMAP and bar plot of scRNA-seq of progenitor-derived NK cells grouped by developmental origins. **g,** Volcano plot of differentially expressed genes between cluster 0 and cluster 1. **h,** Volcano plot of differentially expressed genes between ENKP_NK and ILCP_NK in cluster 0 (left) and in cluster 1 (right). Arrow highlights the gene "*Klra8*". **i**, Hierarchically clustered Euclidean distance matrix of transcriptomes of NK cells and ILC1s in bone marrow (left) and liver (right). **j**, Heatmap of select genes (see Methods) differentially expressed between NK cells and ILC1s in ENKP_NK cells, ILCP_NK cells and ILC1s in bone marrow (left) and liver (right).

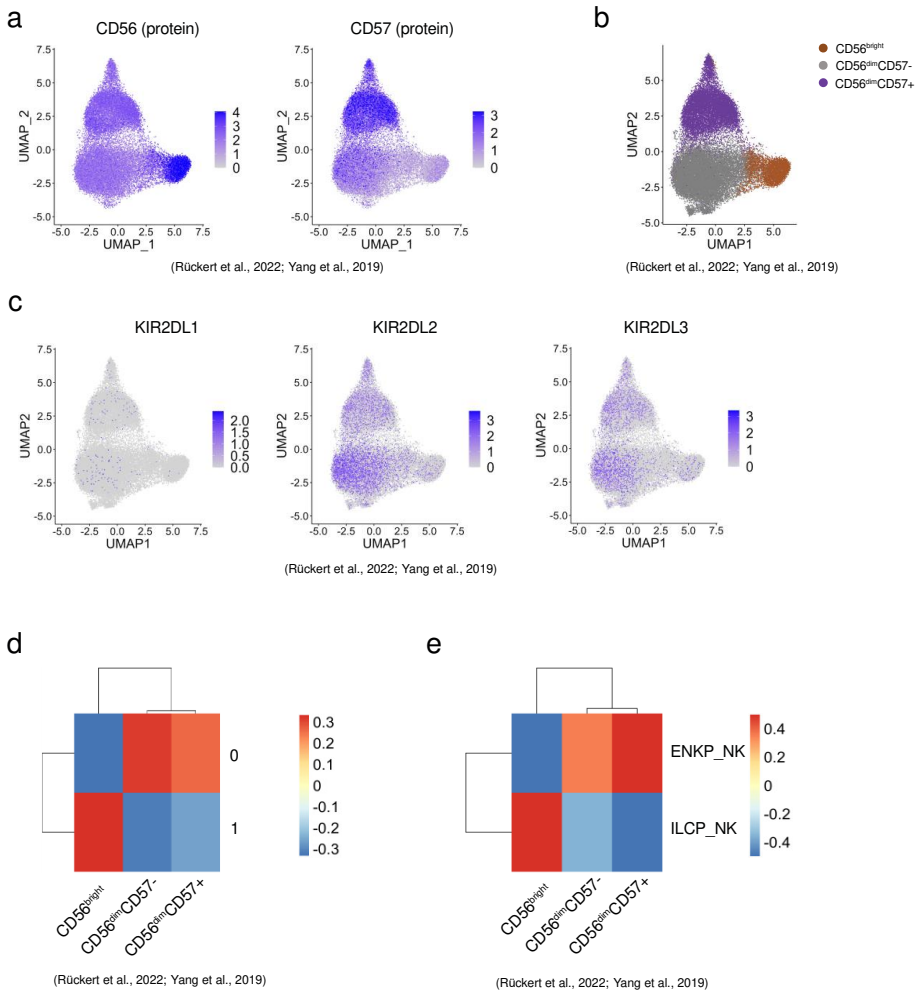
Extended Data Fig. 5 ENKP and ILCP give rise to distinct NK cells



Extended Data Fig. 5 ENKPs and ILCPs give rise to distinct NK cells. a, Quantification of chimerism in competitive assay (same strategy as in **Fig. 2e**) from 2 independent experiments. ALP group, n = 8; ENKP group, n = 9; ILCP group, n = 8. **b**, Representative flow plots, and quantification of Ly49H$^+$ ratios in ENKPs and NK cells in bone marrow combined from 2 independent experiments, n = 5. **c**, Quantification of Ly49 receptor positive ratios in ENKP_NK cells and ILCP_NK cells from NSG mice transferred with ENKPs or ILCPs, n = 10 combined from 2 independent experiments. **d**, Representative histogram of TCF1-YFP expression and quantification of TCF1-YFP$^+$ ratios in ENKP_NK cells and ILCP_NK cells from NSG mice transferred with ENKPs or ILCPs, n = 10 combined

