

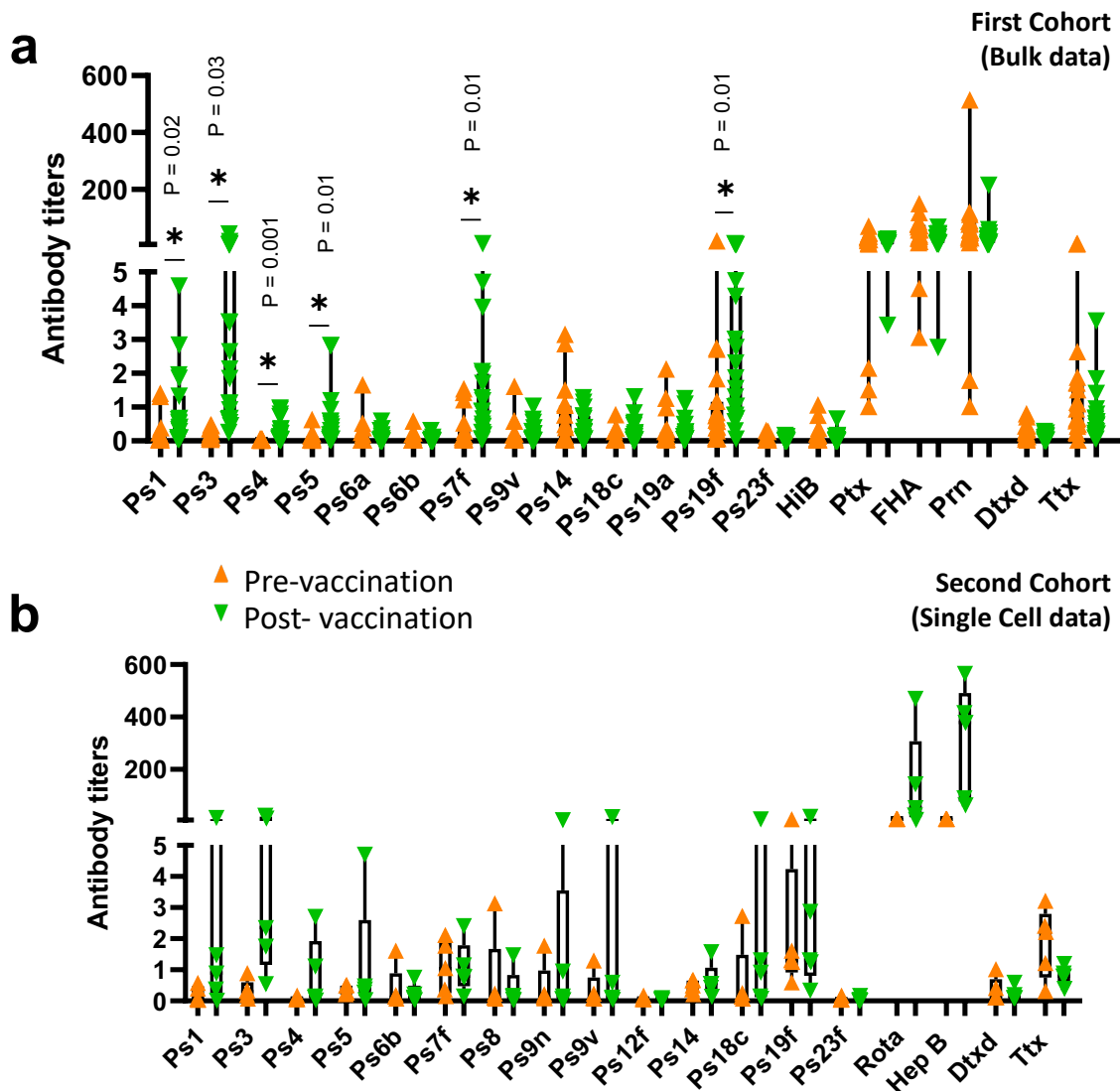
# **Young infants display heterogeneous serological responses and extensive but reversible transcriptional changes following initial immunizations**

Nima Nouri<sup>1,2,#</sup>, Raquel Giacomelli Cao<sup>3,4,#</sup>, Eleonora Bunsow<sup>3,#</sup>, Djamel Nehar-Belaid<sup>1,#</sup>, Radu Marches<sup>1</sup>, Zhaohui Xu<sup>3,5</sup>, Bennett Smith<sup>3</sup>, Santtu Heinonen<sup>3,6</sup>, Sara Mertz<sup>3</sup>, Amy Leber<sup>7</sup>, Gaby Smits<sup>8</sup>, Fiona van der Klis<sup>8</sup>, Asunción Mejías<sup>3,4,5</sup>, Jacques Banchereau<sup>1,9,†</sup>, Virginia Pascual<sup>10,†,\*</sup>, Octavio Ramilo<sup>3,4,5,†,\*</sup>

