Supplemental Methods

Cell lines

Daudi (RRID:CVCL_0008) and derivatives transfected to express MICA or β_2m[1] were maintained in culture media (RPMI-1640 (Sigma Aldrich), 10% fetal calf serum, 2 mM L-glutamine, 50 U/ml penicillin streptomycin (Gibco)) at 37°C, 5% CO_2.

Functional Assays: CD107a degranulation and IFN- γ ELISA

For degranulation assays, unstimulated or IL-2 stimulated NK cells were co-incubated with target cells (Daudi, Daudi-MICA, Daudi-β_2m, or Daudi pre-coated with 10μg/ml rituximab for 30 min at 37°C), at 1:1 ratio in the presence of monensin (GolgiStop, 1/1,000 dilution: BD Biosciences) and PE-labelled anti-Lamp-1 mAb (RRID:AB_396135) or isotype matched control mAb (PE-labelled IgG1 RRID:AB_326435) for 5 hours at 37°C and 5% CO_2. After incubation, cells were washed and stained with viability dye (Zombie UV; Biolegend), FITC-labelled anti-CD3 mAb (RID:AB_314060), APC-labelled anti-CD19 mAb (RRID:AB_314242), BV711-labelled anti-CD56 mAb (RRID:AB_2562417) and PE-labelled anti-LAMP-1 mAb or isotype matched control mAbs (FITC-labelled IgG1k (RRID:AB_326429), and APC-labelled IgG1k (RRID:AB_326443**)).** Finally, cells were washed in 1% FBS/PBS, fixed in 2% PFA/PBS for 10 min then assessed by flow cytometry (BD LSR Fortessa flow cytometer), and analysed (FlowJo V10 software).

For IFN-γ ELISA assay, flat-bottomed 96-well plates were coated with 0.01% poly-L-lysine (Sigma Aldrich) according to the manufacturer's instructions. Wells were then coated with the following proteins in PBS: anti-CD16 mAb (5μg/ml; RRID:AB_395804:), MICA-Fc (3μg/ml; 1300-MA-050, R&D Systems), anti-NKp30 mAb (1μg/ml; RRID:AB_2149445) or isotype matched controls (Biolegend), all with ICAM-1 (2.5μg/ml; ADP4-050, R&D Systems); or ICAM-1 alone. These quantities had been shown to induce robust activation of NK cells. After overnight incubation at 4°C, plates were washed

in PBS and then blocked in culture media for 5 min. This was removed and 50,000 NK cells in culture media were added per well, in triplicate. After overnight culture at 37°C, cell supernatants were harvested by centrifugation at 350 x*g* for 10 min at 4°C. IFN-γ secretion was quantified from the supernatants by sandwich ELISA as previously described.[2]

NK cell isolation and staining for ILCs

Primary human NK cells were isolated from the peripheral blood of healthy donors acquired from the National Health Service blood service using negative magnetic bead selection (Miltenyi Biotec). Ethical approval was obtained from the Research Ethics Committee 05/Q0401/108 and 2017-2551-3945 (University of Manchester). After isolation, NK cells were plated into media (Dulbecco's Modified Eagle Medium, 30% F12 Ham, 10% human serum, 1% non-essential amino acids, 1mM sodium pyruvate (Sigma), 2mM L-glutamine, 50U/ml penicillin streptomycin, 50μM 2-mercaptoethanol (Gibco)) supplemented with 1ng/ml IL-15 overnight. To test for the presence of ILCs, NK cells were washed in 1% FBS/PBS and blocked with 1% human AB serum (Sigma-Aldrich) for 15 min at 4°C before staining. NK cells were stained for 20 min at 4°C with viability dye (Zombie UV; Biolegend), PE-labelled anti-CD56 mAb (RRID:AB_395906) and BV785-labelled anti-CD127 mAb (RRID:AB_2563605). Cells were then washed in 1% FBS/PBS and fixed in 2% PFA/PBS for 15 min. Cells were then washed in 1% FBS/PBS and permeabilsed using Triton X-100 (0.01%) for 20 min at room temperature before staining for AF488-labelled anti-PRF1 mAb (δG9; BD Biosciences) or isotype matched control mAb (AF488-lablelled IgG2b (MPC-11; Biolegend)) for 30 min. Finally cells were washed in 1% FBS/PBS and assessed by flow cytometry (BD LSR II flow cytometer).

Processing of scRNA-seq data

Raw FASTQ files were processed using 10X Genomics CellRanger software (v 2.0.1) in-order to map reads against the human hg38 reference genome, filter out unexpressed genes, and count barcodes and unique molecular identifiers (UMIs). After alignment, more in-depth quality control,

normalisation and analysis were conducted using the Bioconductor package Seurat (v 2.3.0)[3] in R (v 3.3). Only data from high quality cells was used for further analysis.

