A multicolor map of NK and T cell diversity: from bulk to single-cell

Melsen, J.E.

Chapter 7

Single-cell transcriptomics of bone marrow delineates CD56^{dim}GranzymeK^{+} subset as intermediate stage in NK cell differentiation

Janine E. Melsen, Monique M. van Ostaijen-ten Dam, Dorenda J.A. Schoorl, Pieter J. Schol, Daphne A.L. van den Homberg, Arjan C. Lankester, Gertjan Lugthart, Marco W. Schilham

In preparation

Abstract

Human natural killer (NK) cells in lymphoid tissues can be categorized into three subsets: $CD56^{bright}CD16^{+}$, $CD56^{dim}CD16^{+}$ and $CD69^{+}CXCR6^{+}$ lymphoid tissue-resident (lt)NK cells. How the three subsets are functionally and developmentally related is currently unknown. Therefore, we performed single-cell RNA sequencing combined with oligonucleotide-conjugated antibodies against CD56, CXCR6, CD117 and CD34 on fresh bone marrow NK cells. A minor $CD56^{dim}GzmK^{+}$ subset shared features with $CD56^{bright}$ and $CD56^{dim}GzmK^{-}$ NK cells based on transcriptome, phenotype ($NKG2A^{high}CD16^{low}KLRG1^{high}TIGIT^{high}$) and functional analysis in bone marrow and blood, supportive for an intermediate subset. Pseudotime analysis positioned $CD56^{bright}$, $CD56^{dim}GzmK^{+}$ and $CD56^{dim}GzmK^{-}$ cells in one differentiation trajectory, while ltNK cells were separated in a distinct lineage. Integrative analysis with bone marrow cells did not demonstrate a developmental connection between $CD34^{+}$ progenitor and NK cells, implicating that bone marrow is not a site of active NK cell development. In conclusion, single-cell transcriptomics provide new insights on development and differentiation of human NK cells.

Introduction

Natural killer (NK) cells are innate immune cells, known for their cytotoxicity and effector molecule production upon target cell recognition and/or stimulation by interleukins.[12,50,58,267] NK cells can be discriminated from the non-cytotoxic innate lymphoid cells (ILCs) based on expression of perforin and Eomes.[107,108,268] Human NK cells in peripheral blood can be categorized into two subsets: $CD56^{bright}CD16^{+/-}$ and $CD56^{dim}CD16^{+}$. In several lymphoid (bone marrow[56,58], lymph node[56,58], spleen[56,58], tonsil[133]) and non-lymphoid tissues (lung[56,269], intestines[56], uterus[109,112], liver[54,113,135,183,184,270]) a third NK cell subset has been described: tissue-resident NK cells.[177] The lymphoid tissue-resident (lt)NK cells, and liver-resident NK cells are characterized by the combined expression of CD69 and CXCR6, and absence of the integrin CD49e.[58,260] The integrins CD103 and CD49a are exclusively expressed by mucosal tissue-resident NK cells.[56,112,133,269]

The functional capacity of the $CD56^{bright}$ and $CD56^{dim}$ NK cells is dependent on the type of stimulation. $CD56^{dim}$ NK cells are potent cytokine and chemokine producers and become highly cytotoxic upon target cell stimulation[12,50] In contrast, the $CD56^{bright}$ NK cells require interleukin activation to kill target cells and produce cytokines.[271] Although, ltNK cells have the highest mRNA expression of *IFNG*, *CCL3*, *CCL4*, *CCL5, XCL1* and *XCL2* compared to both $CD56^{bright}$ and $CD56^{dim}$ NK cells as studied by bulk RNA sequencing, standard NK cell function assays (PMA/ionomycin, interleukin or K562 stimulation) did not reveal abundant production of IFN-γ at protein level.[58,260] Moreover, they are less cytotoxic compared to $CD56^{dim}$ NK cells. Heterogeneity within the ltNK cell population or requirement of yet unidentified stimuli might explain the unresponsiveness of the majority of ltNK cells.

It is assumed that the $CD56^{bright}$ NK cells are the precursors of the $CD56^{dim}$ NK cells based on telomere length[44], reconstitution following hematopoietic stem cell transplantation[47] and in vitro differentiation studies[44,46,48]. Following in vitro differentiation with cytokines a gain of CD16, KIRs, and loss of IL-7Rα (CD127), CD117, CXCR3, and CCR7 was observed.[44,46] However, whether a cytokine-induced effect in vitro is representative for biological differentiation is questionable. The $CD56^{bright}CD16^{+}$, $CD56^{bright}CD27^{-}$, $CD56^{dim}CD94^{high}$, $CD56^{dim}CD62L^{+}$ and $CD56^{dim}CD16^{dim}$ populations have been independently proposed to represent the intermediate subset based on phenotype and function.[46,140,172,173,175] Nevertheless, the detailed sequential steps during differentiation and overlap of these populations remain unclear. In addition, the developmental relation between the ltNK cells, $CD56^{bright}$ and $CD56^{dim}$ NK cells is still unknown. Hypothetically, ltNK cells could either constitute a separate lineage, or could be developed from either the $CD56^{bright}$ or $CD56^{dim}$ NK cells.

Bulk RNA sequencing does not unveil the heterogeneity within the NK cell populations, and therefore single-cell RNA sequencing provides a valuable tool to study both functional characteristics and development of the $CD56^{bright}$, $CD56^{dim}$ and ltNK cells. Multiple single-cell RNA sequencing datasets on blood[263,272–275] and bone marrow[262,272] NK cells have been published.[276] However, the mRNA expression of some important markers used to identify the known NK cell

populations by cytometry is around the lower level of detection for this technology. Therefore we applied the CITE-seq technology, using oligonucleotide-conjugated antibodies for CD56, CXCR6, CD117 and CD34. We demonstrate considerable heterogeneity within the $CD56^{dim}$ population, and identify a $CD56^{dim}GzmK^{+}$ NK cells subset as intermediate differentiation stage between $CD56^{bright}$ and $CD56^{dim}GzmK^{-}GzmB^{+}$ NK cells. In addition, we propose that resident ltNK cells develop independently from the circulating NK cells and found no evidence of NK cell development in the human bone marrow.

Materials and methods

Ethics statement

With approval of the Institutional Review Board (protocols P00.068, P01.028, B17.001 and LUMC healthy voluntary donor service (LuVDS)) and informed consent, blood (fresh or frozen) and bone marrow (fresh or frozen) from healthy controls and one hematopoietic stem cell transplant recipient were analyzed. Fresh blood (n=4) have been collected by the LuVDS, coordinated by the central biobanking facility at the LUMC. Residual fresh bone marrow from one healthy donor was used for single-cell RNA sequencing. Residual splenic tissue from Dutch solid organ transplant donors were used anonymously, in accordance with the Dutch law for organ donation. Tonsils and omental lymph nodes were collected as leftover material from tonsillectomy and bariatric surgery, respectively. Results were evaluated anonymously in accordance with Dutch national ethical and professional guidelines (http://www. federa.org).

Preparation of NK-enriched bone marrow cells

Bone marrow mononuclear cells (BMMC) were isolated by Ficoll density gradient centrifugation (LUMC Pharmacy, Leiden, The Netherlands). Untouched NK cell enrichment was performed by using Mojosort magnetic cell separation (Biolegend, San Diego, CA, USA), according to manufacturer's instructions. Anti-CD34 was not included in this kit, enabling enrichment of NK progenitor cells. Purity of the population was assessed by flow cytometry (Figure S1A). The antibodies used are listed in Table S1. Enriched cells were incubated with Fc block (eBioscience, San Diego, CA USA), after which cells were labeled with 1µg/ml oligonucleotide conjugated antibodies specific for CD34, CD56, CD117 and CXCR6 (TotalSeq-A, Biolegend, Table S2). The labeling was confirmed on a fraction of the cells by a secondary staining with goat-anti-mouse APC (Becton Dickinson (BD), Franklin Lakes, NJ, USA) (Figure S1A). Data was acquired on a LSR-II flow cytometer (BD).

Single-cell RNA sequencing data acquisition

The NK-enriched cell suspension was loaded on a Chromium Single Cell Chip (10x Genomics, Pleasanton, CA, USA) to encapsulate 10,000 cells with barcoded beads. Library preparation of the mRNA and cell surface bounded antibodies was performed by using the Single Cell 3' solution v3 (10x Genomics), according to manufacturer's instructions. The libraries were pooled, and sequencing was performed on one lane of the Illumina NovaSeq 6000 system (Illumina, San Diego, CA, USA).

Single-cell RNA sequencing data analysis

Cell ranger (software v3.0.2, 10x Genomics) was used to align the sequences to the human genome (hg38) and the antibodies. Barcodes associated with cells were selected based on the distribution of barcode counts and number of UMI counts mapped to each barcode (knee plot). The results from a total of 7000 cells were exported for further analysis in R (v4.0, R Foundation for Statistical Computing, Vienna, Austria) using the Bioconductor workflow as guide.[277] The DropletUtils package[278] was used to import the sequence data as SingleCellExperiment. Quality control was performed by using the Scater package[279]. Low-quality cells with <1000 expressed genes, >8300 expressed genes and >12,5% mitochondrial RNA, were removed (Figure S1B). No additional doublets were detected by using the doubletCluster function from Scran.[280] In total, 6979 cells, 33538 genes and 4 antibodies were included in subsequent analyses. Normalization of the antibody data was performed by using the centered log ratio, as implemented by Seurat[281]. The gene expression data was log-transformed and normalized by using deconvolution size factors, as implemented by Scran. The top 2000 most variable genes were selected by computing the variance of the log-counts and fitting a trend to the variance with respect to abundance across the genes (Scran) (Figure S1C). Principal component analysis (PCA) was applied and the top 20 PCs were retained (Figure S1D). Graph-based infomap clustering (k=30, type=rank) was performed by Igraph[282]. To visualize the clusters and expression data, a UMAP (n_neighbors=30, min.dist=1.0) embedding was calculated. Heatmaps were generated using the pheatmap package.[283] The cell-cycle score was calculated using the cyclone function in Scran. Subclustering of populations was performed by recalculating variable features and principle components. Pseudotime analysis was performed using Slingshot (omega=1.3).[240] The starting clusters were manually determined after lineage identification. To determine which genes change their expression over pseudotime a negative binomial general additive model (GAM) was fitted using the Tradeseq package[284].

Integration of public single-cell RNA sequence NK cell datasets

The single-cell RNA sequencing datasets (GSE130430) of Yang et al.[272] containing 9367 cells were aligned and aggregated (without normalization) using Cell ranger (v3.1.0). Cells that expressed less than 200 genes or more than 6000 genes were removed. In addition, cells containing less than 10% ribosomal protein coding genes or more than 10% mitochondrial genes were filtered out. The Human Cell Atlas (HCA) bone marrow dataset[285] containing 378.000 cells was downloaded using the HCAData package.[286] For each individual donor, cells containing less than 3 median absolute deviations (MAD) of median number of total UMIs or total genes and/or more than 3 MAD of the median percentage of mitochondrial genes were removed, using the quickPerCellQC function in scater. All the files were log transformed and normalized using deconvolution size factors. Integration of our dataset with the Yang dataset required correction of sequencing depth by recomputing log-normalized expression values after adjusting the size factors (multiBatchNorm function in Batchelor[265]). Integration of the Yang and HCA donors was performed using mutual nearest neighbors (MNN) correction[265], based on the top 2000 most variable genes. Graph-based walktrap clustering was performed using kmeans in Scran. Both the clustering and the UMAP were based on the corrected PC scores. Reference-based analysis, using the Blueprint[287] and ENCODE[288] datasets, or our own annotated dataset was performed to annotate cell clusters by SingleR[289].

NK cell stimulation

Peripheral blood mononuclear cells (PBMC) and BMMC were thawed in AIMV-medium (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 20% Fetal Calf serum (FCS, Sigma-Aldrich, Saint Louis, MI, USA) and 1600 IU/ml DNAse (Merck, Darmstadt, Germany). Alternatively, NK cells were enriched from fresh PBMC or BMMC by Mojosort (Biolegend) or EasySep (StemCell Technologies, Vancouver, Canada) magnetic cell separation according to manufacturer's instructions. Cells were cultured in AIM-V medium supplemented with 10% FCS and 1% Penicillin/Streptomycin in the presence of 10 ng/ml IL-12 (PeproTech, Rocky Hill, NJ, USA), 10 ng/ml IL-15 (CellGenix, Freiburg, Germany) and 20 ng/ml IL-18 (20 ng/ml, MBL International, Woburn, MA, USA) for 4h, or K562 cells (overnight) at 37 °C. For the CD16 and NKp46/2B4 stimulation, a flat-bottom plate was coated with 5 ug/ml goat anti-mouse antibody (BD) and subsequently with 1 ug/ml anti-CD16 (clone 3G8, Biolegend), or a combination of anti-NKp46 (clone 9E2, BD) and anti-2B4 (clone C1.7, eBioscience) in PBS. 1 ug/ml mouse IgG1 isotype (Biolegend) served as negative control. Cells were transferred to the flat-bottom plate and cultured for 4h (CD16) or overnight (NKp46+2B4) at 37 °C. For each cell culture condition, Golgistop (BD) was added after 1h of culture.