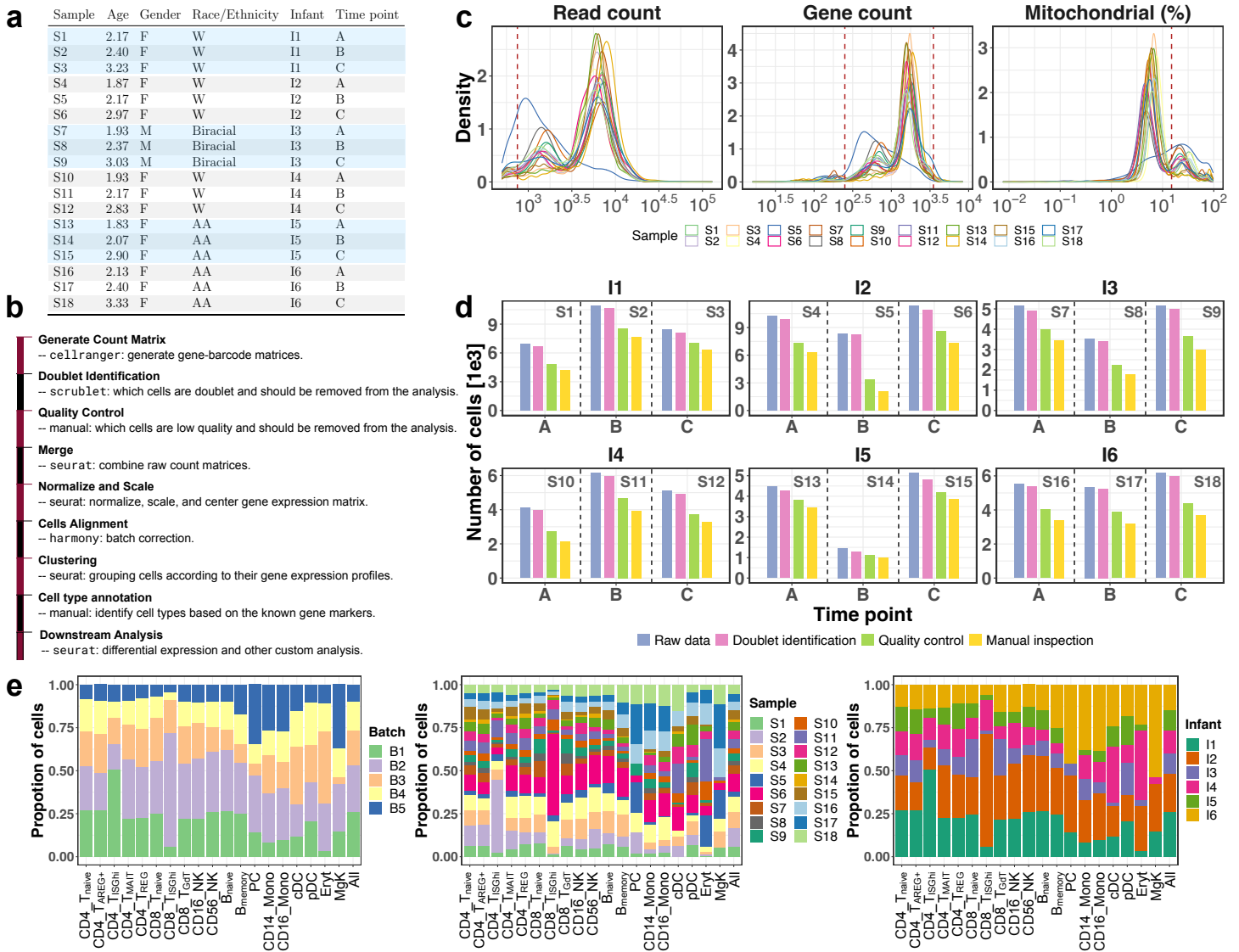
# These authors contributed equally

† These authors jointly supervised the work

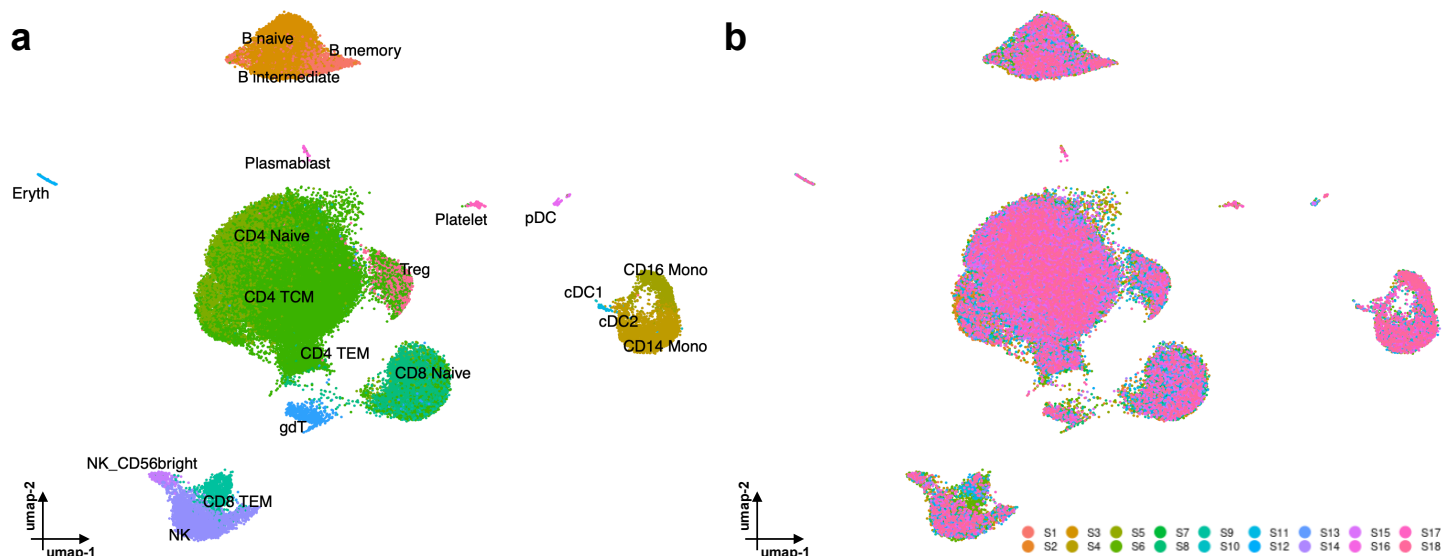
\* Correspondence: Octavio Ramilo (octavio.ramilo@stjude.org) and Virginia Pascual (vip2021@med.cornell.edu)