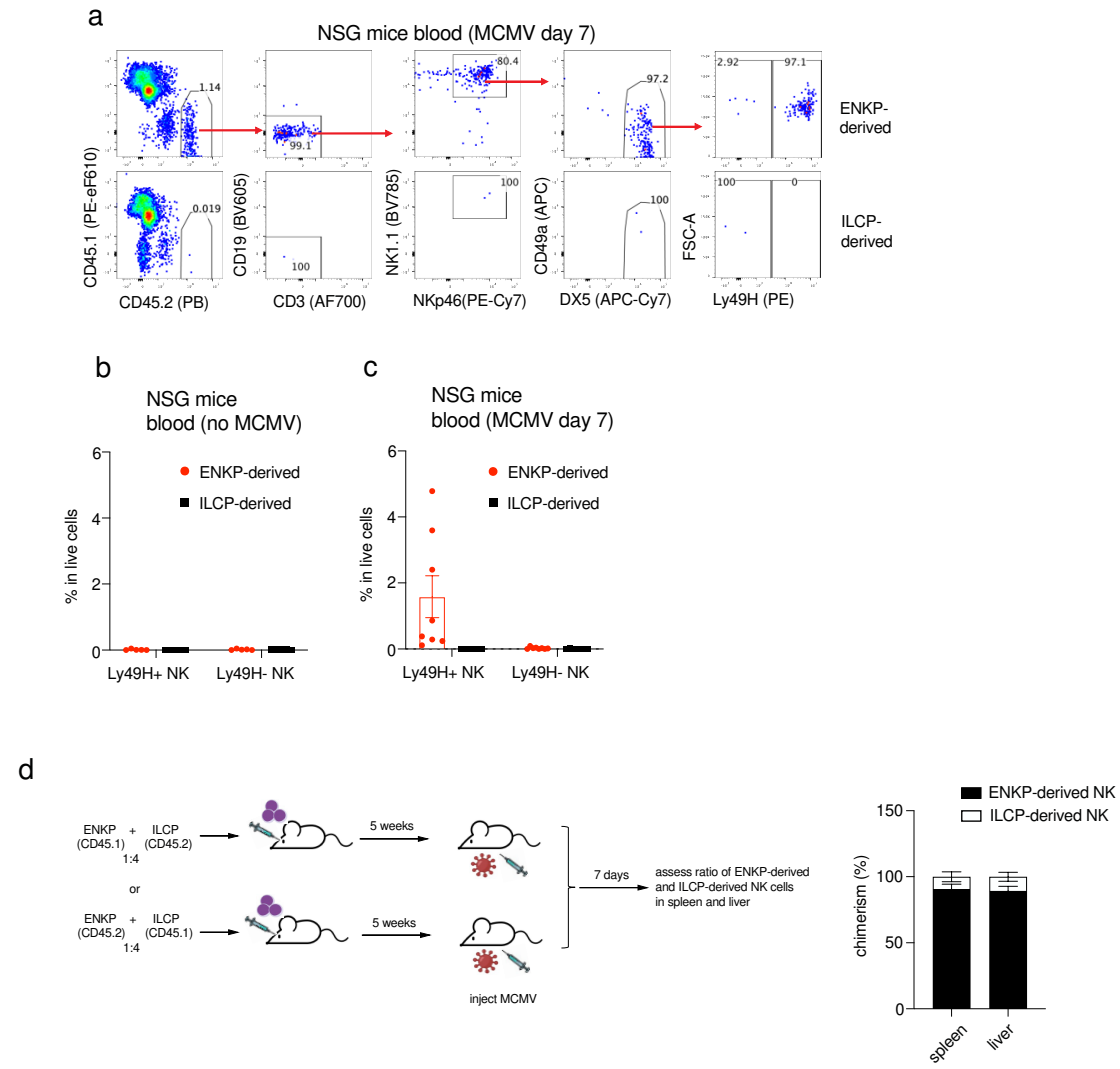
from 2 independent experiments. **e**, Violin plots of *Tcf7* from scRNA-seq grouped by developmental origins. **f**, Quantification of PLZF lineage tracing ratios in different tissues from irradiated wild-type mice transferred with Lin^-Kit^+ progenitors (PLZF-tracing long-term chimera), n = 7 combined from 2 independent experiments. B cells were used as negative control for PLZF labeling. **g,** Quantification of $Ly49H^+$ ratios in NK cells in different tissues from PLZF-tracing long-term chimeras, n = 7 combined from 2 independent experiments. **h - i**, Quantification of PLZF lineage tracing ratios (**h**) and $Ly49H^+$ ratios in NK cells (**i**) in different tissues from irradiated NSG mice transferred with Lin^-Kit^+ progenitors (NSG PLZF-tracing long-term chimera) combined from 3 independent experiments. spleen and salivary glands (SG), n = 7; uterus and bone marrow (BM), n = 6. For **a-d**, **f - i**, data are mean ± s.e.m. For **a**,**c**,**d**, **f - i**, statistical analysis was performed using two-tailed unpaired Student's t-tests.