Spleen, tonsil and lymph node cell isolation

Splenic tissues, lymph nodes and tonsils were stored in University of Wisconsin solution at 4˚C and processed within 12h after surgery. Tissues were dispersed through a 70-mm cell strainer. Mononuclear cells were isolated from spleens and tonsils by ficoll density gradient centrifugation. Samples were analyzed immediately by conventional flow cytometry.

Conventional and spectral flow cytometry

Mononuclear cells were stained with fluorochrome-conjugated antibodies (Table S1) in PBS supplemented with 0.5% Bovine Serum Albumin, 2mM EDTA (Merck) and 0.02% NaN_3 (LUMC Pharmacy), for 30 minutes at room temperature (RT). For spectral cytometry, Brilliant Stain buffer plus (BD) was added to this mix. For the intracellular staining, cells were fixed in 4% paraformaldehyde and permeabilized in 0.1% saponin, as previously described.[202] Next, cells were incubated in Fc block (eBioscience) for 10 minutes at RT. Intracellular staining with antibodies (Table S1) was performed for 30 minutes at 4 °C. For the XCL1 staining, 6.7% donkey serum (Jackson ImmunoResearch, West Grove, PA, USA) was added together with unconjugated polyclonal goat anti-XCL1 to the mix. Finally, cells were incubated with donkey-anti-goat as secondary antibody to detect XCL1 for 30 minutes at 4 °C. A reliable bi-modal population was observed upon 4h PMA (12.5 ng/ml) and ionomycin (1 µg/ml) stimulation (Figure S5B). For cells only stained extracellularly, DAPI was added prior to measurement to detect dead cells. Data was acquired on a LSRII flow cytometer (BD) or 3L/5L Aurora spectral cytometer (Cytek Biosciences, Fremont, CA, USA), using Diva software (v8.0) or SpectroFlo software (v2.2.3, Cytek), respectively.

Cytometry data analysis

All cytometry data were analyzed on the OMIQ data science platform (Omiq, Inc, Santa Clara, CA, USA). For spectral cytometry data, FlowAi[290] was applied to detect anomalous events, based on

changes in flow rate, or outlier events. Data were compensated, arcsinh transformed, and gated to remove dead cells and doublets, as previously described.[291] To remove inter-experiment variation, data was normalized using fda normalization.[255] UMAP was applied for visualization.[102] NK cell and other subsets were either selected by gating on a 2-dimensional plot, or on the UMAP.

Statistics

Differentially expressed genes between clusters and subclusters derived from the single-cell RNA sequence data were determined by performing a paired one-sided and two-sided Wilcoxon test, respectively. The Wald test was applied, using the associationTest function in Tradeseq (default parameters), to find significant genes for each pseudotime lineage. Genes with a false discovery rate (Wilcoxon test) or P-value (Wald test) <0.05 were considered as statistically significant.

Results

In this study we performed single-cell RNA sequencing on fresh NK cell-enriched bone marrow cells. The NK cell-enriched fraction was labeled with oligonucleotide-conjugated antibodies against CD56, CXCR6, CD117 and CD34, which allows identification of subsets at protein level (Figure S1A). Only 21 out of 7000 cells did not fulfill the strict quality control criteria. In total 6979 cells, 33538 genes and 4 antibodies were further analyzed (Figure S1B). With a median of 1635 genes and 3748 unique molecular identifiers (UMIs) per cell, a high resolution single-cell dataset was created. Principal component analysis (PCA) was performed on the top 2000 most variable genes, and the first 20 PCs were used for clustering and uniform manifold approximation and projection (UMAP; Figure S1C-D).

Clustering results in identification of CD34^{+} progenitor cells, ILCs and six distinct NK cell subsets

Clustering of cells based on the gene expression revealed the presence of eight different clusters (Figure 1A). Based on the CD56 and CXCR6 protein expression, cluster 1 (84% of all cells) and 3 (3.8%) represented CD56^{dim} NK cells, cluster 2 (3.7%) represented the CXCR6^{+} lymphoid tissue-resident (lt)NK cells and cluster 4 (3.2%) represented the CD56^{bright} NK cells (Figure 1B). Cluster 7 (1.5%) was a mixed population of CD56^{bright} and CD56^{dim} NK cells. CD34 was exclusively expressed on cluster 5 (1.8%). CD117 was expressed on CD56^{bright} and CD34^{+} cells (cluster 4 and 5) as well as on CD56^{-}CD34^{-} cells in cluster 6 (1.6%). The cells in cluster 8 (0.4%) did not express any of the four proteins.

Based on the differentially expressed genes between clusters, we concluded that CD56^{dim} NK cells were separated into a *GZMK*$^{-}$ (cluster 1) and *GZMK*$^{+}$ (cluster 3) subset (Figure 1C-D). The CD56^{dim}*GZMK*$^{+}$ NK cells were further characterized by higher expression of *SELL* (CD62L), and lower expression of *GZMB*, *PRF1 and FCGR3A (CD16)* and *FGFBP2* compared to CD56^{dim}*GZMK*$^{-}$ cells, suggesting that these cells represent an intermediate stage between CD56^{bright} and CD56^{dim} NK cells (Figure 1C-D). The CD56^{bright} NK cells in cluster 4 had high expression of markers which are known to be expressed at protein level as well: *SELL*, *CD2* and *IL7R* (Figure 1C-D). Among the

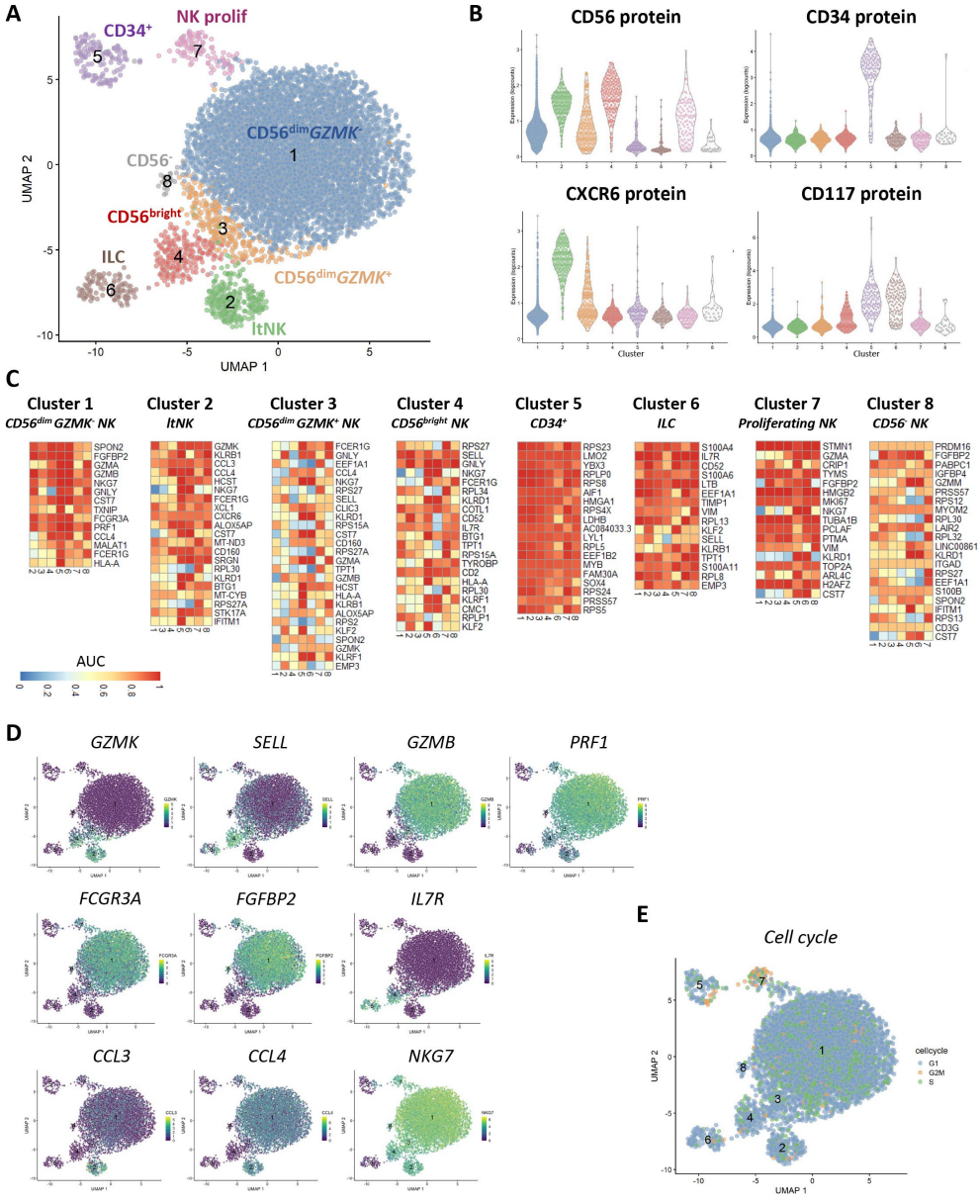

Figure 1. Clustering reveals six distinct NK cell populations

Single-cell RNA sequencing was performed on NK cell-enriched bone marrow cells of one healthy donor. In total 6979 cells, 33538 genes and 4 antibodies were further analyzed. A median of 1635 genes per cell was detected. ***A)*** *The UMAP is based on the top 20 principal components. Clustering on the same components revealed eight distinct clusters.* ***B)*** *The protein expression per cluster as determined by oligonucleotide-conjugated antibodies.* ***C)*** *The top 5 upregulated genes for each cluster comparison is shown.* ***D)*** *For a selection of genes the expression levels have been projected on the UMAP.* ***E)*** *The cell cycle stage was calculated for each cell based on a list of cell-cycle related genes.*

upregulated genes in the ltNK cells in cluster 2, various chemokines (*CCL3*, *CCL4*, *XCL1*) were included, but also *GZMK* and *CD160* (Figure 1C-D). The absence of *NKG7*, and high expression of *IL7R* in cluster 6 suggests that these cells are non-cytotoxic innate lymphoid cells (ILCs, Figure 1C-D). Cluster 7 was characterized by upregulation of cell-cycle related genes as also demonstrated by the cell cycle score, thus representing proliferating NK cells (Figure 1C,E). The CD56$^-$ cells in cluster 8 did express *NKG7* and *FCGR3A*, but had low expression of *PRF1* and *GZMB*, and likely represent CD56$^-$CD16$^+$ NK cells, a population with reduced effector function.[292–296]

The heterogenous CD56^{dim}*GZMK*$^+$ population includes the adaptive-like and terminally differentiated NK cells

Subclustering of the CD56^{bright}, ltNK and CD56^{dim}*GZMK*$^+$ cells did not reveal additional clusters, except for UMAP based separation of *CCL5*$^-$ and *CCL5*$^+$ cells (Figure S2A-C). Importantly, phenotypical markers used for further classification of the CD56^{bright} (i.e. CD117, CD27, CD16) and ltNK cells (i.e. CD16, DNAM-1, NKG2A) did not reflect distinct subsets at the transcriptional level (Figure S2A-B). Within the CD56^{dim}*GZMK*$^+$ NK cell population minor heterogeneity was observed based on *GZMB* and *FGFBP2* expression (Figure S2C). In contrast, an high heterogeneity was observed within the CD56^{dim}*GZMK*$^-$ population, as demonstrated by seven subclusters. Cluster 1 was mainly defined by absence of *CCL5* (Figure 2A-C). Cluster 2 was characterized by lower expression of *XCL2* and *EOMES,* and higher expression of *FGFBP2*. Upregulated genes in cluster 3 included mitochondria associated genes (*MT-CO3*, *MT-CO1*, *MT-CO2*, *CMC1*) which points to increased metabolic activity. Cluster 4 was separated based on the high expression of *KLRC1* (*NKG2A*). The high RNA content and number of detected genes in cluster 5 was suspicious for doublets, therefore this cluster was removed from subsequent analyses (Figure 2B). *S100A4* and *S100A6* were expressed at the highest level in cluster 6 and cluster 7, suggesting terminally differentiated NK cells.[37,297,298] Unfortunately, the enzyme encoded by *B3GAT1*, creating the CD57 epitope, was not sufficiently detected. Nevertheless, genes involved in cytoskeleton remodeling (*ACTB*, *ACTG1*, *CORO1A, PFN1*) were upregulated in cluster 6, a pathway which was earlier shown to be enriched in CD56^{dim}CD57$^+$ NK cells to enable high cytotoxicity (Figure 2C).[298] The upregulation of *NKG2C (KLRC2)* in cluster 7 points to the presence of adaptive-like CMV-associated cells.[299,300] This cluster was further characterized by upregulation of ribosomal protein-coding genes, *IL32*, *GZMH* and *GNLY*, and downregulation of *NKG2A*, *KLRB1* (*CD161), CD160, XCL2* and *CCL3* (Figure 2B-C).

