



MU5BM572 - Advanced cytometry¶  
2021-2022¶

Project 2 - Natural Killer cells¶  
Written report¶

By :¶  
Raphaël BECQUART¶  
Solal BELLAICHE¶  
Minh-Anh HUYNH¶

# Table of contents¶

¶  
¶

<b>Introduction¶</b>	<b>3</b>
<b>Materials &amp; Methods¶</b>	<b>4</b>
Experimental design¶	4
<b>Results¶</b>	<b>6</b>
Clustering cells with SPADE¶	6
Heatmap based on SPADE clustering¶	6
Phenotype viewing of NK metaclusters ¶	8
Kinetics viewing of NK cells metaclusters¶	10
Differentially abundant clusters among NK cells¶	11
Global similarities and differences between biological conditions and individuals¶	12
t-SNE representation¶	13
UMAP representation¶	13
k-means representation¶	14
<b>References</b>	<b>15</b>

¶

**Keywords :** Vaccination, HIV, NK cells, Clustering, SPADE

¶

# Introduction¶

¶

Since the beginning of the HIV/AIDS epidemic in 1981, more than 79.3 million people have been infected with HIV resulting in 36.3 million deaths (World Health Organization, 2021). At the end of 2020, it was estimated that 37.7 million people were living with HIV. Recommended first-line treatments include antiretroviral therapies (ART) that help reduce viral loads and increase the life expectancy of patients without actually curing the infection (National Institute of Health, 2016). Perhaps the best long-term strategy to end the HIV epidemic is the development of vaccines that interrupt HIV's transmission chain. ¶

¶

More than 30 vaccine candidates have moved to clinical trials, yet none of these candidates offered demonstrable vaccination protection in phase II-III studies. The only exception is the 2003 Thai RV144 trial that evaluated the efficacy and safety of a recombinant canarypox vector vaccine with two booster shots of a recombinant glycoprotein 120 vaccine (Rerks-Ngarm et al., 2009). HIV infection rates in the vaccinated group was only 30% lower than in the control group. In absolute numbers, this corresponds to 74 HIV-infected individuals in the placebo group and 51 in the vaccination group. The fact that these numbers are very small compared to the 16,402 total study participants reflects the difficulty and lengthy nature of HIV trials: But most importantly, this trial demonstrated for the first time that HIV vaccines can prevent infections, and that poxvirus vectors are potential HIV vaccine vector candidates. ¶

¶

Poxvirus vectors have since gained significant attention for vaccination against infections, especially the modified vaccinia Ankara (MVA) virus strain that was used in the 1970s to help eradicate smallpox (Gomez et al., 2008). The long history of MVA vaccine development against smallpox helped establish its very good safety profile—likely due to its areplicative nature—and its ability to elicit strong immune responses in animals and humans (Sutter and Staib, 2003). Several clinical trials have assessed the use of MVA vaccines against HIV and found highly immunogenic T-cell responses specific to HIV antigens *gag*, *env*, and *pol* (Sandström et al., 2008; Vasan et al., 2010). In addition, Natural Killer (NK) cell activities and diverse subtypes are impacted by vaccination and infections (Vargas-Inchaustegui et al., 2016; Wilk and Blish, 2018). NK cells are innate immune cells characterized by cytotoxic functions but also by the modulation of other immune cells through cytokine signaling and various cross-talks. While some studies have characterized the roles of NK cells in MVA-induced-immunity using *in vitro* assays (Price et al., 2013), *in vivo* studies are lacking. ¶

¶

¶

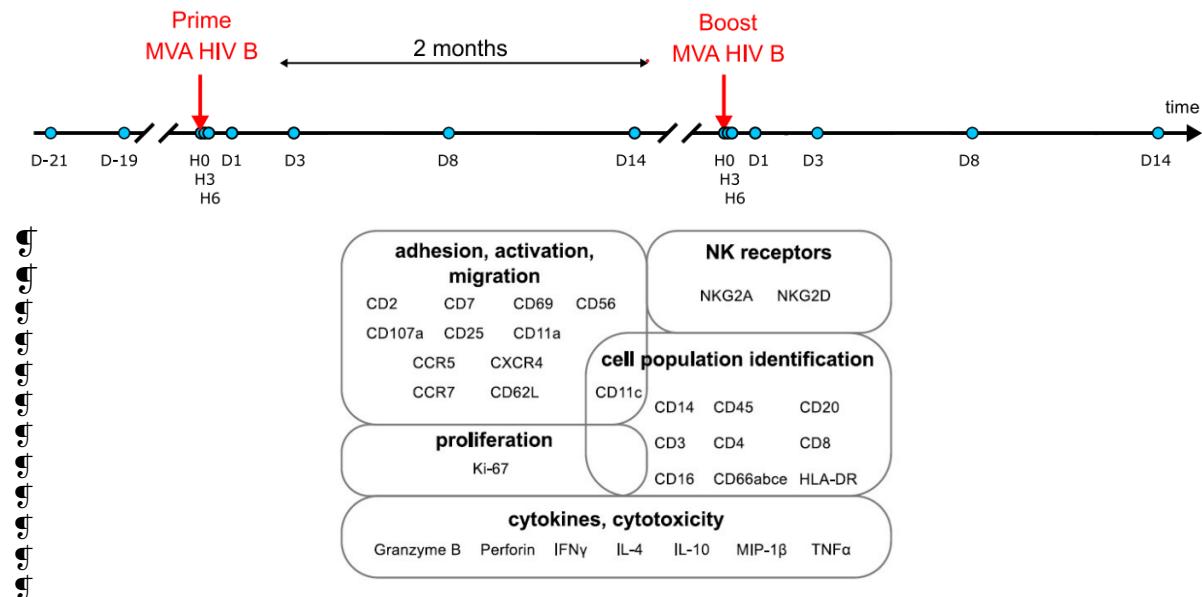
In the present study, we characterize natural killer (NK) cell responses in macaques following MVA-B vaccination. The recombinant MVA-B vaccine vector expresses *gag*, *env*, *pol*, and *nef* HIV-1 antigens, and requires two doses (prime and boost). Blood samples were collected at different time-points before and after prime, and after the booster shot. Myeloid cells were then identified with mass cytometry using a panel of 30 innate immune cell markers. Finally, we identified time-dependent alterations in NK cell tropism following vaccination with the help of the Spanning-tree Progression Analysis of Density-normalized Events algorithm (Qiu et al., 2011) (SPADE) and its derivative package SPADEVizR. ¶