Filtering to remove low quality cells was conducted per datasets using the following features: 1) genes expressed in less than 3 cells; 2) cells with less than 200 genes expressed and 3) cells expressing a high proportion of mitochondrial-associated genes; 4) gene count (nGene) and 5) UMI count; the thresholds of which were based on manual interpretation of the plots (Supplemental Table 1). Following filtering, normalisation was performed using the Seurat global-scaling normalisation method. Briefly, gene expression values for each cell were divided by the total number of transcripts and multiplied by 10,000. These values are then natural-log transformed using log1p for further downstream analyses. After normalization, scaling was performed to regress out cell-cell variation driven by nUMI and the percentage of mitochondrial-associated genes. To do this, scaled z-score residuals were calculated for each gene for downstream dimensionality reduction

Effects of cell cycle

Using the "CellCycleScoring" command, cells were assigned a quantitative score using the expression of 94 cell cycle genes. Principal component analysis (PCA) was then run using the cell cycle genes to visualise cell separation.

Data Analysis of CMV-negative and CMV-positive donors

To identify similarities between blood NK cells between individuals who tested positive or negative for CMV, data were obtained from two CMV-positive healthy donors and an additional CMV-negative healthy donor. Each dataset was processed individually using the same pipeline outlined previously. For comparison with our original analyses, 8,000 random cells (1,600 per donor) from the additional donors and our two original CMV-negative donors (unstimulated cells only) were taken forward for the combined analyses. Due to inherent variability between donors and to

avoid donor specific clusters, we used the Seurat alignment strategy to identify NK cell subsets common across the donors. Briefly, 1,723 genes (highly variable in >2 datasets) were used in the CCA. The CCA subspace was then aligned and cells were clustered using a shared nearest neighbour (SNN) modularity optimisation-based clustering algorithm and visualised using tSNE. Clustering was done using twenty components from the CCA dimensionality reduction, identified by means of the shared correlation strength and using a resolution of 0.7.

Supplemental References

1. Eleme K, Taner SB, Önfelt B, et al. Cell Surface Organization of Stress-inducible Proteins ULBP and MICA That Stimulate Human NK Cells and T Cells via NKG2D. *J. Exp. Med.* 2004;199(7):1005–1010.

2. Oszmiana A, Williamson DJ, Cordoba SP, et al. The Size of Activating and Inhibitory Killer Ig-like Receptor Nanoclusters Is Controlled by the Transmembrane Sequence and Affects Signaling. *Cell Rep.* 2016;15(9):1957–1972.

3. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 2018;36(5):411–420.

4. Yu J, Mao HC, Wei M, et al. CD94 surface density identifies a functional intermediary between the $CD56^{bright}$ and $CD56^{dim}$ human NK-cell subsets. *Blood*. 2010;115(2):274–282.

5. Oki T, Nishimura K, Kitaura J, et al. A novel cell-cycle-indicator, mVenus-p27K-, identifies quiescent cells and visualizes G0-G1 transition. *Sci. Rep.* 2014;4:.

6. Hernandez-Segura A, de Jong T V., Melov S, et al. Unmasking Transcriptional Heterogeneity in Senescent Cells. *Curr. Biol.* 2017;27(17):2652–2660.e4.

7. Hanna J, Bechtel P, Zhai Y, et al. Novel Insights on Human NK Cells' Immunological Modalities Revealed by Gene Expression Profiling. *J. Immunol.* 2004;173(11):6547–6563.

Supplemental Table Legends

Table S1. Individual filtering thresholds. Filtering threshold used per individual datasets during QC and the number of cells filtered out.

Table S2. Variable genes. Individual thresholds used to determine the number of variable genes within the individual datasets.

Table S3. Conserved markers associated with each NK cluster. Spreadsheets of conserved markers associated with each cluster.

Supplemental Figure Legends

Supplemental Figure 1. IFN-γ assay to determine functional responses of NK cells sequenced. IFN-γ secretion (mean± SD) from NK cells was quantified by enzyme-linked immunosorbent assay.

Supplemental Figure 2. Clustering is not influenced by donor variability or cell cycle. (A) t-distributed stochastic neighbour embedding (tSNE), two-dimensional (2D) plot colour coding the individual cells according to which donor and condition they originated from. (B) Quantification showing the percentage of cells within each cluster across the donors and conditions. (C) Quantification showing the percentage of cells within different phases of the cell cycle across the clusters (calculated individually for unstimulated and IL-2 stimulated cells).

Supplemental Figure 3. A strong correlation is observed across all clusters. Correlation analysis using the top 20 markers associated with each cluster (including the excluded clusters 5, 7 and 9) and using all cells (unstimulated and stimulated).

Supplemental Figure 4. Diffusion map dimensionality reduction. Diffusion map to show the underlying structure within the dataset by using non-linear dimensionality reduction to preserve continuous trajectories not observed with linear methods. Cells are colour coded by cluster identity (excluding clusters 5, 7 and 9).