Inhibitory KIRs are not major drivers of CD56^{dim}*GZMK*$^+$ subclustering, but are more frequently detected in terminally differentiated cells

Although terminally differentiated cells more frequently express inhibitory KIRs, the KIRs were not included among the top 5 differentially expressed genes for each cluster comparison (Figure 2C).[174] Within each (sub)cluster, multiple levels of inhibitory *KIR* expression were detected (Figure 2D). The vast majority of the CD56^{bright} and ltNK cells did not express any *KIR*, while the intermediate CD56^{dim}*GZMK*$^+$ population expressed KIRs at levels (16%) between CD56^{bright} and CD56^{dim}*GZMK*$^-$ cells (Figure 2D-E). On the other hand, 35% and 56% of the terminally differentiated CD57$^+$-like and *NKG2C*$^+$ cells in cluster 6 and 7, respectively, expressed at least one KIR (Figure

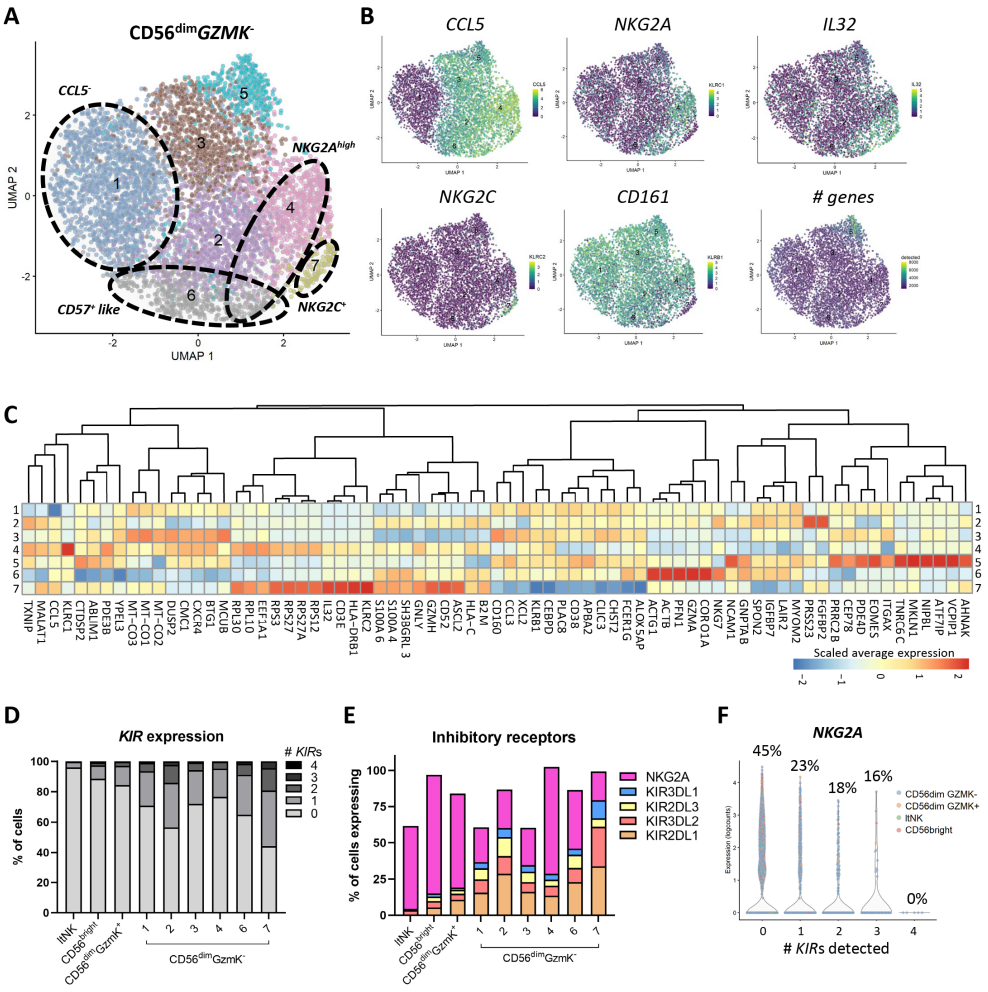

Figure 2. Heterogeneity of CD56^{dim}GZMK^{-} population

A) *The CD56^{dim}GZMK^{-} cells were selected and subclustering was performed. In total 7 clusters were identified.* ***B)*** *The expression levels of several genes, and the number of detected genes are shown.* ***C)*** *The top 5 most differentially expressed genes for each cluster comparison (either up- or downregulated) were selected. The average scaled expression for each cluster is shown.* ***D)*** *The percentage of inhibitory KIR expressing cells for each cluster was based on KIR2DL1, KIR3DL2, KIR2DL3 and KIR3DL1.* ***E)*** *The expression of inhibitory receptors as percentage of total cells, per cluster is shown.* ***F)*** *To study the relation between NKG2A and the inhibitory KIRs, the expression of NKG2A is shown as counts and percentage, sorted by number of KIRs detected.*

2D). Of the four inhibitory KIRs detected in our complete dataset, *KIR2DL1* was most frequently observed (Figure 2E). The percentage of cells expressing *NKG2A* was negatively related to the number of *KIR* molecules detected per cell, confirming previous findings (Figure 2E-F).[174] Overall, CD56^{dim}*GZMK*$^{-}$ subclusters did not represent unique educational states, but inhibitory *KIR*s were more frequently detected in terminal differentiated cells.

Integration with NK cells from eight donors validates the identity of CD56⁺ NK cell subsets

To validate the presence of the identified NK cell subsets in other donors, we integrated our dataset with a public single-cell RNA sequencing dataset of Lin⁻CD56⁺/⁻CD7⁺ cells (total 9071) from bone marrow (n=6) and blood (n=2).[272] In this public dataset multiple subsets of mature NK cells and low numbers of ILCs, T cells, and progenitor cells were detected (Figure S3). After integration with our data, we annotated the cells from the public dataset using our previously defined NK cell clusters as reference. The major 8 subsets as previously described (Figure 1A) again clustered in the UMAP, except the CD56⁻ NK cells, indicating that this population may be donor-specific (Figure 3A). The CD34⁺ and ltNK cells were barely detected in blood, confirming their bone marrow residency (Figure 3B). Except for the proliferating NK cells, the identified CD56⁺ NK cell subsets were present in bone marrow of all individual donors (Figure 3A). Within the CD56^{dim}*GZMK*⁻ subset of the public dataset, *CCL5*⁻, CD57⁺-like, *NKG2C*⁺ and activated NK cells (*NFKBIA*↑, *FOS*↑, *JUNB*↑), were identified in most of the donors (Figure S3B-D). A difference in isolation protocol and/or donor characteristics might explain the absence of activated CD56^{dim}*GZMK*⁻ NK cells in our dataset. No subclusters were identified in the CD56^{bright,} ltNK and CD56^{dim}*GZMK*⁺ population (Figure S2D-F). In conclusion, the identified CD56⁺ NK cell subsets in our dataset are not donor-specific which confirms that they are common populations.

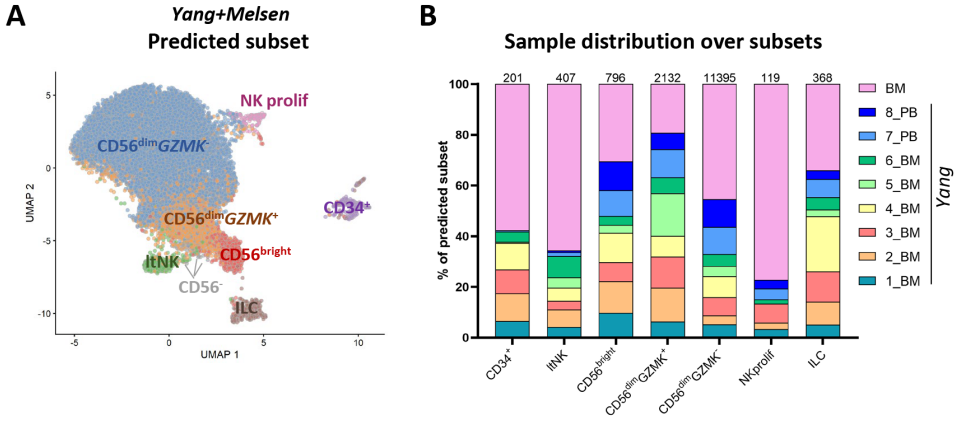

Figure 3. Integration of dataset with public single-cell RNA sequence data validates clusters

A) *Our bone marrow NK cells were integrated with single-cell RNA sequencing data from sorted NK cells (total 9071) from eight donors (two blood, six bone marrow samples) of the Yang et al. dataset. The UMAP was based on the corrected principal components scores. Cells were labeled using an automatic annotation method based on our own dataset as reference dataset.* ***B)*** *The distribution of our own dataset (pink) and individual samples from Yang et al. over each predicted subset is shown. The number of total cells per subset is indicated at the top of each bar.*

CD56^{dim}*GZMK*⁺ cells have an intermediate but distinct NKG2A^{high}CD16^{low}TIGIT^{high}KLRG1^{high} phenotype

To validate the presence of the newly identified CD56^{dim}*GZMK*⁺ NK cells using protein expression and further phenotypically profile them, we developed a 26-marker panel for spectral cytometry. Among the NK cells in bone marrow we identified the CD56^{bright}, CD56^{dim}GzmK⁺, CD56^{dim}GzmK⁻ and ltNK cells (Figure 4A). In the classical density plot of CD56 against CD16, the CD56^{dim}GzmK⁺

cells (orange) are positioned between the CD56^{bright} and CD56^{dim}GzmK^{-} population. A mean of 3.9% and 4.7% of CD56^{dim} NK cells in bone marrow and blood, respectively, expressed granzyme K (Figure 4B). The UMAP embedding of the NK cells from 14 healthy bone marrow donors, based on 21 NK cell markers, was comparable to the UMAP of the transcriptomic data, with the CD56^{dim}GzmK^{+} subset in between the other three NK cell subsets (Figure 4C, Figure S4A). The phenotype of CD56^{dim}GzmK^{+} cells was characterized by variable expression of CD56^{bright}-related markers CD127, CD27 and CD56^{dim}-related markers CX3CR1 and granzyme B (Figure 4C-E). The vast majority of CD56^{dim}GzmK^{+} cells expressed high levels of NKG2A and TIGIT, two markers also expressed by ltNK cells (Figure 4D-E). KLRG1, usually associated with mature CD56^{dim} NK cells, was also highly expressed by CD56^{dim}GzmK^{+} cells, underscoring the distinctiveness of this intermediate NK cell subset. No difference in phenotype of CD56^{dim}GzmK^{+} cells between blood and bone marrow was found (Figure S4B). Interestingly, in an IL2RG deficient patient who received a hematopoietic stem cell transplantation (HSCT) more than 50 years ago and is affected with chronic HPV disease, 26% of the CD56^{dim} NK cells expressed granzyme K and CD27 (Figure 4F).[301] While a similar CD56^{dim}CD27^{+} subpopulation with a GzmK^{+}NKG2A^{high}CD16^{low}TIGIT^{high}KLRG1^{high} phenotype was also present at a low frequency in healthy controls, it was significantly expanded in this post HSCT patient (Figure 4D-F). Combined, the high expression of KLRG1 and TIGIT, and expansion in a specific clinical condition, suggest that CD56^{dim}GzmK^{+} NK cells represent an intermediate but discrete stage during NK cell differentiation.

CD56^{dim}GzmK^{+} cells produce intermediate levels of chemokines and cytokines upon interleukin or target cell stimulation

The identification of the CD56^{dim}GzmK^{+} NK cell subset raised the question on its functional capacity compared to the CD56^{bright}, CD56^{dim}GzmK^{-} and ltNK subset. Based on the transcriptome, the CD56^{dim}*GZMK*$^{+}$ cells had an intermediate chemokine profile (based on *CCL3*, *CCL4*, *CCL5*, *XCL1*, *XCL2)* matching characteristics of both CD56^{bright}, CD56^{dim}*GZMK*$^{-}$ cells and ltNK cells (Figure 5A). ltNK cells had the highest chemokine expression, with each chemokine detected in at least 85% of the cells, while CD56^{bright} NK cells had the lowest overall chemokine expression, with absence of *CCL3* and *CCL4* (Figure 5A). Unfortunately, *IFNG* and *TNF* were underrepresented in our dataset. Notably, all those effector molecules, including *IFNG* and *TNF*, were previously shown to be expressed in each major NK cell subset but at different levels, by bulk mRNA sequencing (Figure S5A).[260]

The mRNA levels for cytokines were determined under steady state conditions. To study whether XCL1, CCL4, CCL5, IFN-γ and TNF-α were also produced at the protein level we stimulated NK cells in bulk mononuclear cells (MNC), or as an enriched NK fraction, with either interleukins (IL-12, IL-15, IL-18) or target-cell(-like) stimulations, anti-CD16, anti-NKp46 & anti-2B4 or K562 cells (Figure S5B, 5B). XCL1 was abundantly produced by all subsets in response to all stimuli (Figure 5C). In contrast, CCL4 and TNF-α were mainly produced by CD56^{dim}GzmK^{-} cells in response to target-cell(-like) stimulations, with a median of 48% and 5.5% positivity upon CD16 crosslinking, respectively. The highest IFN-γ production was observed upon interleukin stimulation in all subsets (Figure 5C). Distinct from the other effector molecules, CCL5 was spontaneously produced by a fraction of