Extended Data Fig. 6 ENKP-derived and ILCP-derived NK cells are similar to human CD56^{dim} and CD56^{bright} NK cells, respectively



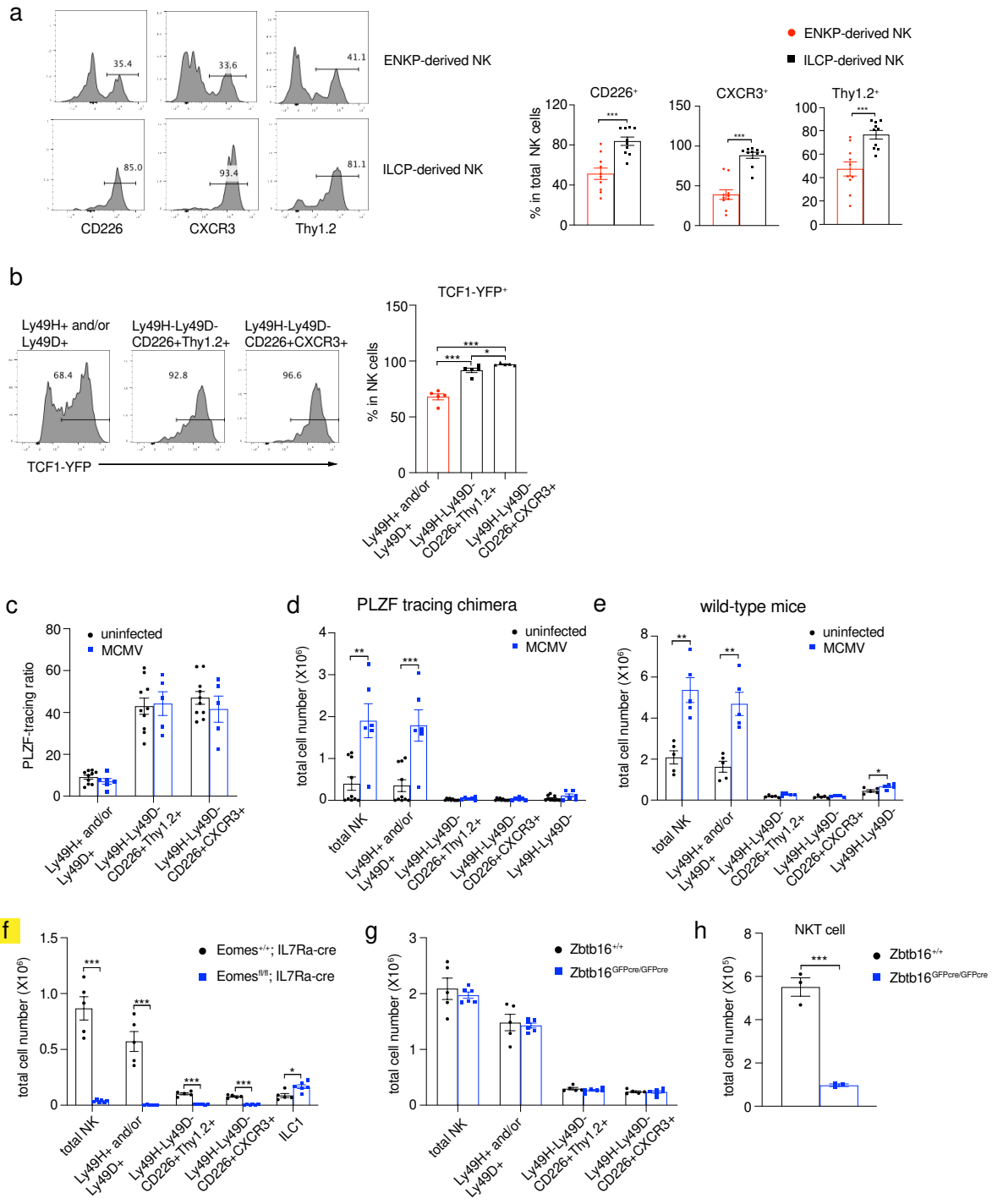
Extended Data Fig. 6 ENKP-derived and ILCP-derived NK cells are similar to human CD56 dim and CD56 bright NK cells, respectively. a. Feature plots of surface protein expression in human blood NK measured by CITE-seq. **b**. UMAP of human blood NK cells analyzed by scRNA-seq. Annotation defined by surface protein expression measured by CITE-seq. **c**. Feature plots of KIR genes in human blood NK cells analyzed by scRNA-seq. **d**. Heatmap of Spearman correlation analysis between mouse NK clusters and human NK clusters. **e**. Heatmap of Spearman correlation analysis between mouse NK developmental originals and human NK clusters. For **a** - **e**, data were integrated from datasets of *Rückert et al., 2022* and *Yang et al., 2019*.