Supplemental Figure 5. Cluster 4 is not a population of non-NK cell ILCs. (A) Flow cytometry plots of negatively isolated human NK cells, showing the expression of CD56 and CD127 on gated live, singlet cells and then showing the expression of intracellular perforin on the $CD56^{+}CD127^{+}$ sup-population. (B) Expression distribution of each cluster (violin plots) using unstimulated cells only, looking at ILC specific markers. The shape represents the distribution of cells based on their log(+1) expression values (*y* axis). The colour scale represents the mean expression. (C) Dotplot (top) of the same ILC

specific markers within unstimulated cells only (columns) across the different clusters (rows). The size of the dot represents the percentage of cells expressing the markers whilst the colour encodes the average scaled expression values.

Supplemental Figure 6. The bright-like signature of cluster 2 is primarily driven by 5 genes. tSNE plots (excluding clusters 5, 7 and 9), showing the expression of genes driving the $CD56^{bright}$ association within cluster 2 (using full dataset). Expression is colour coded from blue (low) to red (high) and cells positively expressing a marker were brought towards the front of the plot.

Supplemental Figure 7. Cluster 6 is not a population of intermediate $CD56^{dim}CD94^{high}$ NK cells. Heatmap of markers which differentiate intermediate $CD56^{dim}CD94^{high}NKG2A^{+}$ NK cells from $CD56^{dim}$ and $CD56^{bright}$ NK cells using flow cytometry[4]. Clusters 0, 1 and 2 were combined and termed "Grouped $CD56^{dim}$"; cluster 4 was termed "$CD56^{bright}$"; clusters 3 and 8 were excluded from this analysis. NK cell "groups" are plotted in columns and markers of interest are shown in rows. Gene expression is colour coded using average scaled expression values per group, based upon z-sore distributions, ranging from low expression (purple) to high expression (yellow).

Supplemental Figure 8. Cluster 8 is a novel population of NK cells. (A) The distribution of percentage of the number genes expressed (nGene) and the number of unique molecular identifiers (nUMI). The shapes represent the distribution whilst the error bars represent median and interquartile range. (B) Module scores for each NK cell cluster at the single cell level calculated using markers of quiescence[5] (left) and markers of scenescence[6] (right) The shapes represent the distribution whilst the error bars represent median and interquartile range. (C) The distribution of the percentage of mitochondrial mRNA across clusters; using unstimulated cells only (left). Module scores for each NK cell cluster at the single cell level calculated using pro-apoptotic markers (right). Module scores were calculated for unstimulated cells only. Violin plots represent the distribution

for each cluster whilst the error bars represent median and interquartile range. (D) Comparison of average gene expression values for cluster 8 between unstimulated (x-axis) and IL-2 stimulated (y-axis) cells. Genes with a fold-change > 0.5 and Bonferroni corrected p-value <0.05 are highlighted.

Supplemental Figure 9. Overview of identified clusters from joint analysis of NK cells isolated from CMV negative and CMV positive donors. (A) Module scores for each NK cell cluster calculated at the single cell level, defined using the top 100 markers from bulk expression profiles of sorted $CD56^{dim}CD16^{+}$ and $CD56^{bright}CD16^{-}$ NK cells.[7] $CD56^{dim}$ module score (left) and custom $CD56^{bright}$ module score, *excluding CTSW, DUSP1, JUN, FOS* and *CD69* (right). Violin plots represent the distribution of the module scores for each cluster whilst the error bars represent median and interquartile range. (B) Heatmap of canonical $CD56^{bright}$ NK cell markers. Clusters are plotted in columns and genes are shown in rows. Gene expression is colour coded using average scaled expression values per cluster, based upon a z-sore distribution, ranging from low expression (purple) to high expression (yellow). (C)) Heatmap of top 10 markers distinguishing each NK cell cluster, identified by differential expression analysis and showing a maximum of 500 genes per cluster; excluding stressed cells and APCs. Cells are plotted in columns and genes are shown in rows. Gene expression is colour coded using a scale based on z-score distribution. (D) top markers distinguishing inflamed type 1 IFN NK cells, (E) top markers distinguishing CIML NK cells and (F) top markers distinguishing low ribosomal NK cells and IEGs. Clusters are plotted in columns and genes are shown in rows. Gene expression is colour coded using average scaled expression values per cluster, based upon a z-sore distribution, ranging from low expression (purple) to high expression (yellow).

Supplemental Table 1: Filtering thresholds determined by manual interpretation per dataset

	DonorA unstim		DonorA + IL2		DonorB unstim		DonorB + IL2	
Thresholds used	**min**	**max**	**min**	**max**	**min**	**max**	**min**	**max**
mitochondrial genes expression	-	10%	-	8%	-	8%	-	8%
nGENE	200	-	200	3,000	200	-	200	-
nUMI	-	8,000	-	8,000	-	7,500	-	7,000
Cells filtered out	**42**		**40**		**47**		**57**	

Supplemental Table 2: Thresholds used to determine the number of variable genes within each dataset

	DonorA unstim		DonorA + IL2		DonorB unstim		DonorB + IL2	
Thresholds used	**min**	**max**	**min**	**max**	**min**	**max**	**min**	**max**
Variable gene expression	0.025	5	0.025	4	0.025	5	0.025	6
Number variable genes	**2,582**		**2,563**		**2,328**		**2,118**	

Figure S1.