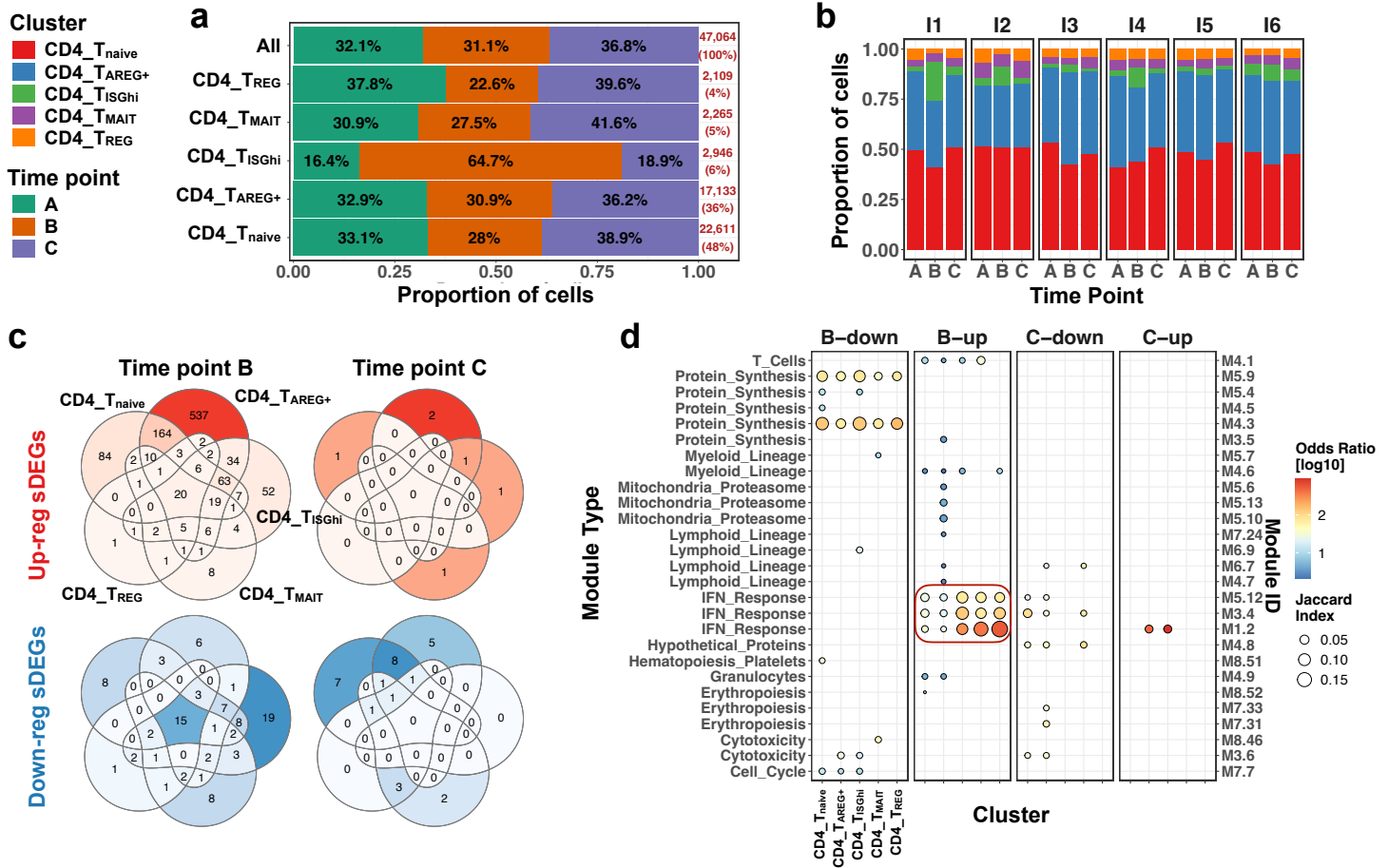
**Supplementary Figure 1: Antibody response in 2-month-old infants from bulk transcriptome cohort (A) and single sell cohort (B).** Antibody titers (y-axis) before vaccination (day 0; orange triangles) and approximately 4 weeks later (green inverted triangles) against different vaccine antigens (x-axis) in each individual infant (n=19 with serologic responses from the first cohort and n=5 from the second cohort). An asterisk (\*) indicates a significant difference of unadjusted  $p < 0.05$  (labels), calculated by a two-sided t-test. Ps: Pneumococcus, Rota: Rotavirus, Hep B: Hepatitis B, Dtxd: Diphtheria toxoid, HiB: Haemophilus influenza type b, Ptx: Pertussis toxin, FHA: Filamentous hemagglutinin, Prn: Pertactin. Please note that the antibody titers measured were slightly different in the two separate cohorts.



**Supplementary Figure 2: Overview of single-cell gene expression analysis.** (a) Dataset demographics: F: Female, M: Male, W: White, AA: African American. Ages are shown in months. (b) Overview of the computational pipeline. (c) Quality control metrics: number of reads per cell (left panel), number of expressed genes per cell (middle panel), and mitochondrial percentage (right panel) for each of the n=18 color-coded samples. Dashed lines indicate the thresholds used in quality control (Read count = 750; Genes count between 250 and 3500; and percentage of mitochondrial < 15%). (d) Number of cells before and after filtration, including doublet removal, quality control removal, and manual removal. (e) Proportion of cells per batch (left panel), per sample (middle panel), and in each individual infant (right panel) for each of the different cell types.