Extended Data Fig. 7 ENKP-derived NK cells respond to MCMV infection



Extended Data Fig. 7 ENKP-derived NK cells respond to MCMV infection. a-b, Analysis of ENKP_NK cells and ILCP_NK cells in peripheral blood on day 7 after MCMV infection. Representative flow plots (**a**) and quantification of frequencies (**c**) of ENKP_NK cells and ILCP_NK cells, n = 8 combined from 2 independent experiments. **b**, Quantification of frequencies of ENKP_NK and ILCP_NK cells in blood without MCMV infection, n = 5 combined from 2 independent experiments. For **b** and **c**, 300 ENKPs or ILCPs were transferred into each NSG mice. **d**, Strategy and quantification of ratios of ENKP and ILCP-derived NK cells day 7 after MCMV infection combined from 2 independent experiments. n = 9. For **b** - **d**, data are mean ± s.e.m.

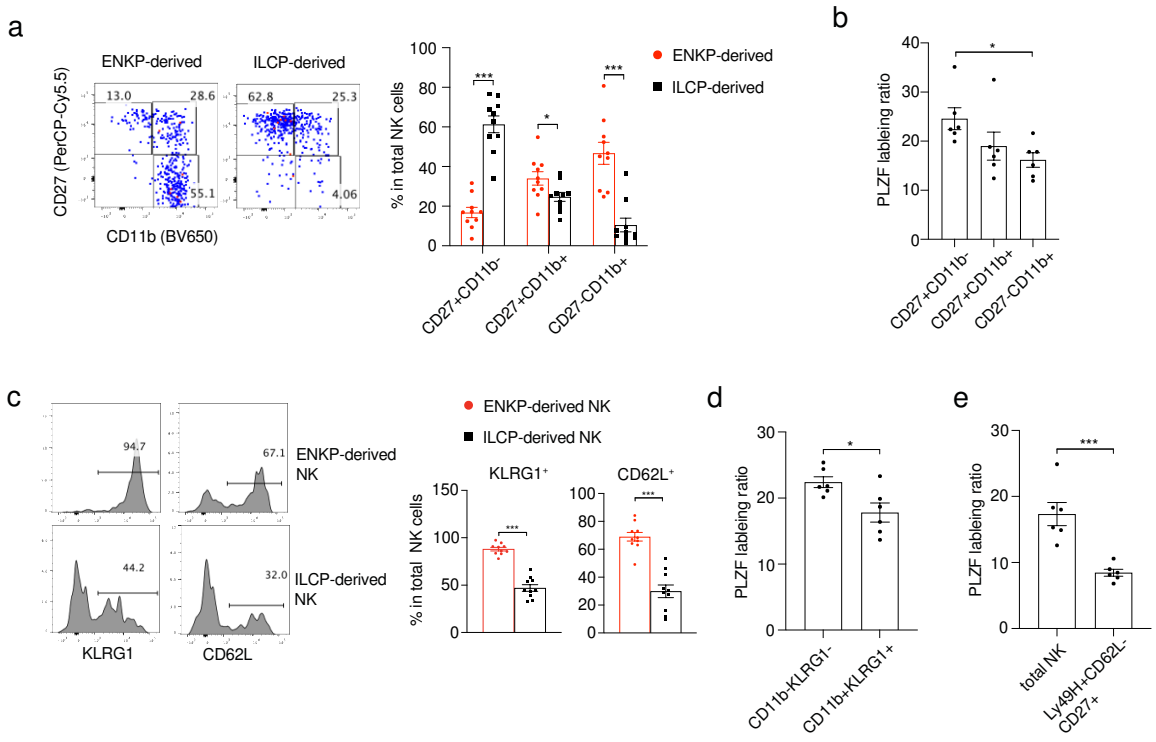
Extended Data Fig. 8　Verifcation of ENKP_NK and ILCP_NK enriched populations