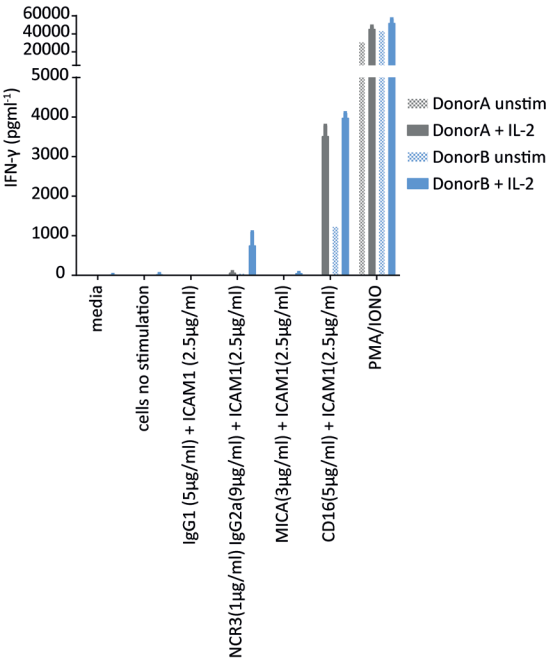

60000
40000
20000
5000
4000
3000
2000
1000
0
IFN-γ (pgml-1)
DonorA unstim
DonorA + IL-2
DonorB unstim
DonorB + IL-2
media
cells no stimulation
IgG1 (5µg/ml) + ICAM1 (2.5µg/ml)
NCR3(1µg/ml) IgG2a(9µg/ml) + ICAM1(2.5µg/ml)
MICA(3µg/ml) + ICAM1(2.5µg/ml)
CD16(5µg/ml) + ICAM1(2.5µg/ml)
PMA/IONO

Figure S2.