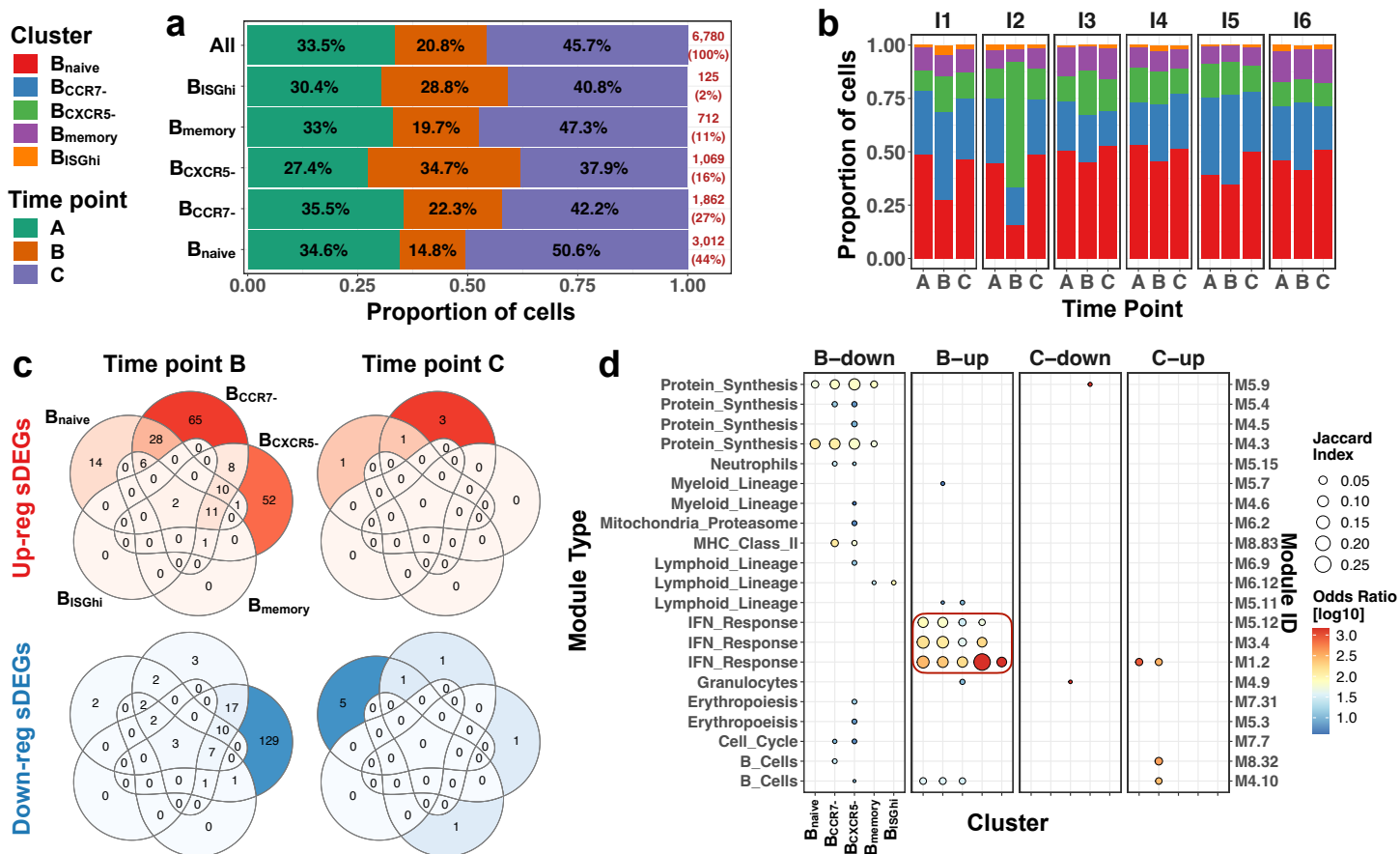


**Supplementary Figure 3: Additional supporting UMAPs.** (a) Automated reference-based (Stuart 2019) annotation results using Azimuth web application (Hao 2021). (b) UMAPs demonstrating good overlap between different samples (color-coded cells grouped by sample).