Extended Data Fig. 8 Verification of ENKP_NK and ILCP_NK-enriched cell population. a, Representative histograms, and quantification of frequencies of cells expressing specific surface markers in spleen NK cells derived from ENKPs or ILCPs combined from 2 independent experiments. n = 10. **b**, Representative histograms and

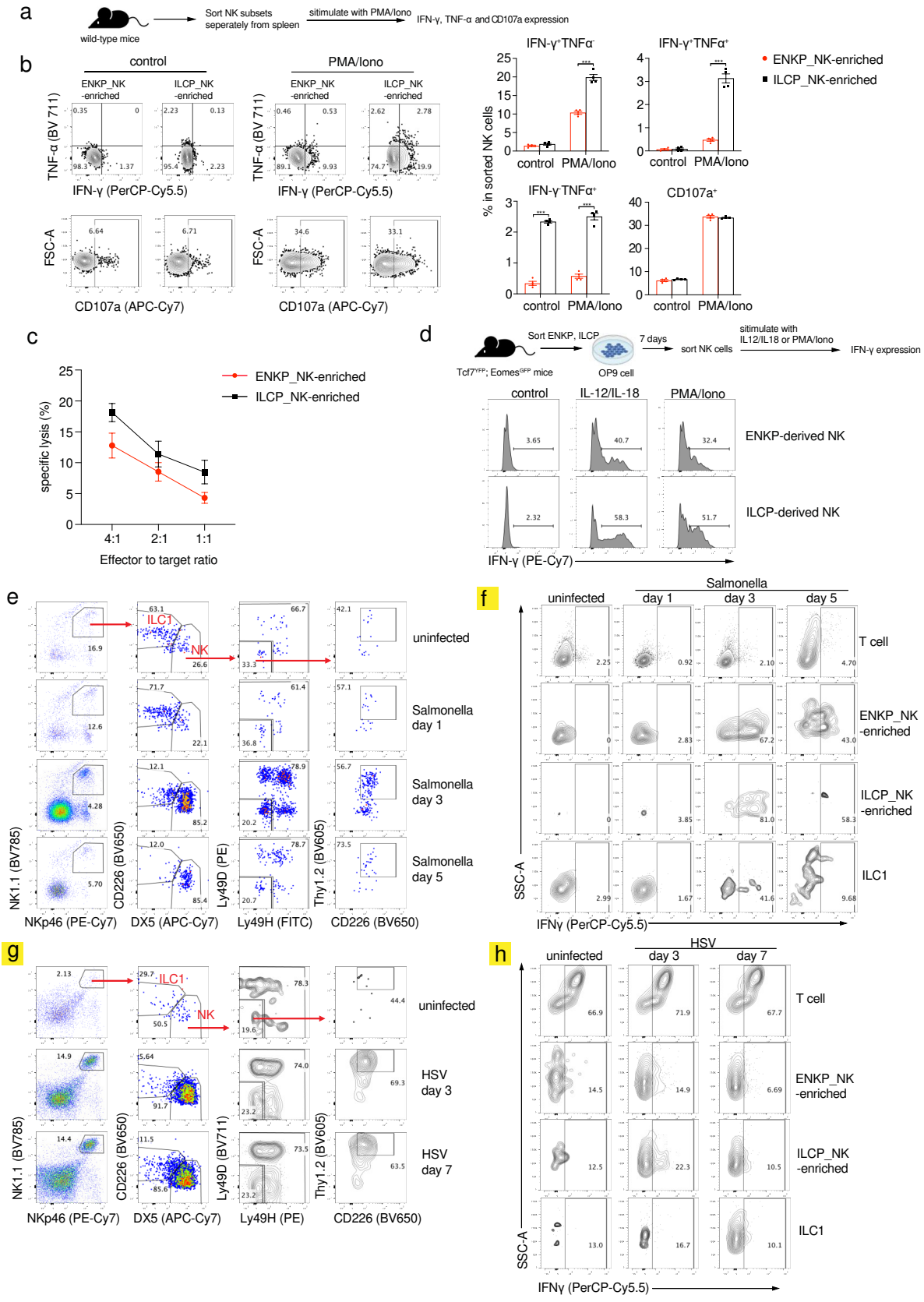
quantification of frequencies of cells expressing TCF1-YFP in spleen NK subsets from TCF1-YFP reporter mice combined from 2 independent experiments. n = 5. **c** - **d**, Quantification of PLZF traced ratios (**c**) and NK cell numbers (**d**) in spleen NK subsets from NSG PLZF-tracing long-term chimera day 7 after MCMV infection combined from 2 independent experiments. control, n = 10; MCMV, n = 5. **e**, Quantification of NK cell numbers in spleen NK subsets in wild-type B6 mice on day 7 after MCMV infection combined from 2 independent experiments. n = 5. **f**, Quantification of cell number in spleen of total NK cells, NK subsets and ILC1s in *Eomes*-deficient mice combined from 2 independent experiments. $Eomes^{+/+}$; IL7Ra-cre, n = 5; $Eomes^{fl/fl}$; IL7Ra-cre, n = 6. **g**, Quantification of cell number in spleen of total NK cells and NK subsets in *Zbtb16*-deficient mice combined from 2 independent experiments. $Zbtb16^{+/+}$, n = 5; $Zbtb16^{GFPcre/GFPcre}$, n = 6. **h**, Quantification of cell numbers of spleen NKT cells in *Zbtb16*-deficient mice combined 2 independent experiments. n = 3. For **a** - **h**, data are mean ± s.e.m., and statistical analysis was performed using two-tailed unpaired Student's t-tests.