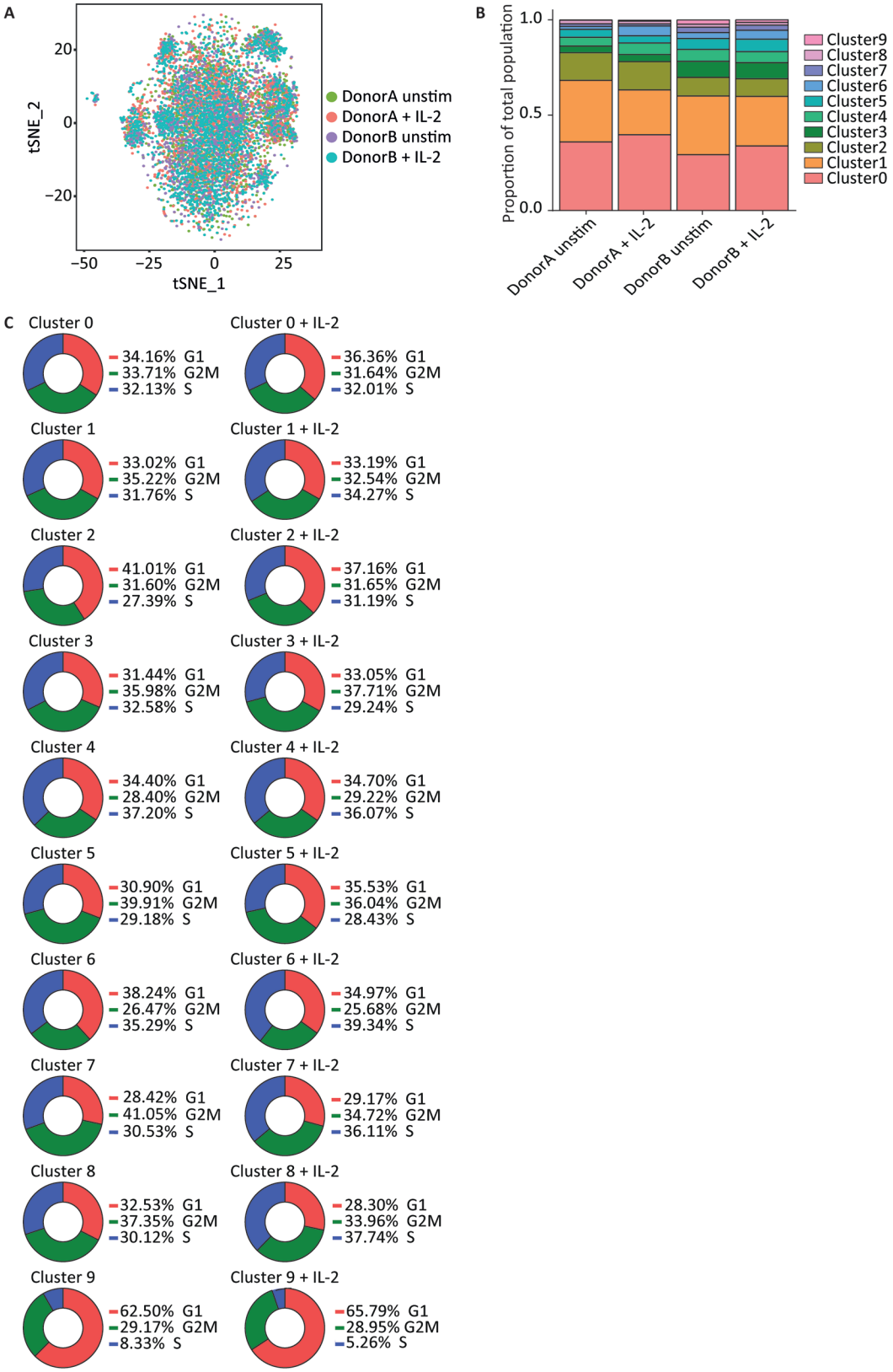

A
tSNE_2
tSNE_1
DonorA unstim
DonorA + IL-2
DonorB unstim
DonorB + IL-2
B
Proportion of total population
Cluster9
Cluster8
Cluster7
Cluster6
Cluster5
Cluster4
Cluster3
Cluster2
Cluster1
Cluster0
DonorA unstim
DonorA + IL-2
DonorB unstim
DonorB + IL-2
C
Cluster 0
34.16% G1
33.71% G2M
32.13% S
Cluster 0 + IL-2
36.36% G1
31.64% G2M
32.01% S
Cluster 1
33.02% G1
35.22% G2M
31.76% S
Cluster 1 + IL-2
33.19% G1
32.54% G2M
34.27% S
Cluster 2
41.01% G1
31.60% G2M
27.39% S
Cluster 2 + IL-2
37.16% G1
31.65% G2M
31.19% S
Cluster 3
31.44% G1
35.98% G2M
32.58% S
Cluster 3 + IL-2
33.05% G1
37.71% G2M
29.24% S
Cluster 4
34.40% G1
28.40% G2M
37.20% S
Cluster 4 + IL-2
34.70% G1
29.22% G2M
36.07% S
Cluster 5
30.90% G1
39.91% G2M
29.18% S
Cluster 5 + IL-2
35.53% G1
36.04% G2M
28.43% S
Cluster 6
38.24% G1
26.47% G2M
35.29% S
Cluster 6 + IL-2
34.97% G1
25.68% G2M
39.34% S
Cluster 7
28.42% G1
41.05% G2M
30.53% S
Cluster 7 + IL-2
29.17% G1
34.72% G2M
36.11% S
Cluster 8
32.53% G1
37.35% G2M
30.12% S
Cluster 8 + IL-2
28.30% G1
33.96% G2M
37.74% S
Cluster 9
62.50% G1
29.17% G2M
8.33% S
Cluster 9 + IL-2
65.79% G1
28.95% G2M
5.26% S

Figure S3.

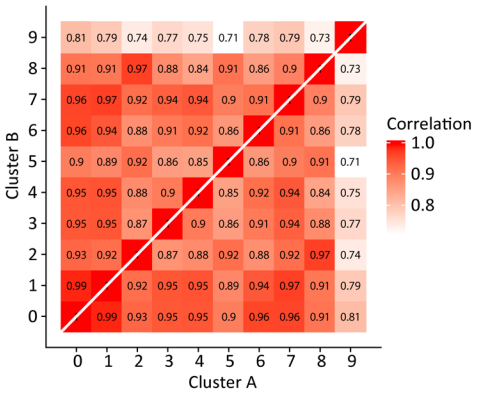

Cluster B
Cluster A
Correlation
1.0
0.9
0.8
0 1 2 3 4 5 6 7 8 9
9 0.81 0.79 0.74 0.77 0.75 0.71 0.78 0.79 0.73
8 0.91 0.91 0.97 0.88 0.84 0.91 0.86 0.9 0.73
7 0.96 0.97 0.92 0.94 0.94 0.9 0.91 0.9 0.79
6 0.96 0.94 0.88 0.91 0.92 0.86 0.91 0.86 0.78
5 0.9 0.89 0.92 0.86 0.85 0.86 0.9 0.91 0.71
4 0.95 0.95 0.88 0.9 0.85 0.92 0.94 0.84 0.75
3 0.95 0.95 0.87 0.9 0.86 0.91 0.94 0.88 0.77
2 0.93 0.92 0.87 0.88 0.92 0.88 0.92 0.97 0.74
1 0.99 0.92 0.95 0.95 0.89 0.94 0.97 0.91 0.79
0 0.99 0.93 0.95 0.95 0.9 0.96 0.96 0.91 0.81

Figure S4.