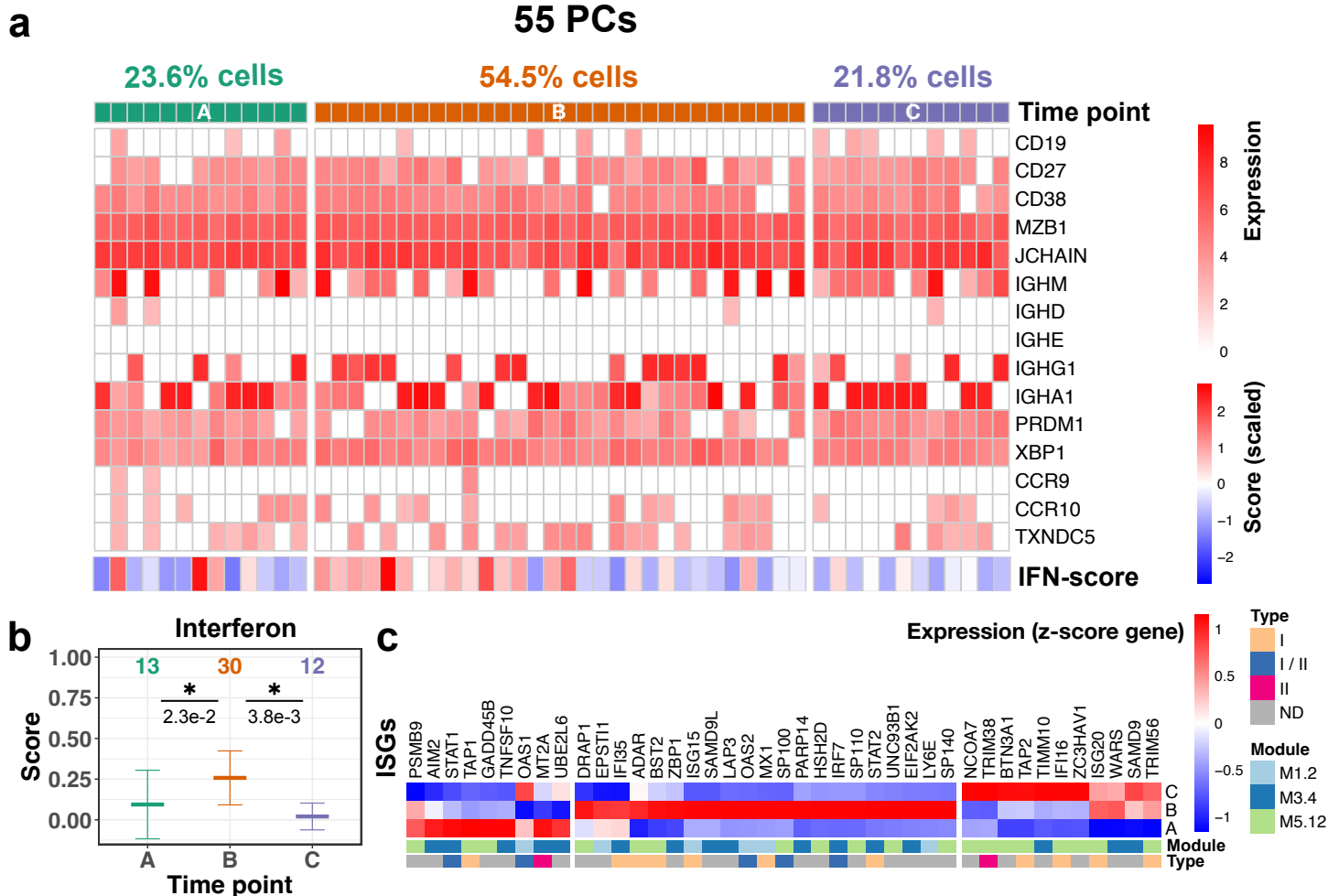


**Supplementary Figure 4: CD4<sup>+</sup> T cells supplementary panel.** Subcluster (SC) and time-point color codes are shown in the top-left. **(a)** Proportion of cells per SC at each time point. **(b)** Proportion of cells at each time point for each SC and infant. **(c)** Venn-diagram showing the number of significantly differentially expressed genes (sDEGs) for each comparison among SCs per time point and regulation type (up/down-regulated). sDEGs are determined as those with a log<sub>2</sub>-fold-change>0.25 and adjusted p<0.05, calculated using two-sided Wilcoxon rank-sum tests. **(d)** Significant overlaps (determined by Fisher's exact test, p<0.05) between sDEGs and modules for each time point and regulation type within each SC. The odds ratio (color-bar legend) represents the strength of association, and the Jaccard index (legend circles) measures the similarity between two lists.