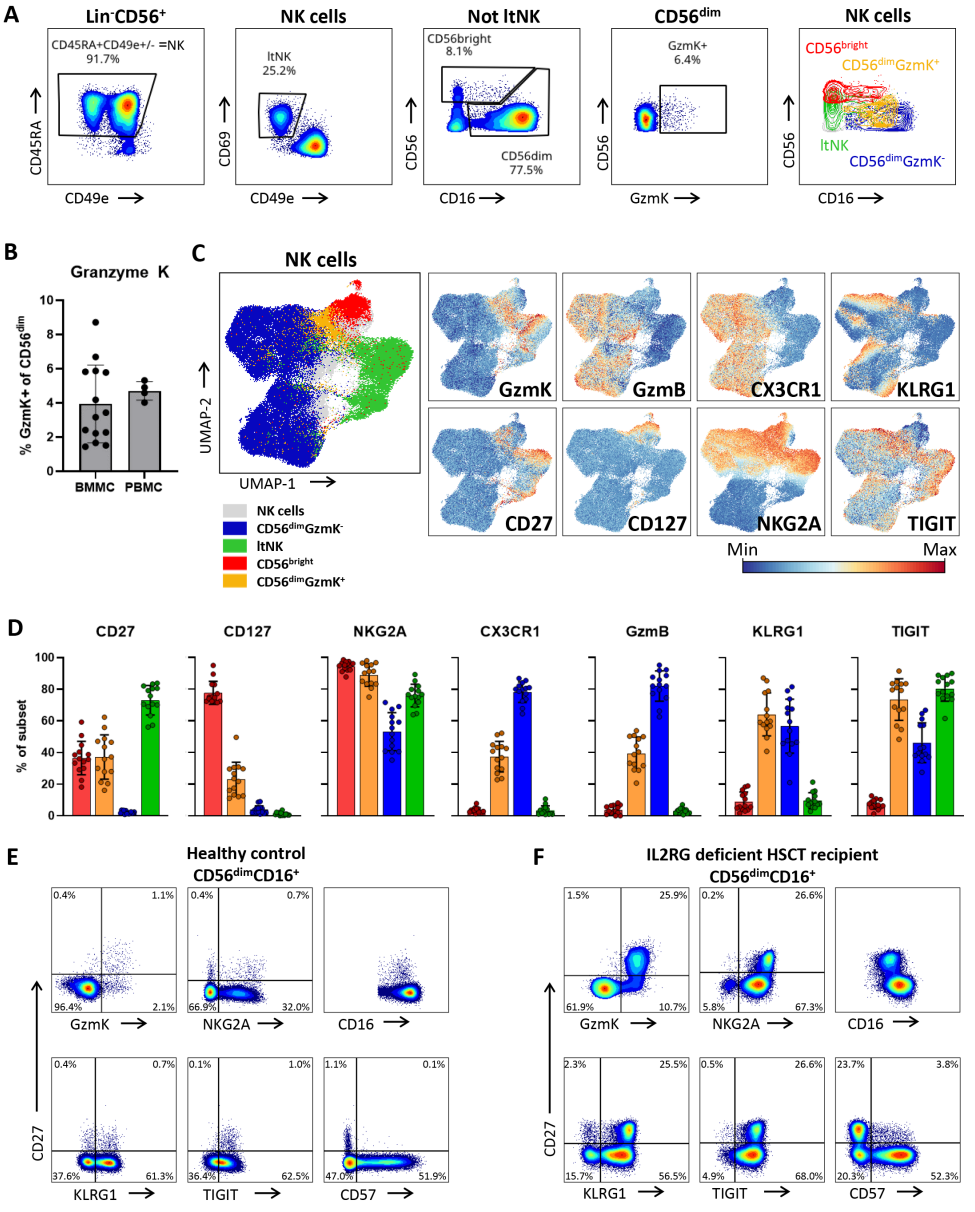

Figure 4. CD56^{dim}GzmK^{+} cells have an intermediate but distinct phenotype

A) *NK cells were selected as CD3^{-}CD14^{-}CD19^{-}CD34^{-}CD45^{+}CD56^{+}CD45RA^{+}. ltNK cells were defined as CD69^{+}CD49e^{-} (green). The non ltNK cells were subdivided into CD56^{bright}CD16$^{+/-}$ (red), CD56^{dim}CD16^{+}GzmK^{+} (orange) and CD56^{dim}CD16^{+}GzmK^{-} (blue) NK cells. One representative bone marrow donor is shown.* ***B)*** *Quantification of the CD56^{dim}GzmK^{+} cells in bone marrow (n=14) and blood (n=4).* ***C)*** *The UMAP embedding of 91079 NK cells from 14 thawed BMMC healthy control samples, based on 21 NK cell markers (Figure S4). The expression of a selection of markers is shown.* ***D)*** *Expression of a selection of markers on bone marrow NK cell subsets (n=14). The mean and SD are indicated in the bar graphs.* ***E)*** *Expression of a selection of markers plotted against CD27 on CD56^{dim}CD16^{+} NK cells of a representative healthy blood donor and* ***F)*** *IL2RG deficient patient that received a hematopoietic stem cell transplantation (HSCT) 51 years ago. In the patient an enrichment of the CD56^{dim}GzmK^{+} cells was observed.*

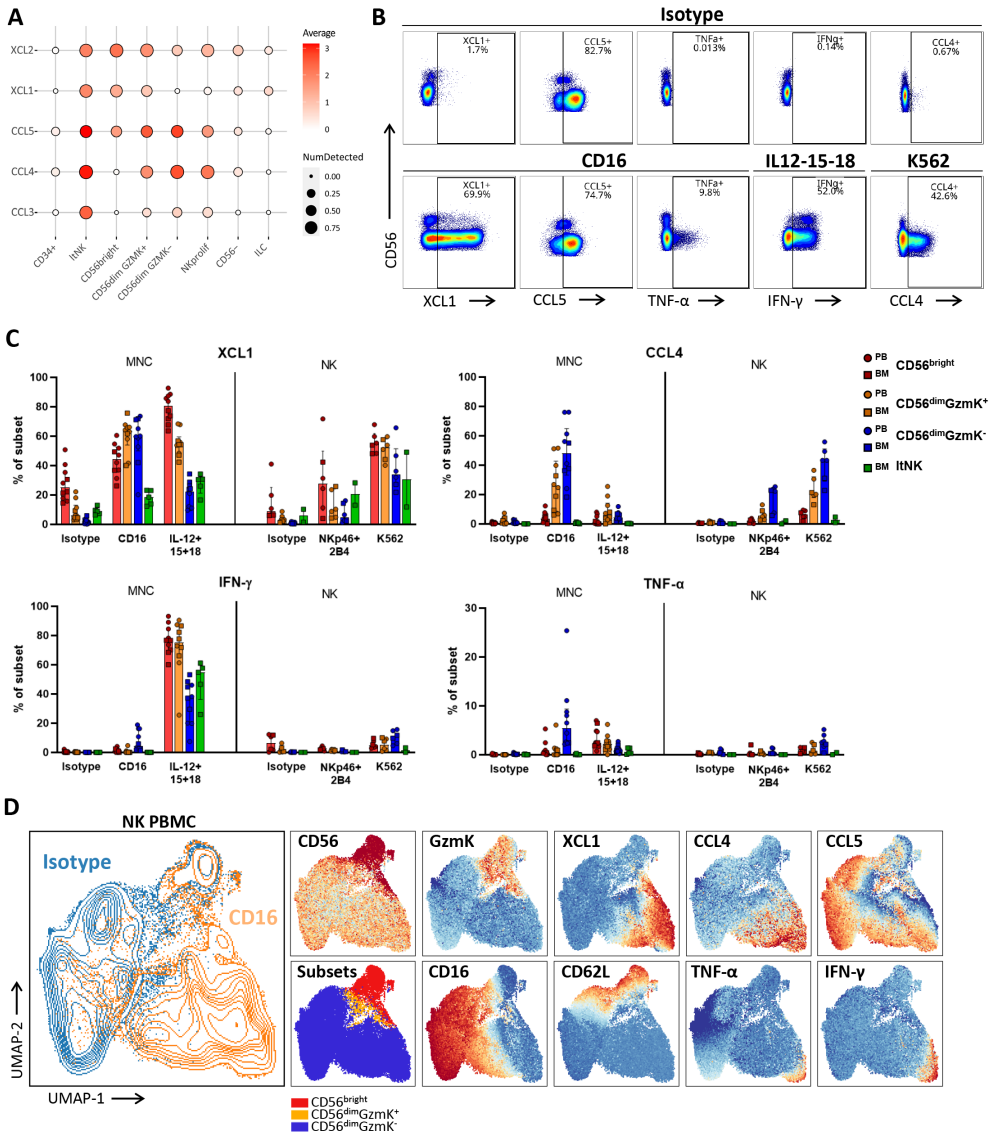

Figure 5. Chemokine and cytokine production by NK cell subsets

A) *The dotplot visualizes for each chemokine the percentage of cells within a NK cell subset expressing the chemokine, and the average mRNA expression.* ***B)*** *Intracellular staining pattern, as measured by spectral cytometry, for each chemokine and cytokine in NK cells cultured in the presence or absence of stimulation. A representative blood donor is shown.* ***C)*** *NK cells were either stimulated in bulk thawed mononuclear cells (MNC, n=5 blood, n=5 bone marrow) or as freshly enriched NK cell fraction (n=4 blood, n=2 bone marrow). MNC were stimulated for 4 hours and enriched NK cells were stimulated overnight. The intracellular effector molecule production was determined in the gated peripheral blood (PB) and bone marrow (BM) derived NK cell subsets. Bars indicate median and interquartile range.* ***D)*** *NK cells from PBMC stimulated with anti-CD16 or isotype control were selected and embedded in a UMAP. A representative donor is shown.*

each NK cell subset upon culture, likely reflecting the *CCL5* based subclustering and UMAP embedding of the transcriptomic data (Figure 5B, S5C). In line with literature, CD56^{dim}GzmK^{-} NK cells were overall most responsive to target-cell(-like) stimulations, while CD56^{bright} NK and ltNK cells were most responsive to interleukin stimulation. CD56^{dim}GzmK^{+} NK cells produced effector molecules at levels in between the production by CD56^{bright} and CD56^{dim}GzmK^{-} NK cells in response to both interleukins and target cell stimulation, reinforcing their position as an intermediate subset (Figure 5C).

To study the repertoire of produced effector molecules at the single-cell level we embedded NK cells cultured with either isotype or anti-CD16 in a UMAP. From this two-dimensional representation it became obvious that upon CD16 crosslinking the NK cells that lost expression of CD16 and were negative for CD62L were the main cytokine and chemokine producers (Figure 5D). We observed a higher frequency of cells positive for XCL1 and CCL4, than IFN-γ and TNF-α, that can be explained by the fact that chemokines are produced earlier compared to cytokines[12] (Figure 5D). Nevertheless, the cells that produced IFN-γ and TNF-α, also produced XCL1 and CCL4, indicating that the repertoire of effector molecules is similar among stimulated NK cells.

Pseudotime analysis suggests that ltNK and CD56^{bright} NK cells are separate lineages

To study the developmental relationship between the mature NK cell subsets CD56^{bright,} CD56^{dim}GzmK^{+}, CD56^{dim}GzmK^{-} and ltNK cells in bone marrow, we performed pseudotime analysis on our NK cell transcriptomic dataset using the Slingshot algorithm. Lineage identification based on these four major NK subsets, resulted in two lineages. The first lineage connected the CD56^{bright} NK cells to the CD56^{dim}GzmK^{-} NK cells, with the CD56^{dim}GzmK^{+} cells as intermediate NK cell subset (Figure 6A). The terminally differentiated CD57^{+} NK cells were positioned at the other end of the principle curve. Therefore, we considered those as most differentiated cells, and the CD56^{bright} NK cells as starting subset. Among the most differentially expressed genes driving the pseudotime, we identified effector molecules *FGFBP2*, *PRF1*, *GZMB*, *GZMK,* the transcription factor *TCF7,* and the adhesion molecule *CD44* (Figure 6B). Interestingly, the ltNK cells were assigned as separate lineage and were not connected to one of the circulating NK cell subsets (Figure 6A). This notion was further supported by adjusting the minimum distance parameter of the UMAP. With a lower minimum distance, the CD56^{bright} NK cells were still connected to the CD56^{dim}GzmK^{-} NK cells via the CD56^{dim}GzmK^{+} subset, while the ltNK cells were not (Figure S6). This suggests that the resident ltNK cells develop independently from the circulating NK cells. Still, there might be NK cell progenitors that give rise to both the ltNK cells and the other NK cell subsets.