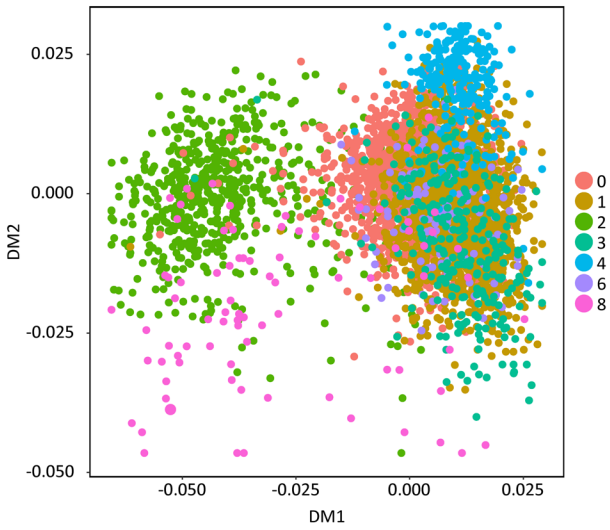


Figure S5.

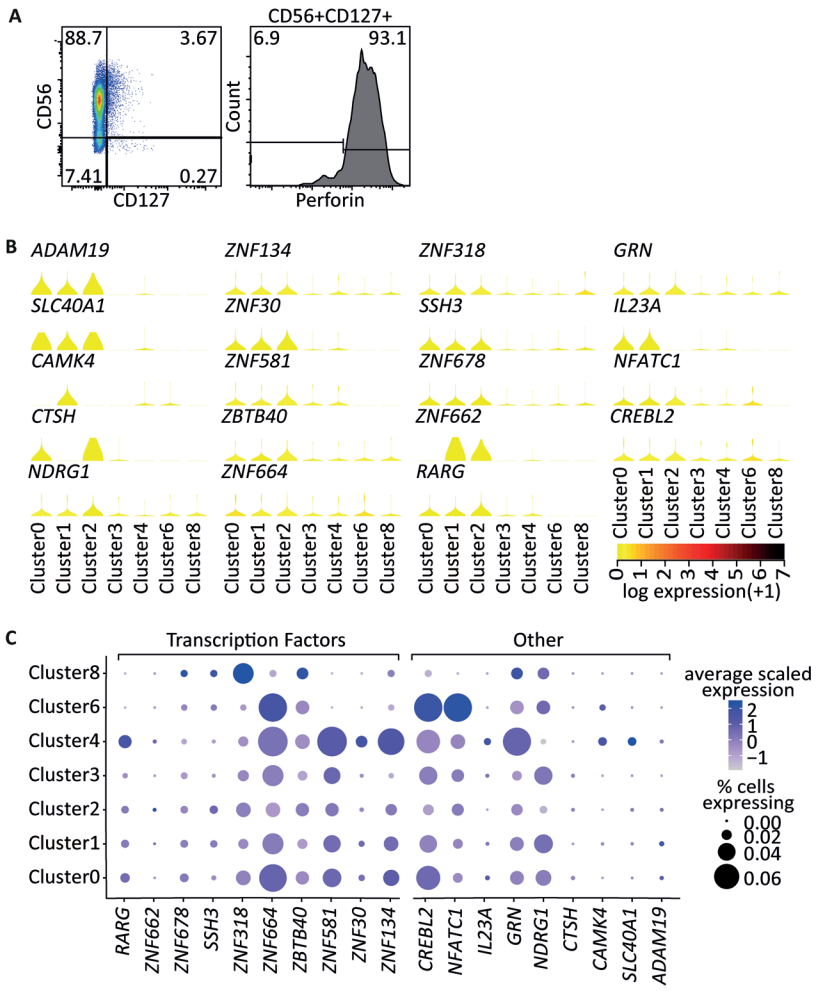

A
CD56+CD127+
88.7
3.67
7.41
0.27
CD56
CD127
Count
6.9
93.1
Perforin
B
ADAM19
SLC40A1
CAMK4
CTSH
NDRG1
ZNF134
ZNF30
ZNF581
ZBTB40
ZNF664
ZNF318
SSH3
ZNF678
ZNF662
RARG
GRN
IL23A
NFATC1
CREBL2
Cluster0
Cluster1
Cluster2
Cluster3
Cluster4
Cluster6
Cluster8
0 1 2 3 4 5 6 7
log expression(+1)
C
Transcription Factors
Other
Cluster8
Cluster6
Cluster4
Cluster3
Cluster2
Cluster1
Cluster0
RARG
ZNF662
ZNF678
SSH3
ZNF318
ZNF664
ZBTB40
ZNF581
ZNF30
ZNF134
CREBL2
NFATC1
IL23A
GRN
NDRG1
CTSH
CAMK4
SLC40A1
ADAM19
average scaled expression
2
1
0
−1
% cells expressing
0.00
0.02
0.04
0.06

Figure S6.