¶

# Materials & Methods¶

## Experimental design¶

Five macaques were vaccinated against HIV-B twice with an MVA vaccine. The second injection (boost) was done two months after the first one (prime). Blood samples were collected at different days and hours before prime, post-prime, and post-boost. Leukocytes were stained with a panel of 32 antibodies for mass spectrometry analysis (*Fig. 1*).¶



**Figure 1. Experimental design for MVA immunization and cell population identification.** 5 macaques were immunized with an MVA HIV B vaccine, and blood samples were collected at different days and hours before & post-prime, and post-boost, for multiparametric mass cytometry analysis. ¶



## Identification of cell populations¶

The identification of cell populations was done with the Spanning-tree Progression Analysis of Density-normalized Events algorithm (SPADE) (Gautreau et al., 2017). The percentage of cells down-sampled was set to 5%, and the desired number of clusters was k=100. The application of SPADE to all samples allowed the construction of a minimum spanning tree (MST) linking cell clusters. ¶

## Quality control of cell clusters¶

The quality of SPADE clusters was assessed with “qcUniformClusters” function included in a derivative R Package called SPADEVizR (Gautreau et al., 2017) to choose an optimal value of the number of clusters (k) to maximize the uniformity of the phenotypes of all clusters. Unimodal marker expressions across all clusters were determined by Hartigan’s Dip test (p-value <5%). With k=100, the percentage of clusters having only uniform phenotypes among all clusters of the dataset was 77%. With k=50 and k=200, we had a lower quality rate. Additionally, we spotted that Perforin was non-unimodal in 13% of the clusters (for example in cluster 78 in Fig. xxx) so we decided to remove it from SPADE analysis. We assessed that the loss of this marker would not affect the final information because we still have the GranzymeB marker that is usually correlated with Perforin. After that the quality improved to 81%. We checked the number of small Clusters with the “qcSmallClusters” function. With k=100, they represented 2% of all clusters whereas it was 13% with k=200. ¶

## **Categorical heatmaps to display cell cluster phenotypes¶**

Heatmaps of cell cluster phenotypes were obtained using SPADEVizR. Marker expression ranges are classified into 5 levels, based on the mean of median marker expressions. ¶

¶

## **Identifying metaclusters and NK cells based on the heatmap¶**

We cut the hierarchical dendrogram above the heatmap after eleven divisions, allowing us to gather similar clusters into twelve metaclusters. By looking at the phenotypic expression of each metacluster via cell markers, we can identify the NK cells metaclusters and the metaclusters corresponding to the other types of leukocytes. ¶

¶

## **Phenotypic characterisation of NK cells metaclusters¶**

To visualize the phenotype of metaclusters of interest, the “phenoViewer” function from SPADEVizR was used.¶

## **Abundance of NK cells metaclusters variation¶**

The function “kineticsViewer” from SPADEVizR was used to display the cell abundances over the different biological conditions for each individual. ¶

## **Identification of differentially abundant clusters¶**

To identify differentially abundant clusters between two biological conditions (e.g post-prime and post-boost), the abundances of each cluster per condition are modeled with a Gaussian distribution centered around the mean of the conditions compared, using the “identifyDAC” function. A t-test is then performed to assess significance between the 2 distributions (p-value <5%), and a fold-change filter was applied ( $FC > |2|$ ).¶

¶

## **Multidimensional Scaling¶**

Our MDS representation is Euclidean-distance-based and was fed with only NK cell clusters from all samples, for the 3 different biological conditions.¶