No evidence for NK cell development in human bone marrow at steady state

The presence of CD34^{+} cells in our dataset allowed us to explore early NK cell development. We re-clustered the CD34^{+} cells and identified multiple lineage committed and progenitor cells guided by the cell subset scores based on reference bulk RNA sequence data[288] (Figure S7A-B). Since there were too few NK progenitor cells to reliably model the NK cell development we integrated our dataset with 316941 bone marrow cells from the human cell atlas. All hematopoietic populations were identified, including the CD34^{+} cells (cluster 4) and the NK cells (cluster 6, Figure

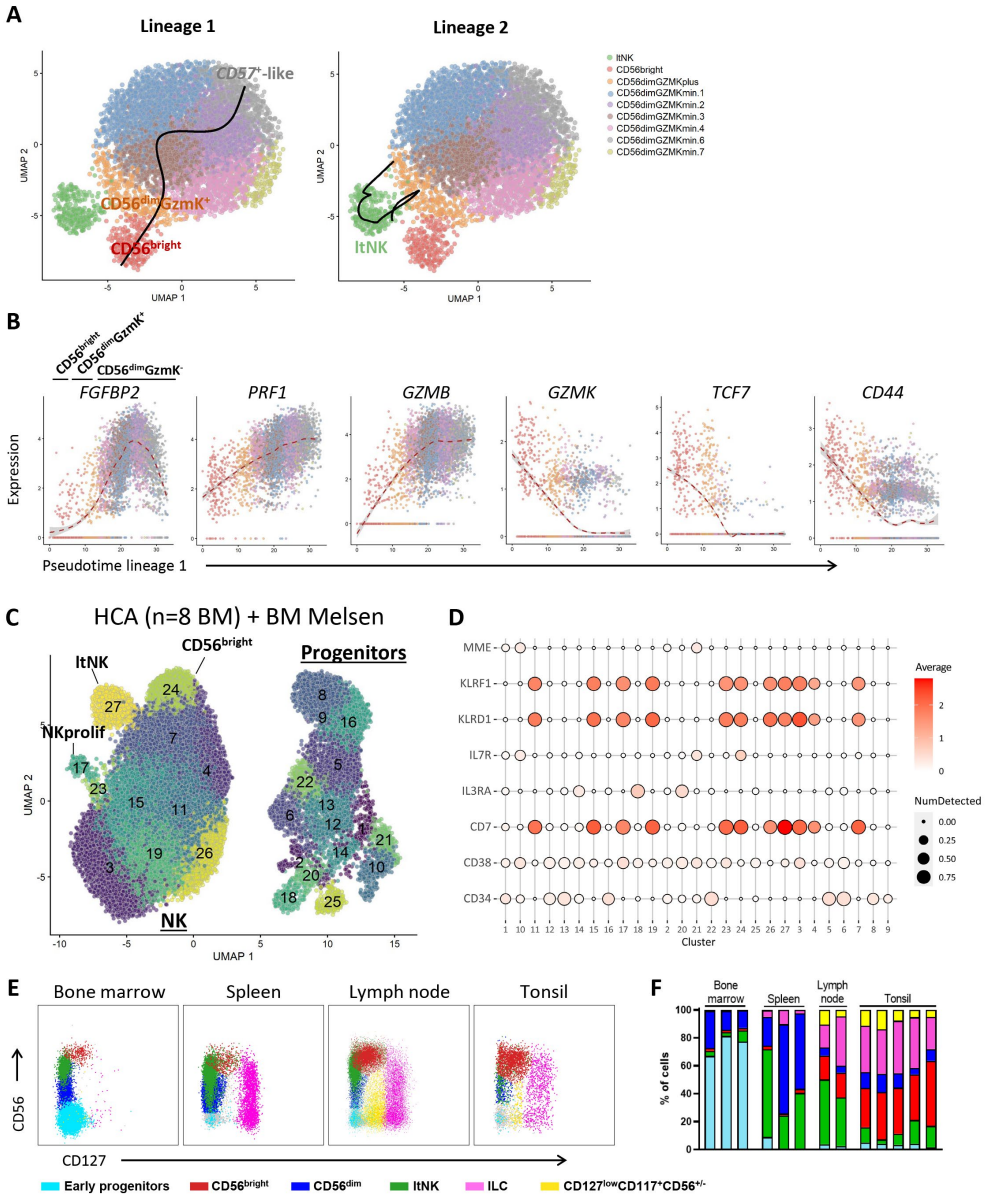

Figure 6. Pseudotime analysis in bone marrow suggests that ltNK cells develop independently from circulating NK cells

A) *Pseudotime analysis was performed on our NK cell dataset using Slingshot, based on the first 20 principal components and four NK cell clusters (CD56^{bright}, CD56^{dim}GzmK^{+}, CD56^{dim}GzmK^{-} and ltNK). The two principal curves of the two lineages are shown.* ***B)*** *Six genes of which the expression changed differentially among pseudotime of lineage 1 are shown.* ***C)*** *To study early NK cell development, our dataset was integrated with the human cell atlas (HCA) bone marrow single-cell RNA sequencing dataset (8 donors). CD34^{+} cells and NK cells were selected and re-clustered. See Figure S8 for the selection procedure.* ***D)*** *For each cluster, the expression as percentage and intensity of a selection of progenitor-related genes is shown. No NK progenitor cluster with combined expression of CD34, CD38, CD7 and CD10 (MME) could be identified.* ***E)*** *Fresh bone marrow (n=3), lymph node (n=2), tonsil (n=5) and spleen (n=3) >>>*

<<< Figure 6. legend continued

samples were analyzed for the presence of NK cell subsets (CD56^{+}CD127$^{low/-}$), ILC (CD127^{high}CD117^{+}) and early progenitor cells (CD45$^{-/dim}$CD117^{+}SSC^{high}). Monocytes, T cells and B cells were excluded based on CD3, CD19, CD16, CD45 and SSC. In the remaining cells a population of CD127^{low}CD117^{+}CD56$^{+/-}$ cells was identified (yellow), present in lymph node and tonsil, but absent in spleen and bone marrow. ***F)*** *subset distribution in the individual tissue samples.*

S8A-B). We selected the CD34^{+} and NK cells, clustered the data and removed the megakaryocyte-erythroid progenitors, megakaryocytes, adipocytes and macrophages (Figure S8C-D). Strikingly, in the new UMAP embedding of the remaining cells we found no connection between the NK and the progenitor cell clusters, suggesting that there were no cells with characteristics representative for intermediate developmental stages (Figure 6C). The common lymphoid progenitors (CLP) were identified in cluster 10 and 21 (*CD34^{+}CD10^{+}IL7R^{+}CD38^{+}*). However, we found no cluster with combined expression of *CD34*, *CD38*, *CD7* and *CD10* (*MME*), as has been previously proposed as human NK progenitor cell (Figure 6D).[302] Multiple articles postulated that development of NK cells occurs in secondary lymphoid organs.[303–305] In agreement, we demonstrate that lymph node and tonsil, but not bone marrow and spleen, harbor a cell population with an early NK cell phenotype (CD127^{low}CD117^{+}CD56$^{+/-}$) that is absent from bone marrow and spleen (Figure 6E, Figure S9). Moreover, the less mature CD56^{bright} NK cell subset was highly enriched in tonsil and lymph node, compared to bone marrow and spleen. In conclusion, these findings provide evidence that NK cell development occurs in tonsil and lymph node, rather than in bone marrow.

Discussion

Single-cell RNA sequencing combined with antibodies of human bone marrow NK cells revealed that within the CD56^{dim} NK cells a subpopulation exists characterized by expression of granzyme K. The transcriptional, phenotypical and functional profiles of the minor CD56^{dim}GzmK^{+} population were intermediate between the CD56^{bright} and CD56^{dim}GzmK^{-} subset. Pseudotime analysis positioned CD56^{bright}, CD56^{dim}GzmK^{+} and CD56^{dim}GzmK^{-} cells in one differentiation trajectory, while ltNK cells were developmentally separated. Progenitor NK cells could not be identified in bone marrow, suggesting that other lymphoid tissues are responsible for NK cell development.

Using trajectory analysis on single-cell RNA sequence data of NK cells, we confirmed that the CD56^{bright} NK cells are precursors of the CD56^{dim} NK cells.[262,272,275] Clustering and trajectory analysis of our single-cell RNA sequence data, and phenotypical and functional analysis identified CD56^{dim}GzmK^{+} cells as intermediate subset, with a CD27$^{+/-}$GzmB^{low}CX3CR1^{low}NKG2A^{high}CD16^{low} KLRG1^{high}TIGIT^{high} phenotype. In literature, multiple populations have been proposed as intermediate differentiation stage, based on flow cytometry data: CD56^{bright}CD16^{+}, CD56^{bright}CD27^{-}, CD56^{dim}CD94^{high} , CD56^{dim}CD62L^{+} and CD56^{dim}CD16^{dim}.[46,140,172,173,175,306] Phenotypically, the CD56^{dim}GzmK^{+} subset resembles the proposed CD56^{dim}CD94^{high} and CD56^{dim}CD16^{dim} NK cell subsets.[172,306,307] However, in contrast to the gradual differences in CD94 and CD16 expression, the bi-modal expression of granzyme K within the CD56^{dim} NK cells allows for strict gating of populations. Moreover, in parallel to CD8 T cells, conversion of granzyme K to granzyme B was driving the differentiation in pseudotime analysis.[82,308,309] Therefore, we consider CD56^{dim}GzmK^{+}

cells as an intermediate NK cell subset. The question arises whether this subset represents a continuum of cells differentiating towards the CD56^{dim}GzmB^{+} subset, or whether it represents a defined differentiation stage. The fact that this subset was expanded in a IL2RG deficient patient post HSCT, but also uniformly expressed high levels of both KLRG1 and TIGIT, is in favor of a defined differentiation stage.

By subclustering of the CD56^{dim}GzmK^{-} NK cells, the known terminally differentiated CD57^{+} and adaptive-like NKG2C^{+} NK cells were most notably different. The functional response of CD56^{dim} NK cells is influenced by the educational state, i.e. the expression of self-inhibitory KIRs and NKG2A.[16,174] Moreover, CD56^{dim} NK cell differentiation is associated with loss of NKG2A and sequential gain of KIR expression.[174] Indeed, in our spectral cytometry data, NKG2A and KIR expression were driving the UMAP embedding, however this was not obvious from the transcriptome-based pseudotime analysis or clustering, except for the lowest and highest inhibitory *KIR* expression observed in a minor *NKG2A*^{high} cluster and *NKG2C*$^{+}$ cluster, respectively. As previously reported, no difference in NKG2A^{+} and NKG2A^{-} CD56^{dim} NK cells was observed by bulk RNA sequencing, and increase of cytotoxicity in educated NK cells was shown to be independent of transcription.[260,310] Thus, although the single-cell RNA sequencing results suggest that expression of genes encoding for the inhibitory receptors are not the main drivers of CD56^{dim} NK cell subset classification, the additional use of KIR-specific oligonucleotide-conjugated antibodies in future experiments will be essential in understanding the NK cell education in relation to the process of differentiation.

Although all the NK cells subsets were identified by clustering, the use of oligonucleotide-antibodies specific for CD56 and CXCR6 was of added value to recognize the tissue-resident ltNK cells in our data set. Previously, this population was unrecognized by single-RNA sequencing on bone marrow NK cells.[262,272,311] Like the CD56^{bright} NK cells, the ltNK cells express granzyme K, but not granzyme B, perhaps explaining their low cytotoxicity compared to CD56^{dim} NK cells at resting state.[58,271] The highest functional response of ltNK and CD56^{bright} NK cells was observed upon interleukin stimulation, as shown by XCL1 and IFN-γ production. Although ltNK cells expressed the highest mRNA levels of multiple effector molecules[260], including CCL4 and TNF-a, neither our target-cell or antibody stimulation, nor interleukin stimuli induced the highest production in ltNK cells compared to the other subsets. Since no subclusters were observed within the ltNK cells, this lower response is probably not explained by the heterogeneity within the ltNK cell population. It requires further investigation to decipher the appropriate physiological stimuli and optimal duration of in vitro stimulation for ltNK cells.

In agreement with murine data on tissue-resident NK cells, the ltNK cells develop independently from the circulating NK cells based on trajectory analysis.[176] Although we hypothesized that the ltNK and CD56^{bright} NK cells diverge from a shared NK cell progenitor, we were not able to decipher the in situ development of ltNK cells nor the CD56^{bright} NK cells, since no NK precursor cells were identified in bone marrow from healthy donors. The hypothesis that bone marrow is the site for human NK cell development originates mainly from in vitro studies, where early bone marrow-

derived progenitor cells were sorted and cultured in the presence of growth factors, interleukins and/or bone marrow stromal cells to generate cytotoxic NK cells.[312,313] The current model describes that hematopoietic stem cells give rise to common lymphoid progenitor cells (CLP), which subsequently downregulate CD34 and upregulate CD56.[314] However, the exact sequential steps especially in early NK cell development are not completely understood. An NK-lineage restricted progenitor has been described in human bone marrow[302], but we were neither able to identify these cells in a human bone marrow, nor did we identify intermediate stages linking $CD34^+$ progenitors to mature NK cells. In contrast, in the lymph node and tonsil, after exclusion of ILCs, $CD127^{low}CD117^+CD56^{+/-}$cells were identified that might represent the direct precursors of the $CD56^{bright}$ NK and ltNK cells. Multiple reports provided evidence for the presence of NK progenitor cells and intermediate stages in secondary lymphoid tissues[303–305,315,316], but the appropriate markers to define these cells remain yet to be identified. Therefore, it would be interesting to decipher the NK cell development by single-cell trajectory analysis of RNA and protein data in these tissues.

Overall, the use of untouched fresh NK cells (including $CD34^+$ cells) and oligonucleotide-labeled antibodies, resulted in a unique single-cell RNA sequencing dataset of high quality based on the number of genes and UMIs detected per cell. Still, the detection limit of single-cell RNA sequencing compromises studying markers with limited mRNA expression, highlighting the use of additional oligonucleotide-conjugated antibodies, and validation by other techniques. Although we performed single-cell RNA sequencing on one healthy donor, data integration with public datasets and spectral cytometry on 14 donors validated our results.