Extended Data Fig. 9 Distinguish ENKP and ILCP-derived NK cells by surface markers



Extended Data Fig. 9 Distinguish ENKP_NK cells and ILCP_NK cells by surface markers. a, Representative flow plots, and quantification of frequencies of NK subsets in spleen NK cells derived from ENKPs or ILCPs combined from 2 independent experiments. n = 10. **b**, Quantification of PLZF-tracing ratios in spleen NK subsets from NSG PLZF-tracing long-term chimeras combined from 2 independent experiments. n = 6. **c**, Representative histograms, and quantification of frequencies of cells expressing specific surface markers in spleen NK cells derived from ENKPs or ILCPs combined from 2 independent experiments. n = 10. **d**, Quantification of PLZF-tracing ratios in spleen NK subsets from NSG PLZF-tracing long-term chimeras from 2 independent experiments. n = 6. **e**, Quantification of PLZF-tracing ratios in spleen NK subsets from NSG PLZF-tracing long-term chimeras combined from 2 independent experiments. n = 6. For **a** - **e**, data are mean ± s.e.m., and statistical analysis was performed using two-tailed unpaired Student's t-tests.

Extended Data Fig.10 Characterization of functions of NK subsets



Extended Data Fig. 10 Characterization of functions of NK subsets. a, Strategy of assessing cytokine expression in sorted NK subsets. **b**, Representative flow plots, and quantification of frequencies of cells expressing cytokines in NK subsets, representative data from 2 independent experiments. n = 4. **c**, Quantification of specific lysis ratios of Yac-1 cells co-cultured with NK subsets, representative data from 2 independent experiments. n = 4. **d**, Strategy of assessing IFN-γ expression of ENKP_NK and ILCP_NK after stimulation in vitro, and representative histograms of IFN-γ expression levels from 2 independent experiments. **e - f,** Representative flow plots of gating (**e**) and IFN-γ expression (**f**) of NK subsets and ILC1s from cecum lamina propria (LP) after Salmonella infection. **g - h,** Representative flow plots of gating (**g**) and IFN-γ expression (**h**) of NK subsets and ILC1s from ear skin after HSV infection. For **b**, **c**, ENKP_NK-enriched population, $Ly49H^+$ and/or $Ly49D^+$; ILCP_NK-enriched population, $Ly49H^-Ly49D^-CD226^+$($Thy1.2^+$ and/or $CXCR3^+$). For **f**, **h**, ENKP_NK-enriched population, $Ly49H^+$ and/or $Ly49D^+$; ILCP_NK-enriched population, $Ly49H^-Ly49D^-CD226^+Thy1.2^+$. For **b - c**, data are mean ± s.e.m, and statistical analysis was performed using two-tailed unpaired Student's t-tests.