## **Code availability¶**

<https://github.com/Minh-AnhHuynh/R-NK-Cytometry-Project-2021-2022>¶

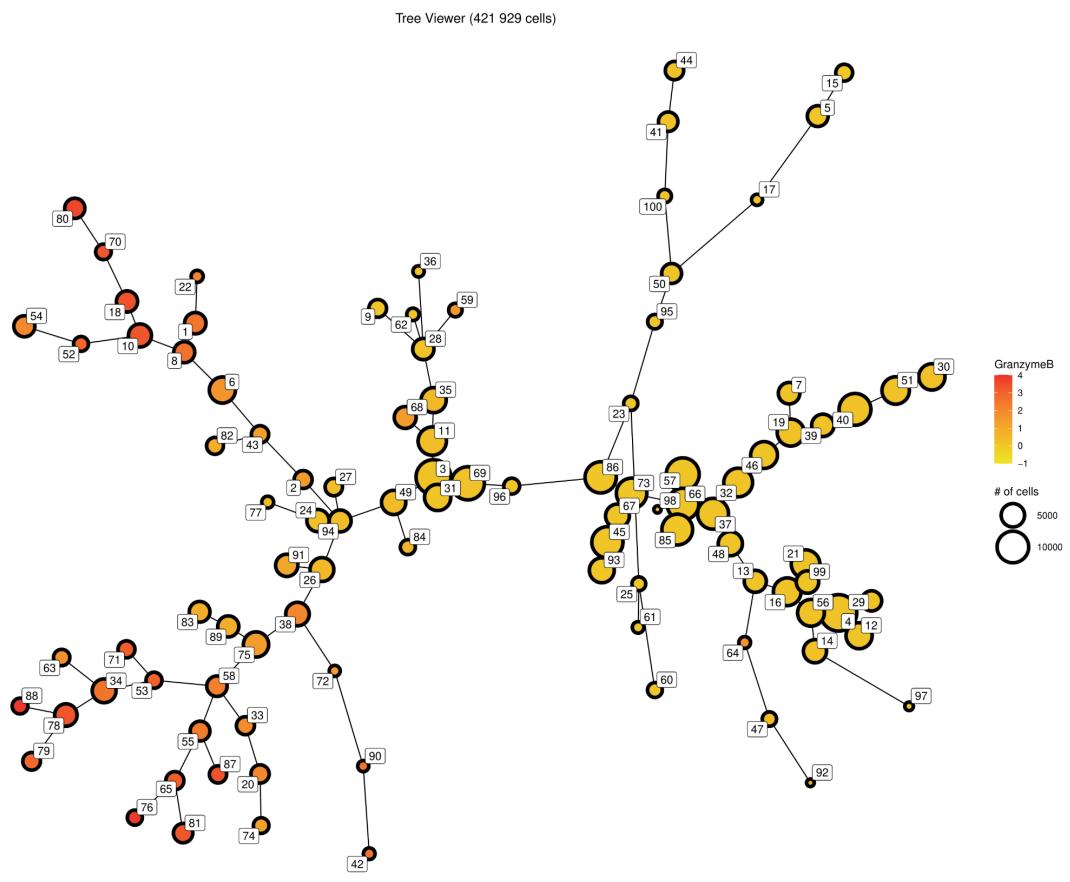
Please read the README.md to initiate the project.¶

¶

# Results¶

## Clustering cells with SPADE¶

To visualize global cluster phenotype similarities in our sample calculated by the SPADE algorithm, a SPADE tree visualization was done (*Fig. 2*). The SPADE tree shows that our clusters can be phenotypically grouped into larger families called metaclusters. Granzyme B is a strong NK cell marker (Cho et al., 1997) and is notably expressed in the left part of the SPADE tree, indicating that potentially two different NK cell populations may exist, as evidenced by the divergent clustering in the left part of the SPADE tree.¶||