In conclusion, we provide detailed analyses on single-cell RNA sequence data of human bone marrow NK cells. Our work challenges the current statement that NK cell development occurs in bone marrow, proposes that tissue-resident ltNK cells develop independently from circulating NK cells, and define $CD56^{dim}GzmK^+$ NK cell population as an intermediate stage in NK cell differentiation.

Data availability

The single-cell RNA sequence data generated in this study have been deposited in the GEO database under accession number GSE199411.

Acknowledgments

The authors thank Dr. Susan Kloet, Yavuz Ariyurek and Emile de Meijer from the Leiden Genome Technology Center (Leiden University Medical Center, LUMC) for technical assistance of the library preparation. The authors acknowledge Flow cytometry Core Facility of LUMC, coordinated by Dr. Koen Schepers and Marjolijn Hameetman, for facilitating the use of flow cytometers (https://www.lumc.nl/research/facilities/fcf). The authors thank Dr. Pauline van Schouwenburg (LUMC) and Martijn Cordes (LUMC) for critical discussion on the single-cell RNA sequencing data analysis. J.E.M. was supported by funding from the Leiden University Medical Center, the

graduate program of Nederlandse Organisatie voor Wetenschappelijk Onderzoek and Stichting Zeldzame Ziekten Fonds (SCID project).

Supplemental data

1. Supplemental tables

Table S1. Antibodies used for flow cytometry (continued at next page>>>)

Specificity		Antibody characteristics					
CD designation	Alternative name	Fluorochrome	Type	Clone	Company	Catalog#	Panel
CD3	CD3	BV421	m-IgG1	UCHT1	BD	562426	LSR
CD7	GP40	AF700	m-IgG1	M-T701	BD	561603	LSR
CD16	FCGR3A	BV711	m-IgG1	3G8	BD	563127	LSR/Au
CD19	CD19	BV510	m-IgG1	SJ25C1	BD	562947	LSR
CD34	CD34	PE	m-IgG1	8G12	BD	345802	LSR
CD45	PTPRC	PE-Cy5.5	m-IgG1	J33	BC	A54139	LSR
CD56	NCAM	APC	m-IgG1	N901	BC	IM2474	LSR
CD56	NCAM	ECD	m-IgG1	N901	BC	A82943	LSR
CD56	NCAM	PE-Cy7	m-IgG1	N901	BC	A21692	LSR
CD69	EA-1	PE-Cy7	m-IgG1	L78	BD	335792	LSR
CD117	c-KIT	PE	m-IgG1	104D2	BD	332785	LSR
CD117	c-KIT	BV510	m-IgG1	104D2	Biolegend	313220	LSR/Au
CD127	IL7R	FITC	m-IgG1	eBioRDR5	eBioscience	11-1278-42	LSR
CD186	CXCR6	Unconjugated	m-IgG2b	56811	R&D	MAB699	LSR
CD186	CXCR6	BV421	m-IgG2a	K041E5	Biolegend	356014	LSR/Au
Secondary	m-Ig	APC	g-Ig	polyclonal	BD	550826	LSR
Secondary	m-IgG2b	AF647	g-IgG	polyclonal	Invitrogen	A21242	LSR
CD2	LFA-2	FITC	m-IgG1	MT910	DAKO	f767	Aurora
CD3	CD3	BV480	m-IgG1	UCHT1	BD	566105	Aurora
CD3	CD3	BV785	m-IgG1	UCHT1	Biolegend	300472	Aurora
CD7	GP40	AF532	m-IgG1	eBio124-1D1	eBioscience	58-0079-42	Aurora
CD8	CD8	BUV395	m-IgG1	RPA-T8	BD	9317565	Aurora
CD16	FCGR3A	Vio515	rec h-IgG1	REA423	Miltenyi	130-119-616	Aurora
CD16	FCGR3A	BV480	m-IgG1	3G8	BD	566108	Aurora
CD27	TNFRSF7	APC-Fire 810	m-IgG1	QA17A18	Biolegend	393213	Aurora
CD56	NCAM	BV650	m-IgG2b	NCAM16.2	BD	564057	Aurora
CD57	HNK-1	PB	m-IgM	HNK1	Biolegend	359608	Aurora
CD45	PTPRC	APC-H7	m-IgG1	2D1	BD	641417	Aurora
CD45RA	PTPRC	BV570	m-IgG2b	HI100	Biolegend	304132	Aurora
CD49e	ITGA5	BUV737	m-IgG1	IIA1	BD	741849	Aurora
CD62L	SELL	BV605	m-IgG1	DREG-56	Biolegend	304834	Aurora
CD69	EA-1	BUV805	m-IgG1	FN50	BD	748763	Aurora
CD127	IL7Ra	PE-Cy5.5	m-IgG1	eBioRDR5	eBioscience	35-1278-42	Aurora
CD158a+h	KIR2DL1, KIR2DS1	PE	m-IgG1	EB6	BC	A09778	Aurora
CD158b1 +b2+j	KIR2DL2, KIR2DL3, KIR2DS2	PE	m-IgG1	GL183	BC	IM2278U	Aurora
CD158e1	KIR3DL1	PE	m-IgG1	DX9	BD	555967	Aurora
CD158i	KIR2DS4	PE	m-IgG2a	FES172	BC	IM3337	Aurora
CD159a	NKG2A	PE	m-IgG2b	Z199	BC	IM3291U	Aurora
CD159a	NKG2A	PE-Cy7	m-IgG2b	Z199	BC	B10246	Aurora
CD335	NKp46	Unconjugated	m-IgG1	9E2	BD	557911	Aurora

<<< Table S1. continued

Specificity		Antibody characteristics					
CD designation	Alternative name	Fluorochrome	Type	Clone	Company	Catalog#	Panel
	CCL4	PerCP-ef710	m-IgG2a	FL34Z3L	eBioscience	46-7540-42	Aurora
	CCL5	PE	m-IgG2b	VL1	Biolegend	515504	Aurora
	CX3CR1	BV786	r-IgG2b	2A9-1	BD	744489	Aurora
	GZMB	AF700	m-IgG1	GB-11	BD	560213	Aurora
	GZMK	AF647	m-IgG1	GM26E7	Biolegend	370503	Aurora
	IFNG	PE-Dazzle594	m-IgG1	4S.B3	Biolegend	502546	Aurora
	KLRG1	PerCP-ef710	m-IgG2a	13F12F2	eBioscience	46-9488-42	Aurora
	TIGIT	AF647	m-IgG1	MBSA43	eBioscience	51-9500-42	Aurora
	TNFa	AF700	m-IgG1	Mab11	Biolegend	502928	Aurora
	XCL1	Unconjugated	g-IgG	polyclonal	R&D	AF695	Aurora
Secondary	g-IgG	AF488	d-IgG	polyclonal	Invitrogen	A-11055	Aurora
Secondary	m-IgG1	AF594	g-IgG	polyclonal	Invitrogen	A-21125	Aurora

Main abbreviations: Type: m=mouse, g=goat, rec=recombinant, h=human, r=rat, d=donkey, Company: BD=Becton Dickinson Biosciences, BC=Beckman Coulter, Panel: Au=Aurora.

Table S2. Oligonucleotide-conjugated antibodies used for single-cell RNA sequencing

Specificity		Antibody characteristics				
CD designation	Alternative name	Oligonucleotide	Type	Clone	Company	Catalog#
CD34		GCAGAAATCTCCCTT	m-IgG1	581	Biolegend	343537
CD56	NCAM	TCCTTTCCTGATAGG	m-IgG1	5.1H11	Biolegend	362557
CD117	c-KIT	AGACTAATAGCTGAC	m-IgG1	104D2	Biolegend	313241
CD186	CXCR6	GACAGTCGATGCAAC	m-IgG2a	K041E5	Biolegend	356021

Main abbreviations: Type: m=mouse.

2. Supplemental figures

Figure S1. Quality control and preprocessing single-cell RNA sequencing (next page >>>)

A) *Bone marrow mononuclear cells (BMMC) were negatively enriched for NK cells and labeled with oligonucleotide-conjugated antibodies (O-AB) against CD56, CXCR6, CD117 and CD34. By flow cytometry the composition of each fraction was tested. By using goat-anti-mouse (GaM) the binding of the oligonucleotide-conjugated antibodies was validated.* ***B)*** *Low-quality cells with <1000 expressed genes, >8300 expressed genes and >12.5% mitochondrial RNA were removed from further analyses (in total 21 cells). 6979 cells were included in subsequent analyses.* ***C)*** *After log-transformation and normalization, the top 2000 most variable genes were selected by computing the variance of the log-counts and fitting a trend to the variance with respect to abundance across the genes. The top 10 most variable genes are marked.* ***D)*** *Principal component (PC) analysis was applied based on the top 2000 most variable genes. The top 20 PCs, explaining the most variance, were retained for further analyses.*

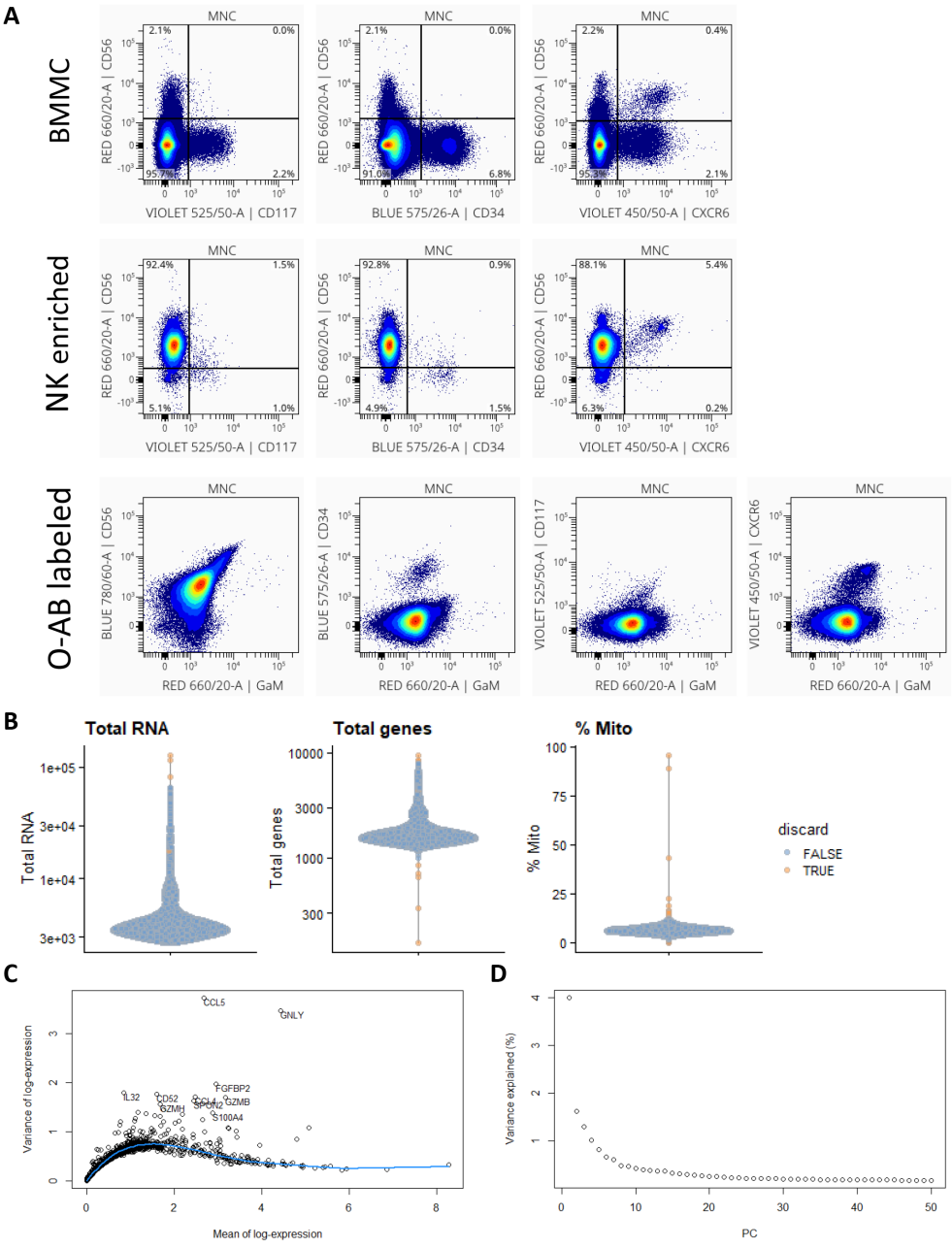

Figure S1. (<<< legend at previous page)