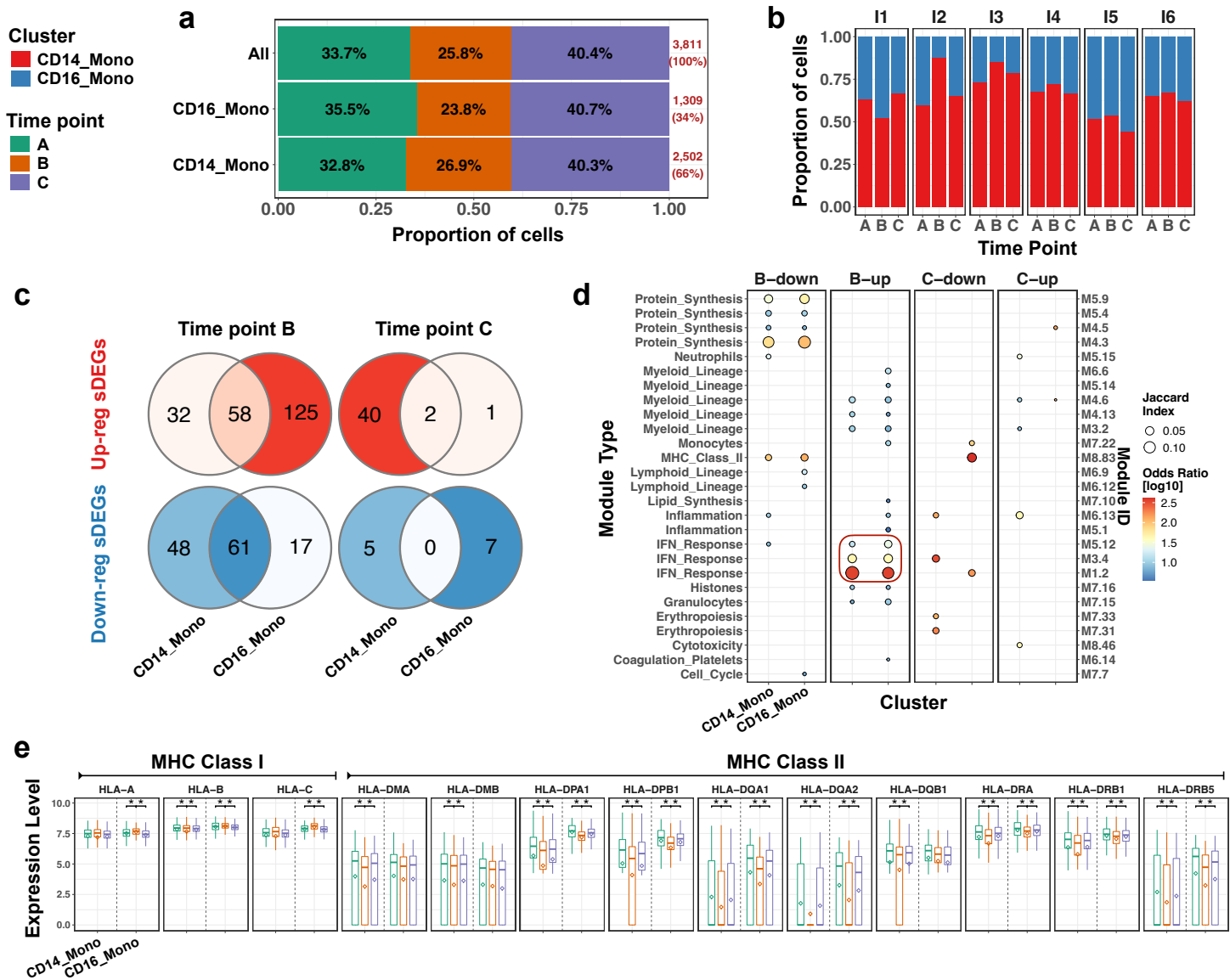




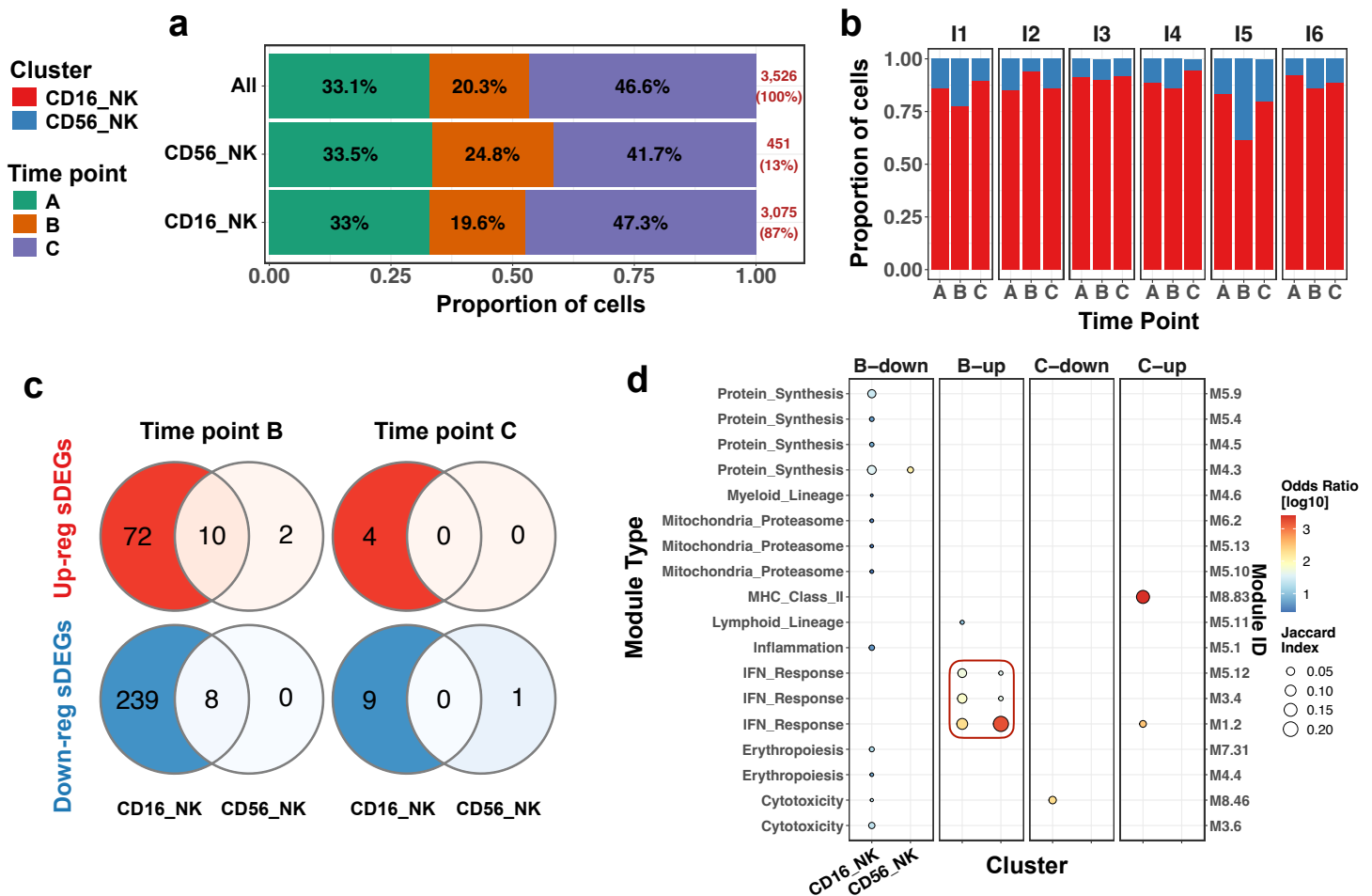
**Supplementary Figure 6: B cells supplementary panel.** Subcluster (SC) and time-point color codes are shown in the top-left. **(a)** Proportion of cells per SC at each time point. **(b)** Proportion of cells at each time point for each SC and infant. **(c)** Venn-diagram showing the number of significantly differentially expressed genes (sDEGs) for each comparison among SCs per time point and regulation type (up/down-regulated). sDEGs are determined as those with a  $\log_2$ -fold-change  $> 0.25$  and adjusted  $p < 0.05$ , calculated using two-sided Wilcoxon rank-sum tests. **(d)** Significant overlaps (determined by Fisher's exact test,  $p < 0.05$ ) between sDEGs and modules for each time point and regulation type within each SC. The odds ratio (color-bar legend) represents the strength of association, and the Jaccard index (legend circles) measures the similarity between two lists.