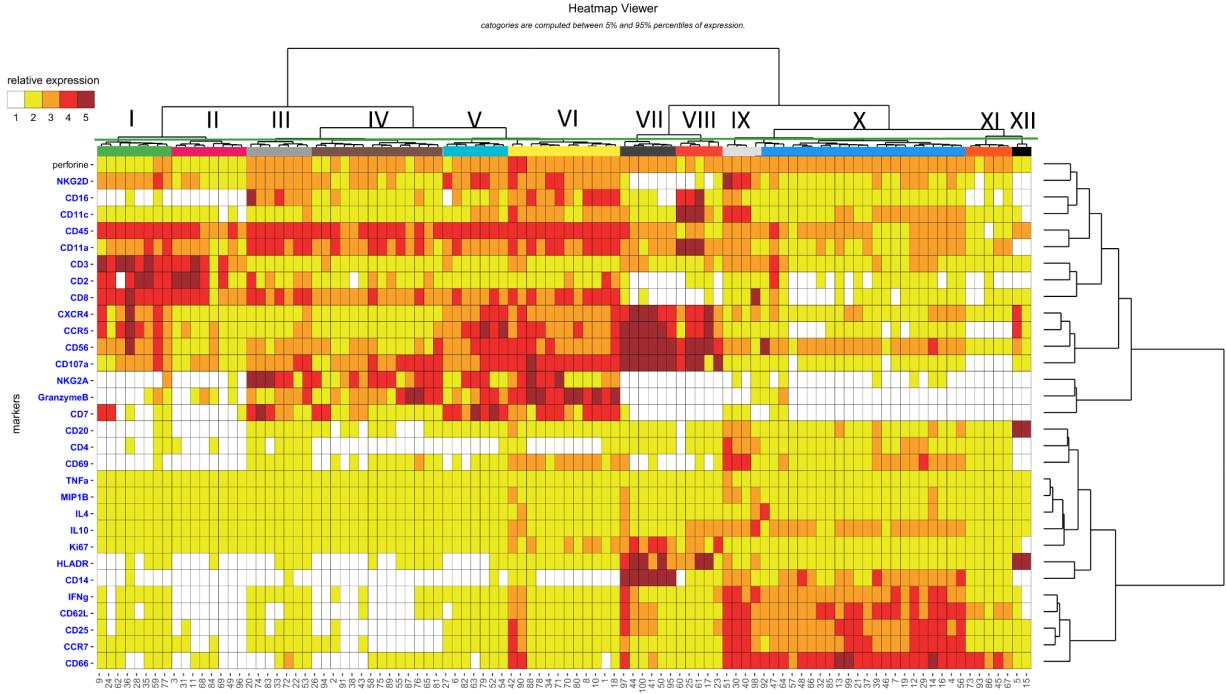


**Figure 2. Minimum spanning tree colored with a strong NK cell marker, Granzyme B.** The tree was generated with SPADE with all macaque samples at all time points, 100 clusters, 30 markers and 95% of cells used in the clustering procedure. Each node represents a cluster of cells and its size is correlated to the number of cells contained (5000 or 10000). The yellow to red gradient is correlated to the intensity of Granzyme B expression in each cluster.¶

## Heatmap based on SPADE clustering¶

We then sought to further characterize phenotypic variations in our samples by generating a heatmap in all 100 clusters generated with the SPADE algorithm (*Fig. 3*). We can note that some clusters sharing a similar phenotype can be merged together in metaclusters. As a matter of fact, we cut horizontally the dendrogram to highlight 12 metaclusters, on top of which we added colored rectangles for visualization. We numbered those metaclusters from I to XII

starting from the left. Using literature, we managed to associate our metaclusters to different cell types (**Table 1**). In fact, we identified 6 immune cell types: T-cells, NK cells, monocytes, dendritic cells, neutrophils, and B-cells. Interestingly, the 4 NK cell meta-clusters do cluster together which adds two extra clusters to the two previously found with our SPADE tree. Notably, two large groups of coexpressed markers can be identified thanks to hierarchical clustering shown by the horizontal dendrogram. One group is composed of cytotoxic (e.g Granzyme B), cell population and chemotaxic markers and the other one is composed of cytokines.



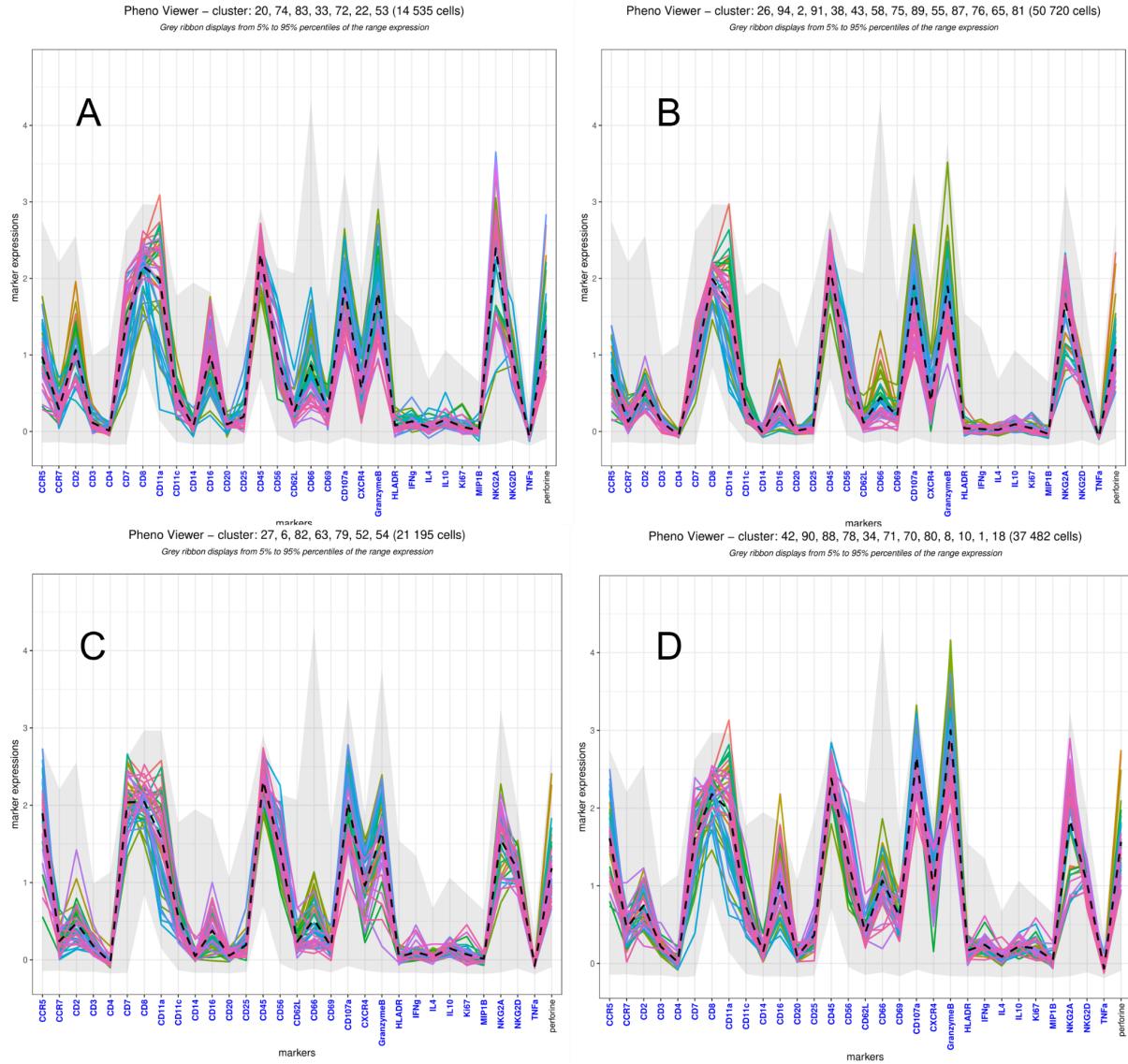
**Figure 3. Visualization of cell cluster phenotypes in the form of a heatmap.** X-axis represents the 100 clusters, Y-axis represents the markers (colored in blue when used in the SPADE analysis) and the color gradient from white to red indicates the relative marker expression. The arbitrary colored rectangles between the dendrogram and the heatmap show the different metaclusters based on the truncation of the dendrogram at the green line level.