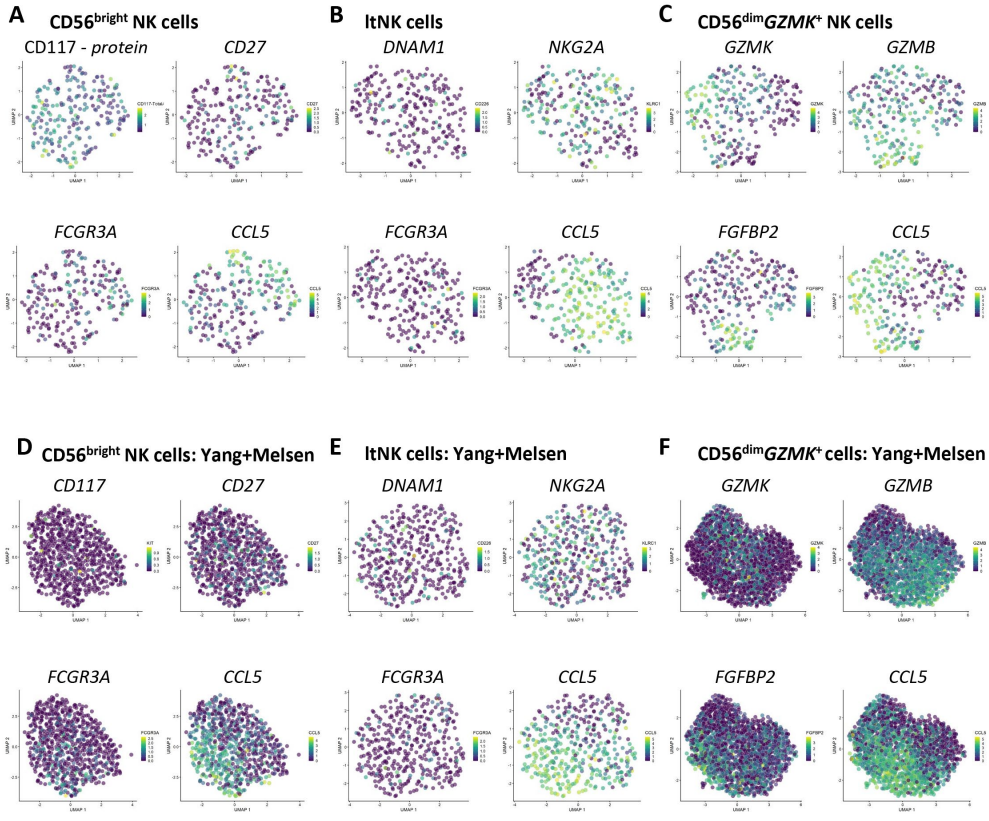

Figure S2. Subclustering of CD56^{bright}, ltNK and CD56^{dim}GZMK^{+} cells does not reveal additional clusters

A) *CD56^{bright} NK cells (n=225),* ***B)*** *ltNK cells (n=255) and* ***C)*** *CD56^{dim}GZMK^{+} NK cells (n=262) were selected and further subclustered based on the first 20 principal components. In both populations only 1 cluster was identified. Genes which are coding for proteins that are only expressed by a fraction of the cells at protein level, and CCL5 are shown.* ***D)*** *To increase the number of cells, our bone marrow NK cells were integrated with single-cell RNA sequencing data from sorted NK cells (total 9071) from eight donors (two blood, six bone marrow samples) of the Yang et al. dataset. The CD56^{bright} NK cells (n=771),* ***E)*** *ltNK cells (n=407) and* ***F)*** *CD56^{dim}GZMK^{+} NK cells (n=2132) were extracted. The UMAP was based on the corrected principal components scores. No additional clusters were identified.*

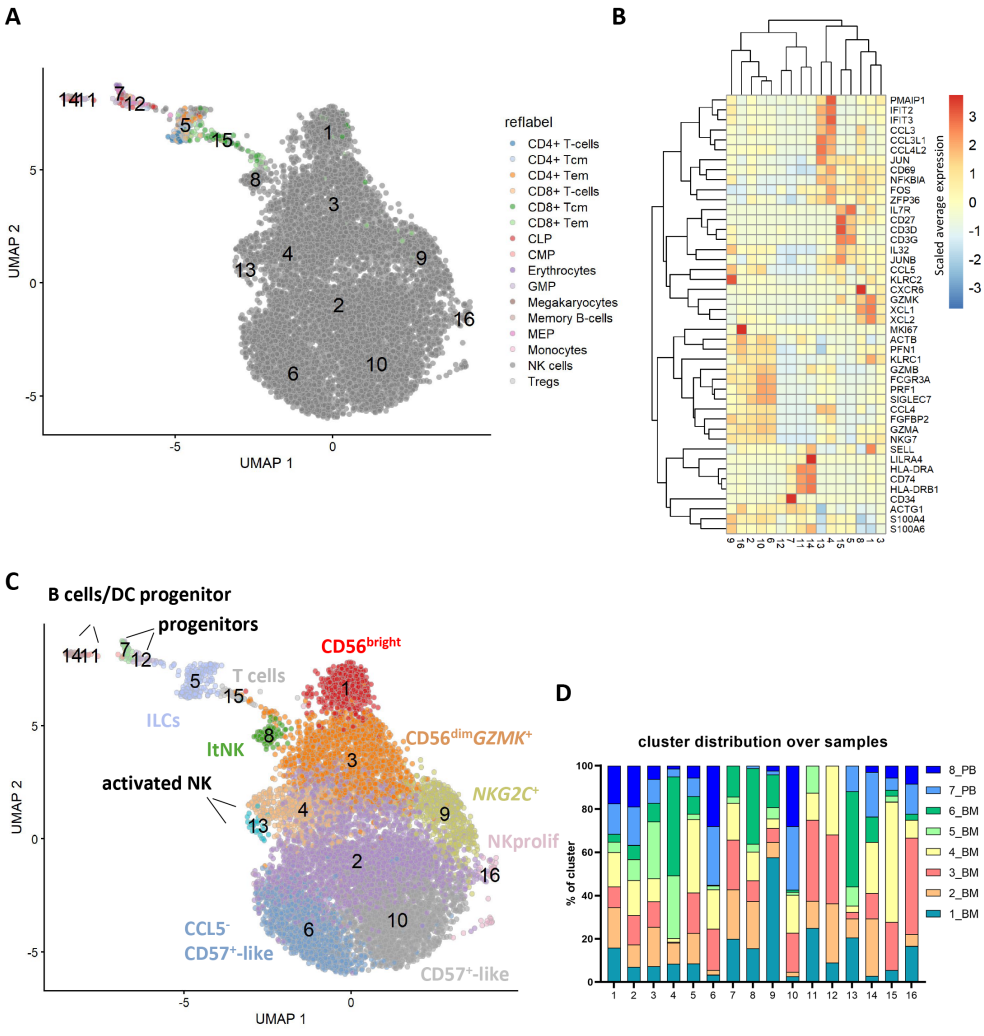

Figure S3. Analysis of the single-cell RNA sequence dataset from Yang et al.

A) *Clustering of the single-cell RNA sequencing dataset (GSE130430) containing 8 donors resulted in 14 clusters. Reference based analysis was performed by using the Blueprint and Encode datasets, containing bulk RNA sequenced hematopoietic populations.* ***B)*** *A selection of genes used to define the clusters are included in the heatmap. The scaled average expression is shown.* ***C)*** *Cluster definitions based on A and B.* ***D)*** *Distribution of individual samples per cluster.*

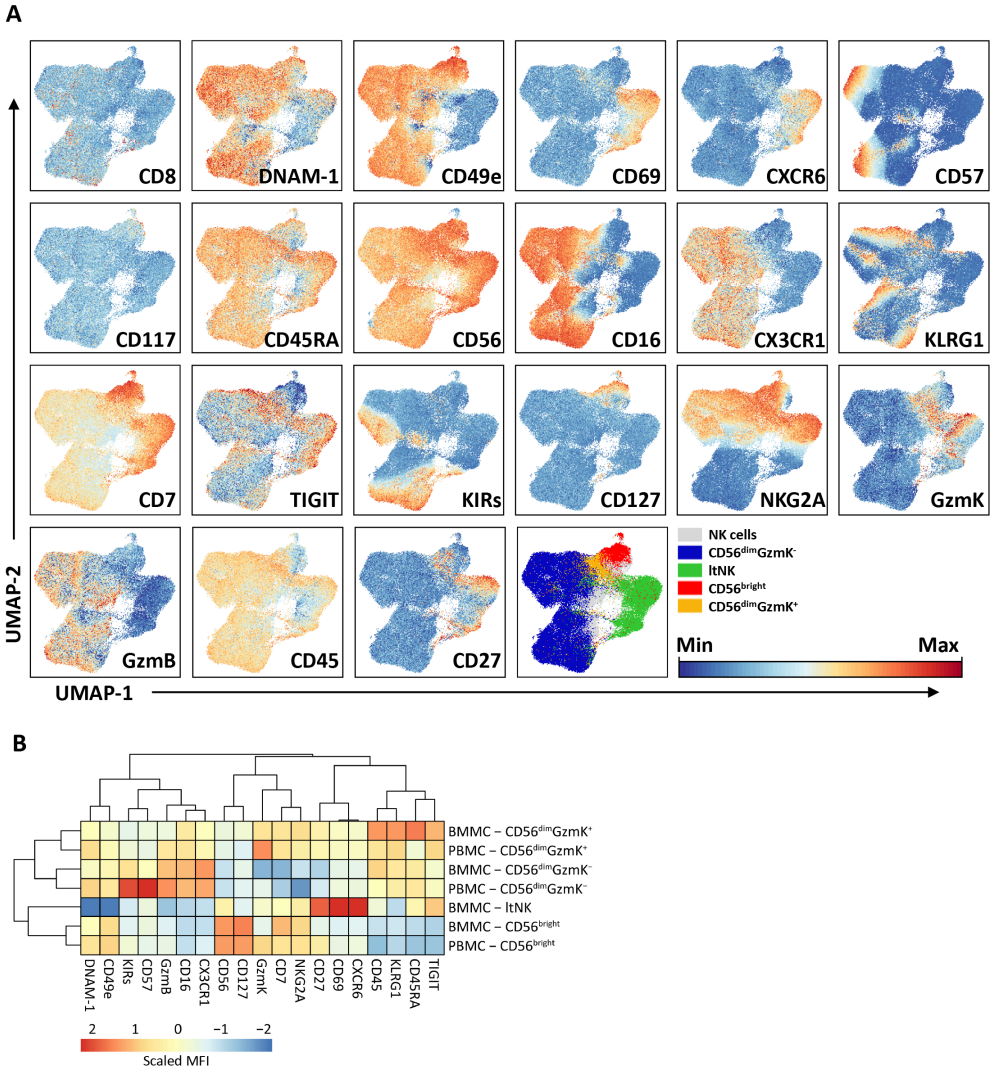

Figure S4. UMAP embedding of spectral cytometry data of 14 bone marrow samples of healthy controls

A) *25-marker panel for spectral cytometry was developed and NK cells were selected as $CD3^-CD14^-CD19^-CD34^-CD45^+CD56^+CD45RA^+$. A UMAP embedding of NK cells derived from 14 healthy bone marrow samples, based on 21 NK cell markers is shown. ltNK cells were defined as $CD69^+CD49e^-$. The non-ltNK cells were subdivided into $CD56^{bright}CD16^{+/-}$ and $CD56^{dim}CD16^+$ NK cells.* ***B)*** *The heatmap indicates the scaled mean fluorescence intensity (MFI) expression of NK cell markers on the NK cell subsets in blood (3 donors) and bone marrow (14 donors).*

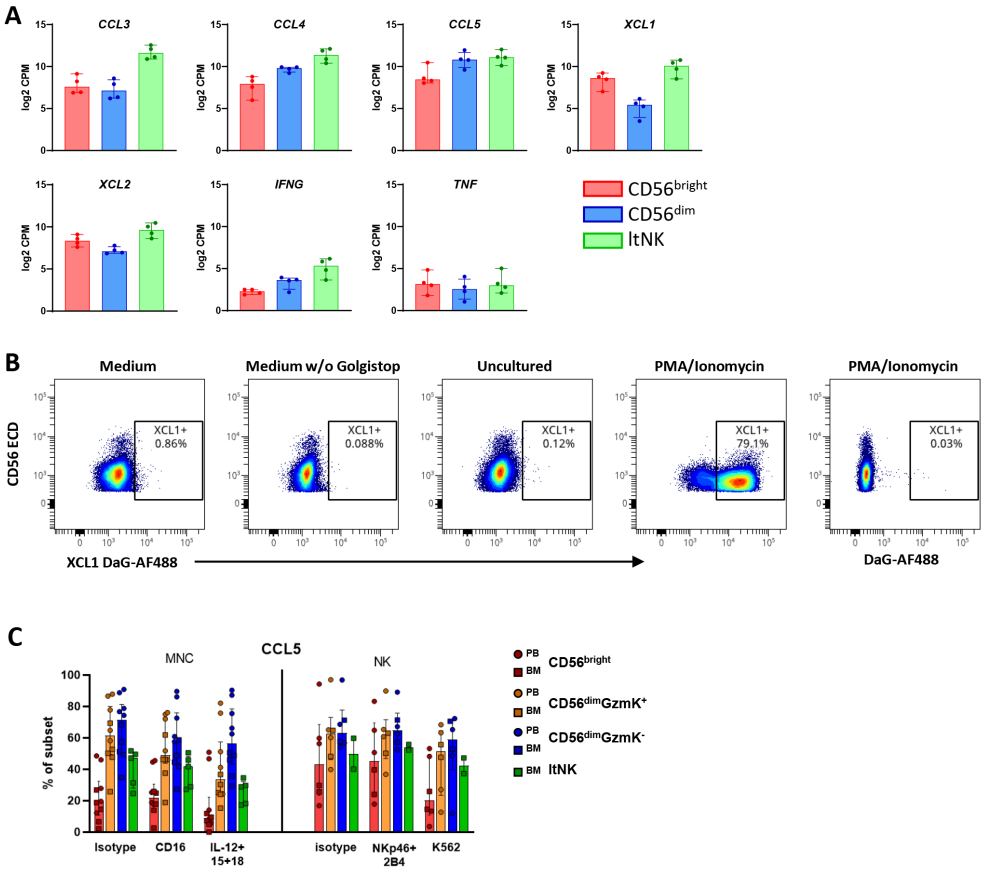

Figure S5. Chemokine and cytokine production of NK cell subsets

A) *Previously published bulk RNA sequence data (GSE116178) showed that each NK cell subset in bone marrow expresses CCL3, CCL4, CCL5, XCL1, XCL2, IFNG and TNF at the mRNA level (n=4). Bars indicate median and range. CPM = counts per million reads* ***B)*** *XCL1 was intracellularly detected by a two-step staining protocol, using unconjugated goat anti-XCL1 and donkey-anti-goat (DaG) AF488. As a negative control, cells were stained with only DaG-AF488 (most right panel). NK cells from a representative blood donor are shown.* ***C)*** *Intracellular CCL5 expression in NK cell subsets, either stimulated in bulk mononuclear cells (MNC, n=5 blood, n=5 bone marrow) or as enriched NK cell fraction (n=4 blood, n=2 bone marrow). PB=peripheral blood, BM=bone marrow.*