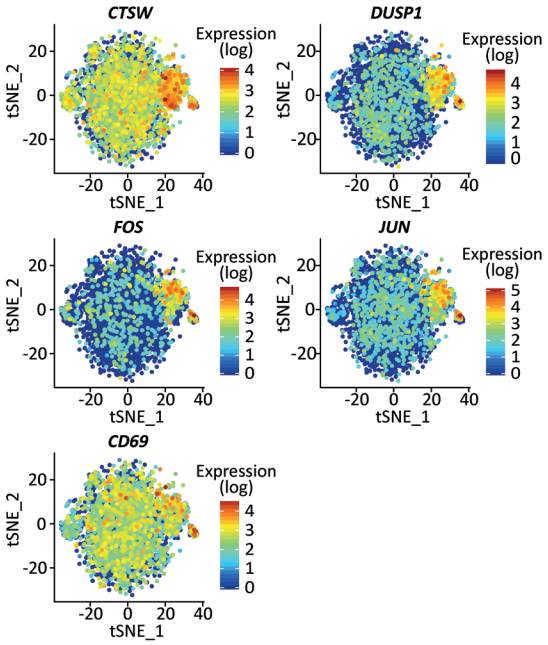


Figure S7.

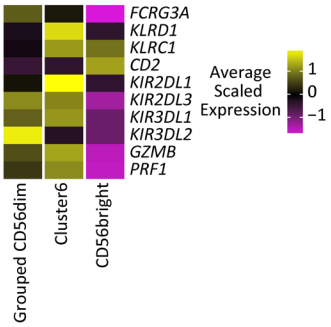

FCRG3A
KLRD1
KLRC1
CD2
KIR2DL1
KIR2DL3
KIR3DL1
KIR3DL2
GZMB
PRF1
Average Scaled Expression
1
0
−1
Grouped CD56dim
Cluster6
CD56bright

Figure S8.