Metaclusters¶	Specific markers¶	Cell type¶
I (green), II (pink)¶	CD3+¶	T-cells ¶
III (medium gray), IV(dark brown), V (light blue), VI (yellow)¶	NKG2D, NKG2A, Granzyme B, CD3 -¶	NK cells¶
VII (dark gray)¶	CD14+, HLA-DR+¶	Monocytes¶
VIII (red)¶	CD11c+ HLA-DR+¶	Dendritic Cells¶
IX (light gray), X (dark blue) and XI (orange)¶	CD66+¶	Neutrophils ¶
XII (black)¶	CD20+ HLA-DR+¶	B-cells¶

**Table 1. Metaclusters phenotypes.¶**

# Phenotype viewing of NK metaclusters ¶

¶



¶

**Figure 4. Representation of mean marker expression of each NK cell metacluster: metacluster III (A), metacluster IV (B), metacluster V (C) and metacluster VI (D). Each line corresponds to a biological sample. The dashed line indicates the mean marker expressions of all samples. The gray ribbon indicates the bounds of marker expressions in the overall dataset. Markers used as clustering markers are shown in blue. The color of the line corresponds to the biological condition (red for before prime, blue to purple gradient for post-prime, orange to green gradient for post-boost). With this representation we can highlight phenotypic differences between our NK cell metaclusters. ¶**

- individuals
- BB078
  - BB231
  - BO641
  - BD620

To characterize the four NK cell metaclusters identified previously, we computed their mean expressions for each marker (**Fig. 4**). Our first observation is that marker expression profiles are homogeneous between individuals of the same metaclusters. ¶

¶

Metaclusters III¶	<b>CD16<sup>bright</sup> CD8<sup>+</sup> CD2<sup>bright</sup> CXCR4<sup>dim</sup> NKG2A<sup>bright</sup>.</b> ¶
Metaclusters IV¶	<b>CD16<sup>dim</sup> CD8<sup>+</sup> CD66<sup>dim</sup> CD2<sup>dim</sup> CXCR4<sup>dim</sup> NKG2A<sup>bright</sup>.</b> ¶
Metaclusters V¶	<b>CD16<sup>dim</sup> CD8<sup>+</sup> CD66<sup>dim</sup> CD2<sup>dim</sup> GranzymeB<sup>dim</sup> CD56<sup>bright</sup> CCR5<sup>bright</sup> NKG2A<sup>bright</sup>.</b> ¶
Metaclusters VI¶	<b>CD16<sup>bright</sup> CD8<sup>+</sup> GranzymeB<sup>bright</sup> CD56<sup>bright</sup> CD107a<sup>bright</sup> NKG2A<sup>bright</sup>.</b> ¶