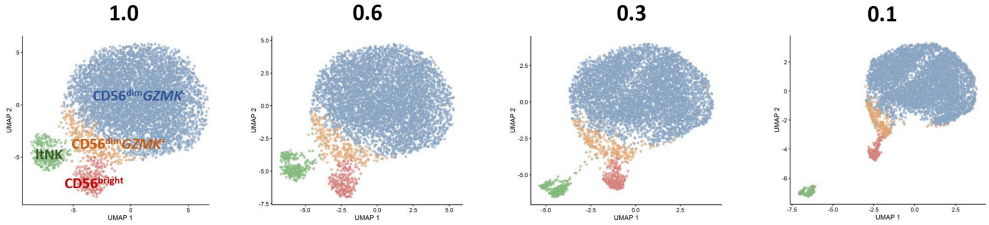

Figure S6. NK cell development

The minimum distance parameter of UMAP (in this article set at 1.0) was adjusted to study the effect on the embedding. With a lower minimum distance, the CD56^{bright} NK cells were still connected to the CD56^{dim} NK cells, while the ltNK cells were not.

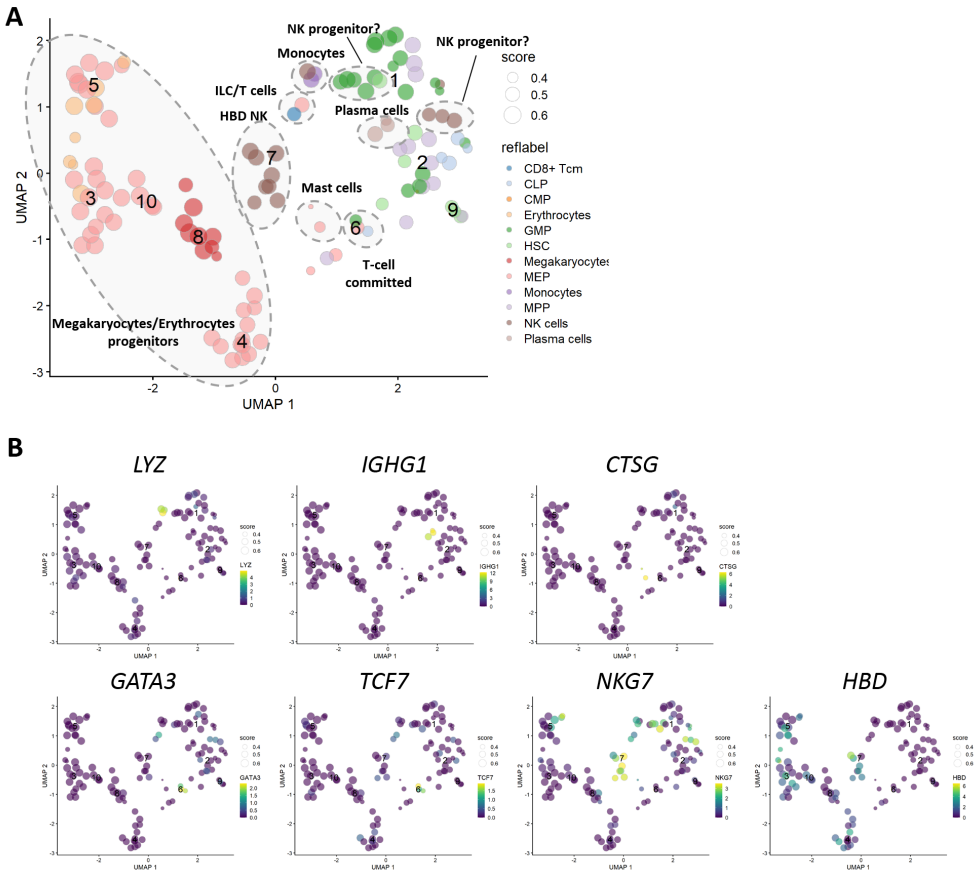

Figure S7. Subclustering of CD34+ population

A) *The CD34^{+} cells in our dataset were selected and subclustered. In addition, reference-based analysis was performed based on bulk RNA sequence data of hematopoietic populations. The size of the cells reflects the probability score (the bigger, the more likely the assignment of the reference label is correct). The HBD^{+}NKG7^{+} cells in cluster 7 are likely to represent NK cells contaminated with lysed erythrocytes.* ***B)*** *The expression of a selection of genes is indicated on the UMAP.*

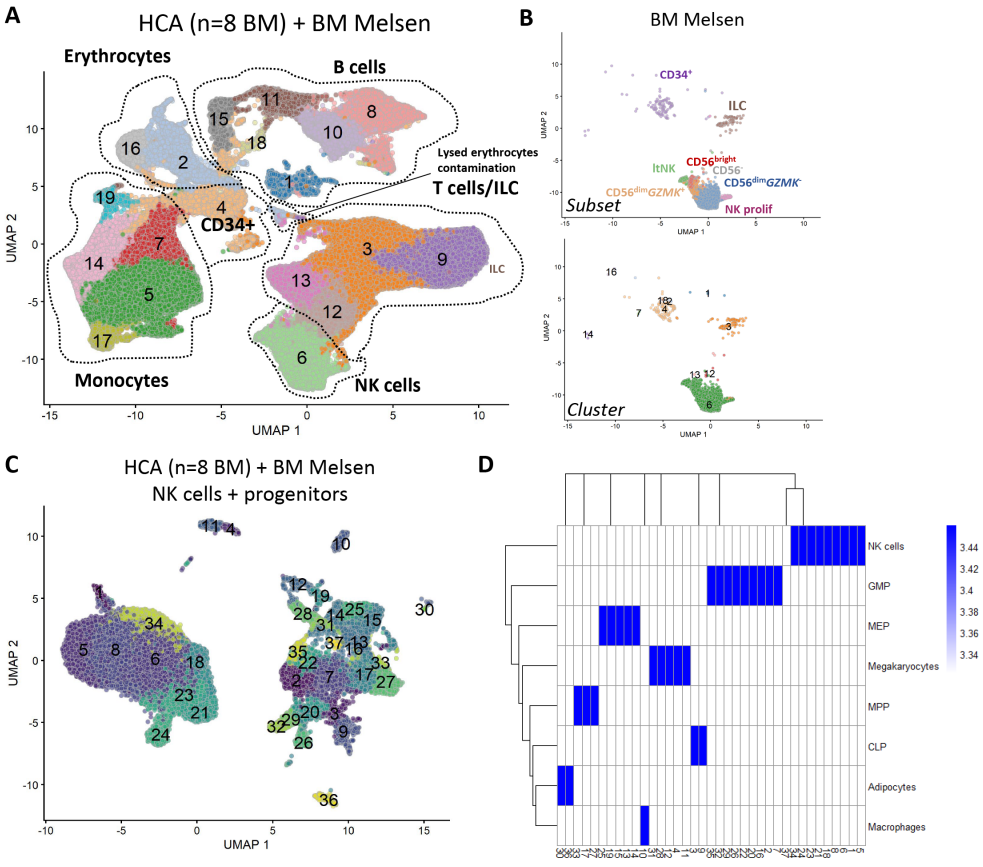

Figure S8. Integration of NK cells with bone marrow cells from the human cell atlas

A) *The bone marrow cells of 8 donors from the human cell atlas (HCA) single-cell RNA sequence dataset were integrated with our bone marrow NK cell dataset. Each color on the UMAP represents a distinct cluster.* ***B)*** *Cells from our bone marrow NK cell dataset were selected and plotted in the same UMAP embedding as in A. Cells were colored by their original subset assignment, or cluster assignment as indicated in A.* ***C)*** *Cluster 4 and 6 were selected and re-clustered.* ***D)*** *Reference-based analysis using the bulk RNA sequence Blueprint and Encode datasets was performed on the clusters in C. The MEP (megakaryocyte-erythroid progenitor), megakaryocytes, adipocytes and macrophages were removed from further analyses.*

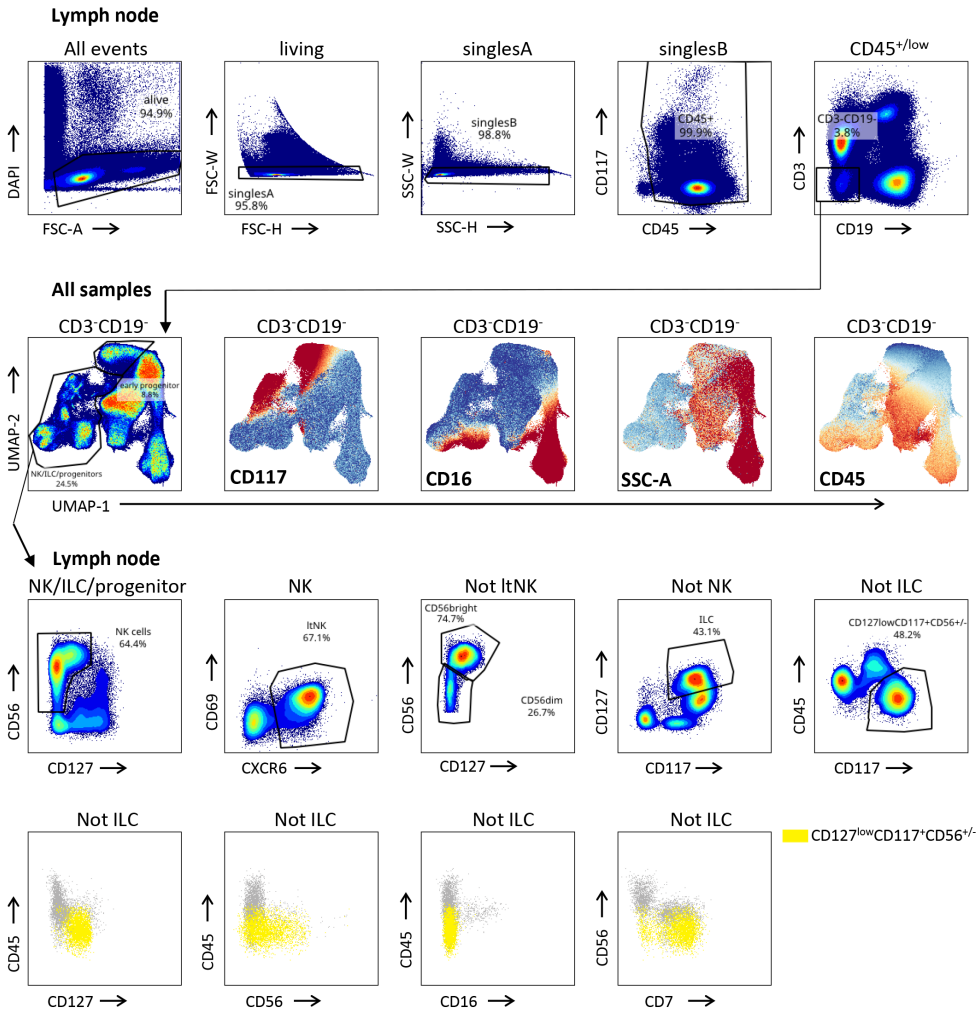

Figure S9. NK cell development in secondary lymphoid tissues

Gating strategy applied to study presence of NK progenitor cells in bone marrow, spleen, lymph node and tonsil. Dead and doublet cells were excluded. CD45$^{+/low}$CD3^{-}CD19^{-} cells of all samples were gated and embedded in a UMAP. In the UMAP, early progenitors (CD117^{+}CD16^{-}SSC^{high}CD45^{low}) and a combination of NK, ILC and other progenitors were selected. Within the latter population, NK cell subsets, ILC and the CD127^{low}CD117^{+}CD56$^{+/-}$ potential NK progenitor cells were defined. A representative lymph node donor is shown in the upper and bottom plots.