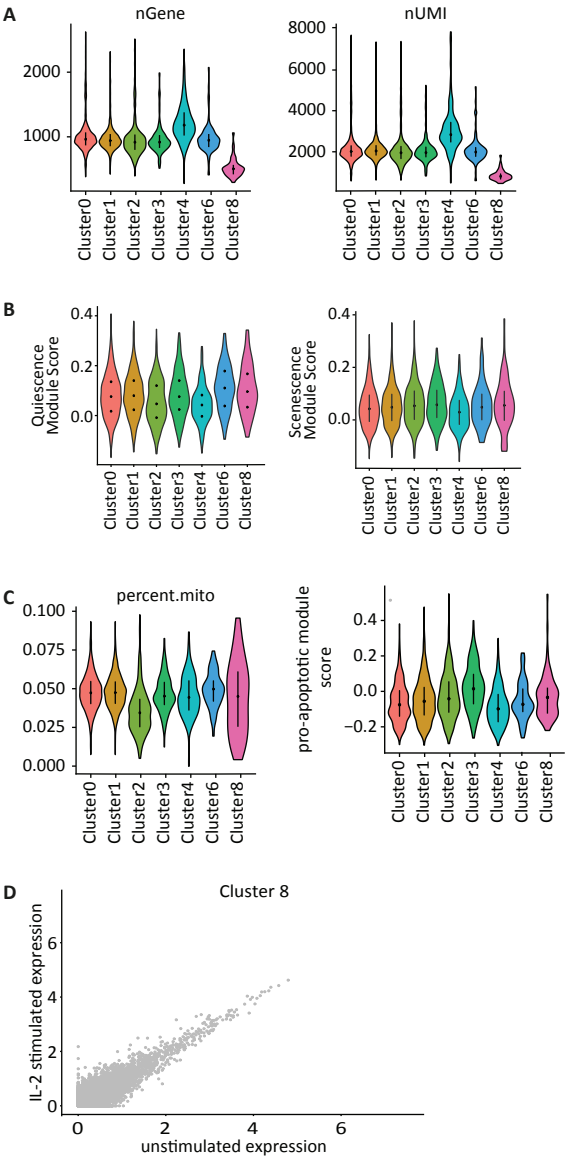
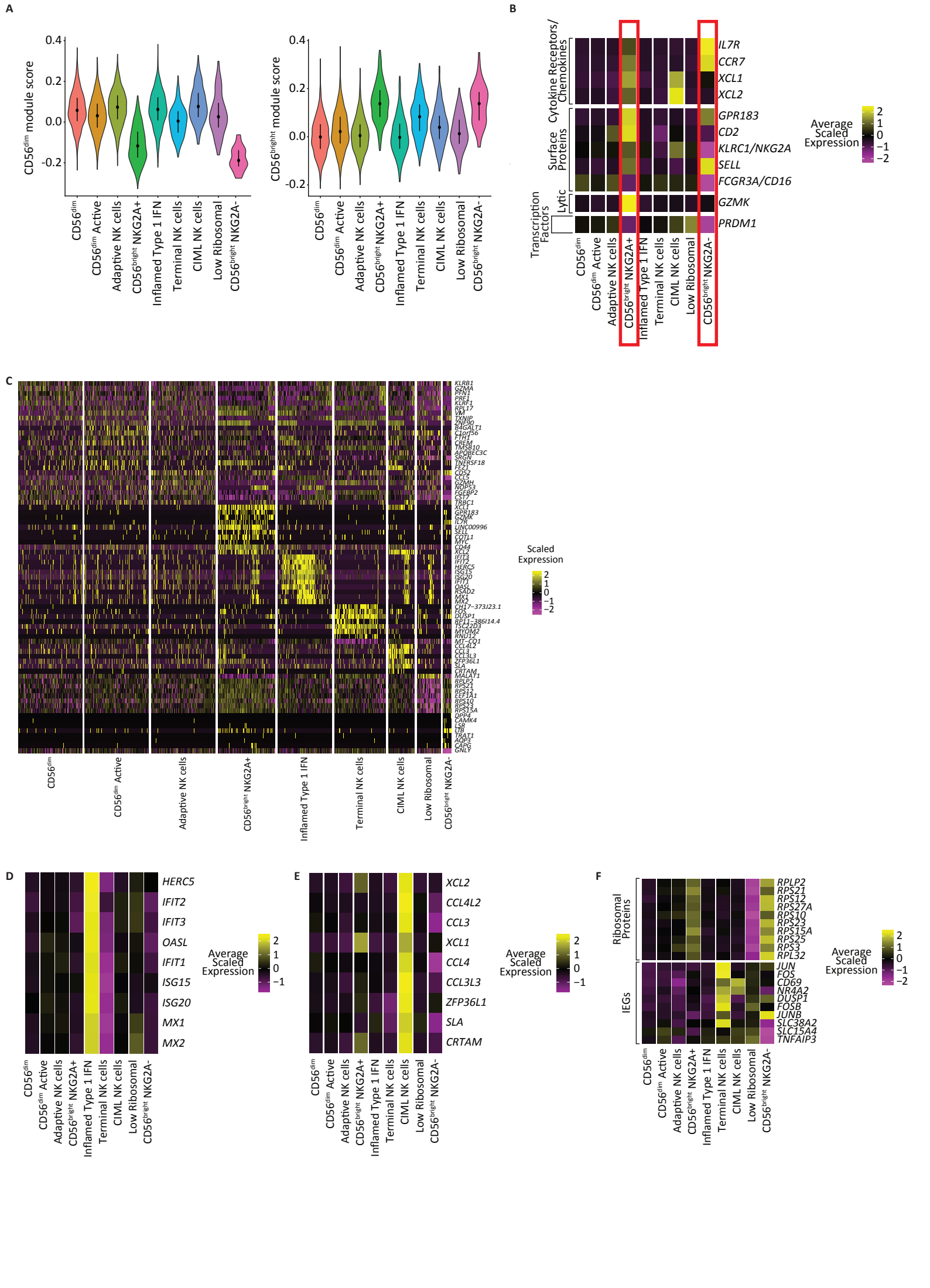

A
nGene
nUMI
2000
1000
8000
6000
4000
2000
Cluster0
Cluster1
Cluster2
Cluster3
Cluster4
Cluster6
Cluster8
B
Quiescence Module Score
Scenescence Module Score
0.4
0.2
0.0
C
percent.mito
0.100
0.075
0.050
0.025
0.000
pro-apoptotic module score
0.4
0.2
0.0
−0.2
D
Cluster 8
IL-2 stimulated expression
unstimulated expression
0
2
4
6

Figure S9.


A
CD56dim module score
CD56brightNT module score
CD56dim
CD56dim Active
Adaptive NK cells
CD56bright NKG2A+
Inflamed Type 1 IFN
Terminal NK cells
CIML NK cells
Low Ribosomal
CD56bright NKG2A-
B
Cytokine Receptors/ Chemokines
Surface Proteins
Lytic
Transcription Factors
IL7R
CCR7
XCL1
XCL2
GPR183
CD2
KLRC1/NKG2A
SELL
FCGR3A/CD16
GZMK
PRDM1
Average Scaled Expression
C
Scaled Expression
D
HERC5
IFIT2
IFIT3
OASL
IFIT1
ISG15
ISG20
MX1
MX2
Average Scaled Expression
E
XCL2
CCL4L2
CCL3
XCL1
CCL4
CCL3L3
ZFP36L1
SLA
CRTAM
Average Scaled Expression
F
Ribosomal Proteins
IEGs
RPLP2
RPS21
RPS12
RPS27A
RPS10
RPS23
RPS15A
RPS25
RPS3
RPL32
JUN
FOS
CD69
NR4A2
DUSP1
FOSB
JUNB
SLC38A2
SLC15A4
TNFAIP3
Average Scaled Expression