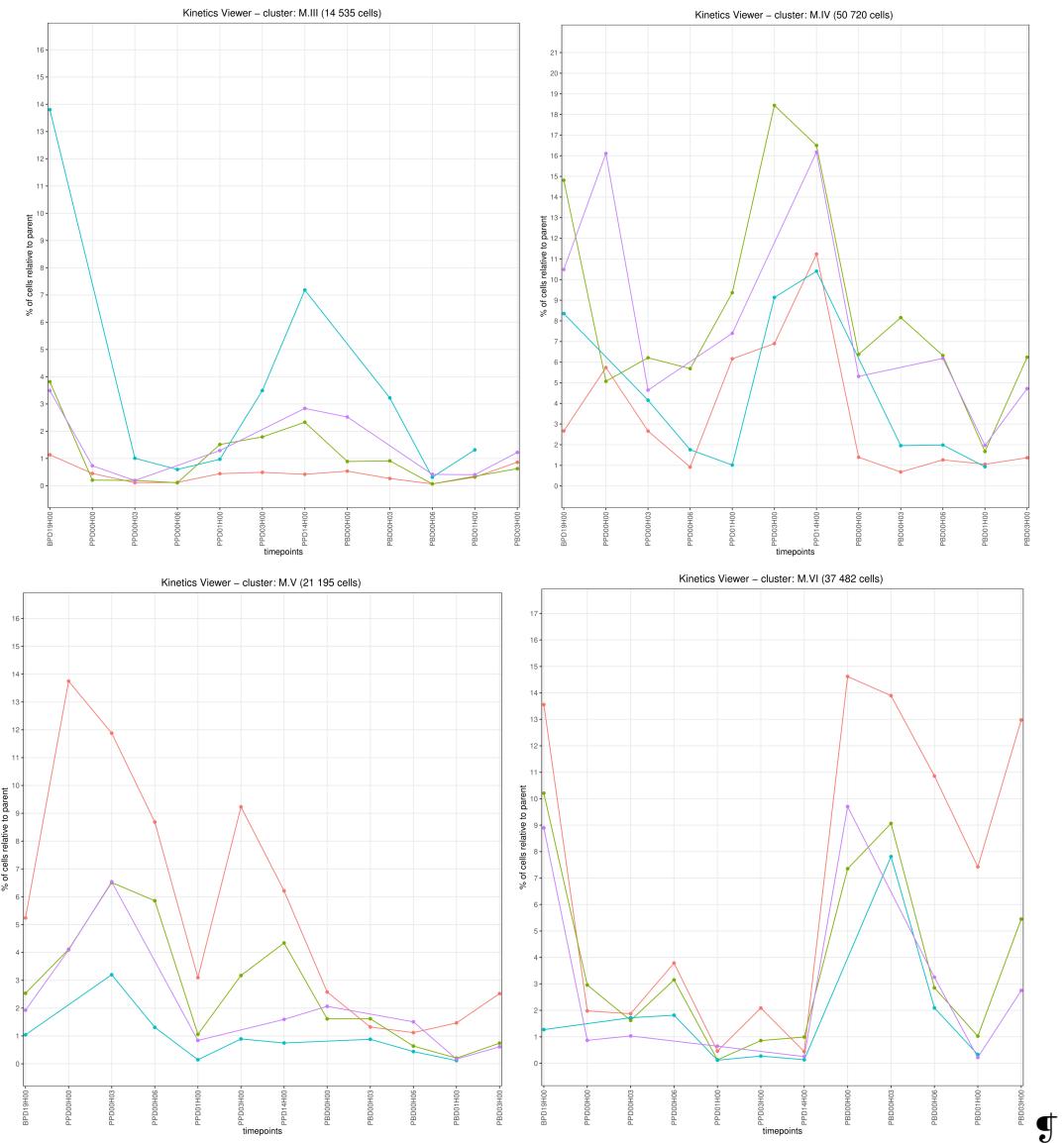
**Table 2.** NK cell metaclusters with their associated relative markers as seen compared to each other. Markers that are not relevant or very low were not shown. ¶In green are colored relatively more expressed markers.

Between the four observed metaclusters, two subtypes of NK cells population can be identified, one comprising metaclusters III and VI characterized by CD16<sup>bright</sup> and CD8<sup>+</sup> and the other one of metaclusters IV and V as CD16<sup>dim</sup> and CD8<sup>+</sup>. ¶

NKG2A, Granzyme B, CD16, CD8 are good markers for NK cells phenotyping (Webster and Johnson, 2005) and associated with NK cell cytotoxicity. Interestingly, the CD8 marker is constant and highly expressed among every NK cell as opposed to humans where CD8 expression is variable (Vargas-Inchaustegui et al., 2016). We can notably see that metacluster VI has the most highly expressed markers among the NK metaclusters. Furthermore, it is the only metacluster with a strong marking of CD107a. This marker is known to be significantly upregulated on the surface of NK cells following stimulation with MHC devoid targets, and correlates with both cytokine secretion and NK cell-mediated lysis of target cells (Alter et al., 2004).

Interestingly there is no difference in NKG2D expression between metaclusters, which is relevant as a staple NK cell marker. Another pointNK cells are on average expressing low CD66 compared to the general sample which is possibly biased as we have seen the cluster of neutrophils expressing CD66 impacting highly the relative average of the marker expression, as seen in **Fig 4**.

# Kinetics viewing of NK cells metaclusters¶



**Figure 5. Representation of the abundance of cells included in metacluster III (A), metacluster IV (B), metacluster V (C), metacluster VI (D) at each biological condition and for each individual.** X-axis represents the biological conditions in chronological order, Y-axis represents the percentage of cells included in the metacluster. Each colored line corresponds to a different individual. ¶