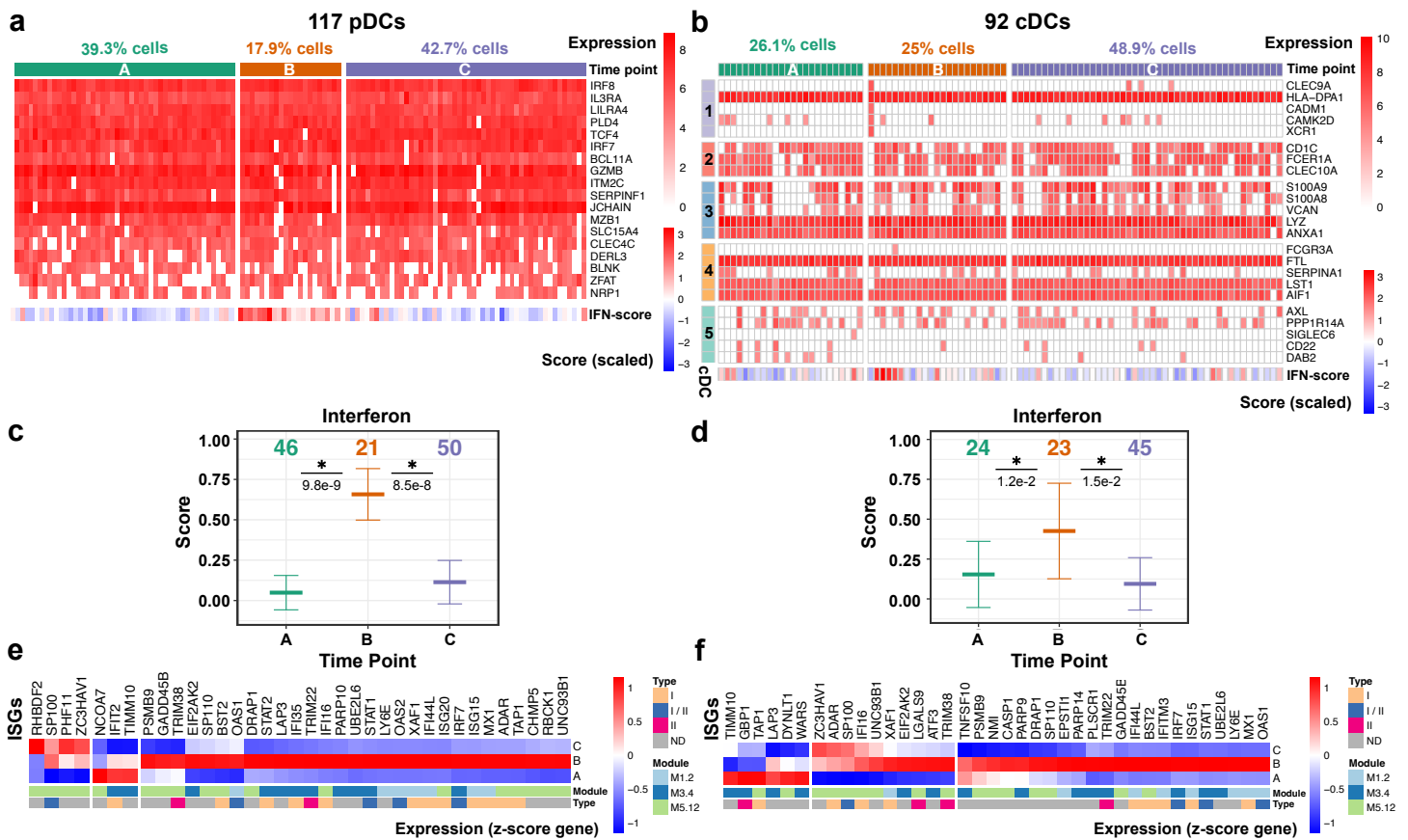
**Supplementary Figure 7: Vaccinated infants display an expansion of plasma cells expressing interferon-stimulated genes at day 7 post-vaccination. (a)** Expression levels of selected genes for each cell. Scaled interferon scores calculated for each cell are shown in the bottom row. Proportion of cells for each time point is indicated by a color-coded label on top. **(b)** Interferon scores calculated for each cell per time point. Error bars indicate standard deviation across cells, and the horizontal bar represents the mean. An asterisk (\*) indicates a significant difference between two consecutive time points, determined using two-sided t-tests, with unadjusted  $p < 0.05$  (labels). The number of cells for each time point is shown by a color-coded label on top. **(c)** Average expression levels of the top 30 Interferon-stimulated genes (ISGs) per each time point. The color keys at the bottom represent gene associations.