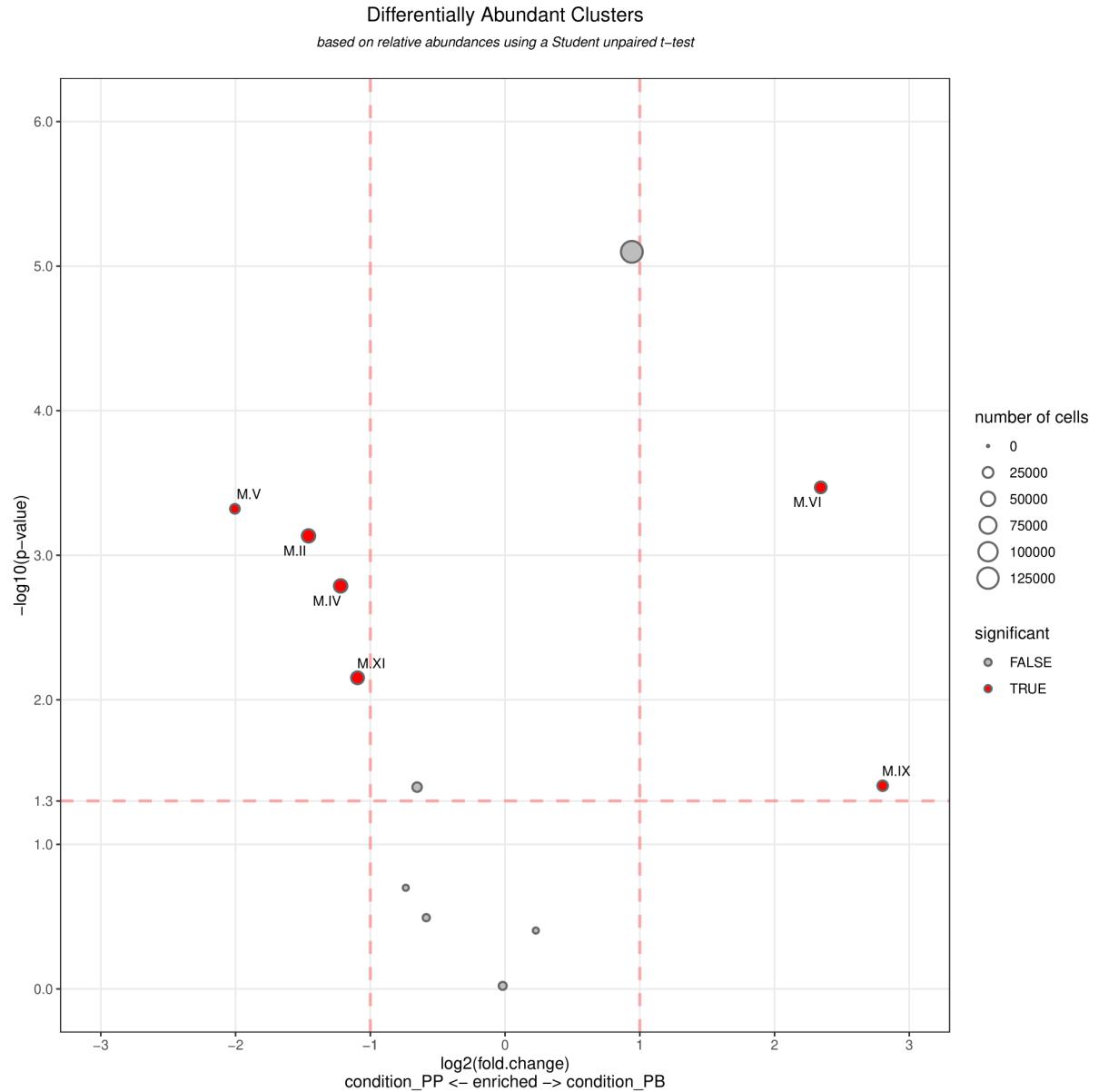
Next, we studied the abundance of each of the 4 identified metaclusters over time to infer their modulation post-prime and post-boost (**Fig. 5**). Despite the significant inter-individual abundance kinetics of NK cells, even before prime, we noticed abundance kinetics in all metaclusters. For metaclusters III, IV, and V, there's a decrease in cell abundances post-prime that continues for 6 hours before increasing at day 1 and spiking right before boost (day 14), before returning to normal levels. In fact, according to our heatmap on **Fig. 3**, these metaclusters express relatively low levels of GranzymeB, suggesting a low cytotoxic phenotype. By contrast, metaclusters VI behaves with a spike in abundance at post-boost. This metaclusters could correspond to effector memory NK cells with high cytotoxic activity

10

as seen by its strong GranzymeB expression (**Fig. 3**). In all, these results suggest that the vaccine's effects on NK cells are most pronounced 2 weeks post-infection for metaclusters III-V and after boost for metaclusters VI. ¶

¶

## Differentially abundant clusters among NK cells¶



¶

**Figure 6.** Representation in the form of a volcano plot of differentially abundant clusters between post-prime and post-boost, among the forty NK cells clusters. X-axis represents the  $\log_2$  of fold-change, Y-axis represents the  $-\log_{10}$  of p-value. Each node represents a cluster generated with SPADE, the size of the node is correlated to the number of cells. Red filling indicates clusters differentially abundant with a p-value  $> 0.05$  and a fold-change  $> |2|$ . ¶