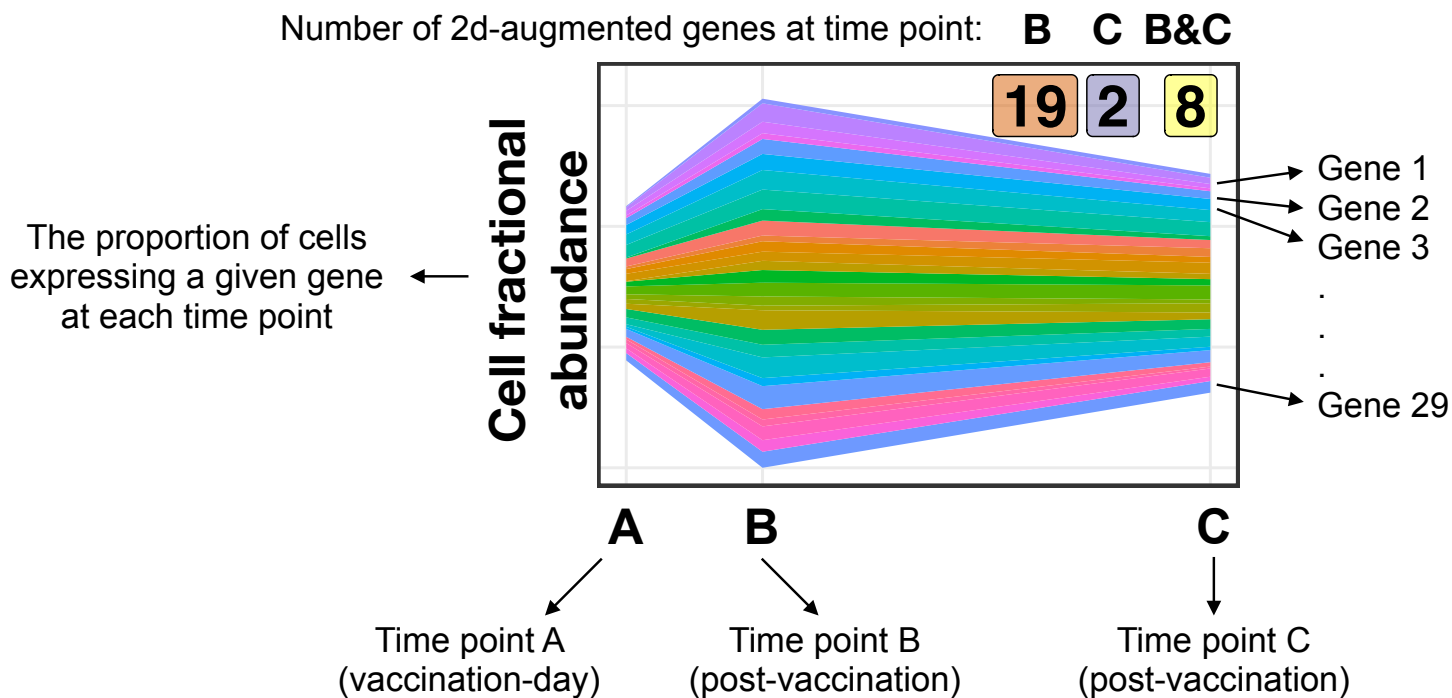
**Supplementary Figure 8: Monocytes supplementary panel.** Subcluster (SC) and time-point color codes are shown in the top-left. **(a)** Proportion of cells per SC at each time point. **(b)** Proportion of cells at each time point for each SC and infant. **(c)** Venn-diagram showing the number of significantly differentially expressed genes (sDEGs) for each comparison among SCs per time point and regulation type (up/down-regulated). sDEGs are determined as those with a log2-fold-change > 0.25 and adjusted  $p < 0.05$ , calculated using two-sided Wilcoxon rank-sum tests. **(d)** Significant overlaps (determined by Fisher's exact test,  $p < 0.05$ ) between sDEGs and modules for each time point and regulation type within each SC. The odds ratio (color-bar legend) represents the strength of association, and the Jaccard index (legend circles) measures the similarity between two lists. **(e)** MHC class I and II related genes expression pattern over time for each SC. The results are depicted in boxplots for all cells, in which the upper and lower bounds represent the 75% and 25% percentiles, respectively. The center bars indicate the medians, and the whiskers denote values up to 1.5 interquartile ranges above the 75% or below the 25% percentiles. The diamonds indicate the mean. An asterisk (\*) indicates a significant difference between two consecutive time points, determined using two-sided t-tests, with unadjusted  $p < 0.05$ .



**Supplementary Figure 9: Natural killer cells supplementary panel.** Subcluster (SC) and time-point color codes are shown in the top-left. **(a)** Proportion of cells per SC at each time point. **(b)** Proportion of cells at each time point for each SC and infant. **(c)** Venn-diagram showing the number of significantly differentially expressed genes (sDEGs) for each comparison among SCs per time point and regulation type (up/down-regulated). sDEGs are determined as those with a  $\log_2$ -fold-change  $> 0.25$  and adjusted  $p < 0.05$ , calculated using two-sided Wilcoxon rank-sum tests. **(d)** Significant overlaps (determined by Fisher's exact test,  $p < 0.05$ ) between sDEGs and modules for each time point and regulation type within each SC. The odds ratio (color-bar legend) represents the strength of association, and the Jaccard index (legend circles) measures the similarity between two lists.



**Supplementary Figure 10: Plasmacytoid dendritic cells (pDCs) and conventional dendritic cells (cDCs) panels. (a-b)** Expression level of selected genes for each pDC (a) and cDC (b). Scaled interferon scores calculated for each cell are shown in the bottom row. Proportion of cells per each time point is shown by color-coded label on top. **(c-d)** Interferon score calculated for each cell per time point for pDCs (c) and cDCs (d). Error bar indicates standard deviation. Horizontal bar indicates mean. An asterisk (\*) indicates a significant difference between two consecutive time points, determined using two-sided t-tests, with unadjusted  $p < 0.05$  (labels). **(e-f)** Average expression levels of top 30 Interferon-stimulated genes (ISGs) per each time point for pDCs (e) and cDCs (f). The color keys on the bottom represent genes association.



**Supplementary Figure 11: Stream plot for the visualization of 2D-augmented genes.** Each ribbon spans three time-points A, B, and C (x-axis) and represents an augmented gene. The thickness of each ribbon (y-axis) indicates the proportion of cells expressing the given gene at each time point. Labels at the top-right indicate the number of identified augmented genes at time points B (orange), C (purple), or both B and C (yellow).



## References

- Hao, Yuhao and Hao, Stephanie and Andersen-Nissen, Erica and Mauck III, William M and Zheng, Shiwei and Butler, Andrew and Lee, Maddie J and Wilk, Aaron J and Darby, Charlotte and Zager, Michael and others. 2021. "Integrated analysis of multimodal single-cell data." *Cell* (Elsevier) 184 (13): 3573--3587.
- Stuart, Tim and Butler, Andrew and Hoffman, Paul and Hafemeister, Christoph and Papalexi, Efthymia and Mauck III, William M and Hao, Yuhao and Stoeckius, Marlon and Smibert, Peter and Satija, Rahul. 2019. "Comprehensive integration of single-cell data." *Cell* (Elsevier) 177 (7): 1888--1902.