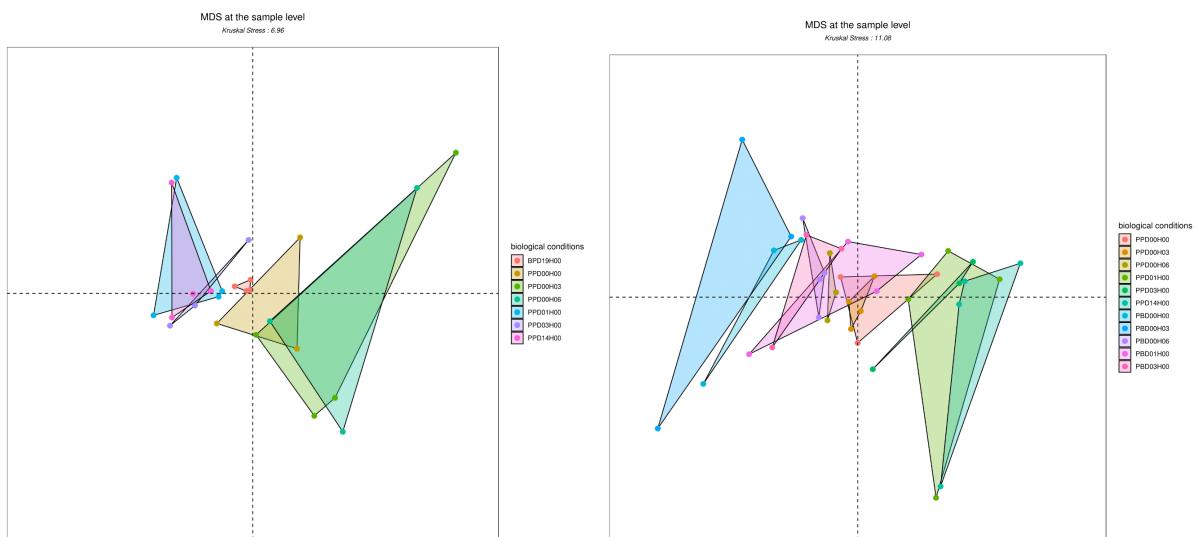
We then sought to find out whether the variations in metacluster abundances were significant or not post-prime and post-boost (**Fig. 6**). In line with our kinetics analyses, metacluster VI

—supposedly memory effector NK cells—is differentially abundant (up-regulated) between these conditions. We also confirmed that metaclusters IV and V are downregulated post-prime. Metacluster III is not differentially abundant.¶

¶

## Global similarities and differences between biological conditions and individuals¶

To visualize the global temporal changes in NK cell compositions caused by the first and second MVA-B doses, MDS representations on all samples and all timepoints were performed (**Fig. 7**). The first one compares before-prime and post-prime conditions to illustrate the effect of the first dose, while the second MDS contrasts the post-prime and post-boost conditions to determine the booster’s effect. Our MDS representations had low Kruskal Stress percentages of 6.96% and 11.08% for first and second MDS respectively, indicating that little information was lost in the dimensionality reduction step. ¶



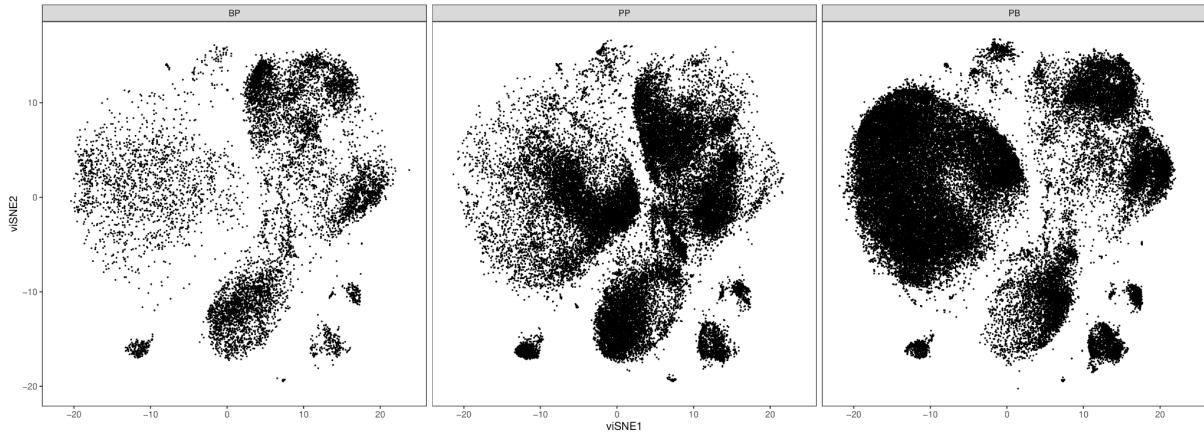
**Figure 7. Visualization of the similarity between samples and between biological conditions in the form of Multidimensional Scaling.** Each color corresponds to a biological condition and each dot to an individual. The clusters used for the analysis are the differentially abundant clusters among NK cells clusters, between before prime and post-prime (A) and between post-prime and post-boost (B). ¶

¶

On **Fig. 7A**, we observe that individuals are similar before prime. After the prime, we see a modification of the phenotype in certain way (to the right side of the plot) between day 0 hour 0 and day 0 hour 6 while there is a divergence among individuals, and starting from day 1 the phenotype is modified in another way (to the left side of the plot) while there is a convergence between individuals. ¶

The second MDS (**Fig. 7B**) shows that individuals almost completely overlap 0-3 hours post-prime and begin diverging after 6 hours. On days 1-14, individuals form a “late post-prime” cluster that doesn’t overlap with “early post-prime”. Remarkably, the same pattern is observed post-boost with early and late post-boost clusters. For any same time point post-prime and post-boost we observe no overlap suggesting that NK cell responses differ between prime and boost. ¶

## t-SNE representation¶



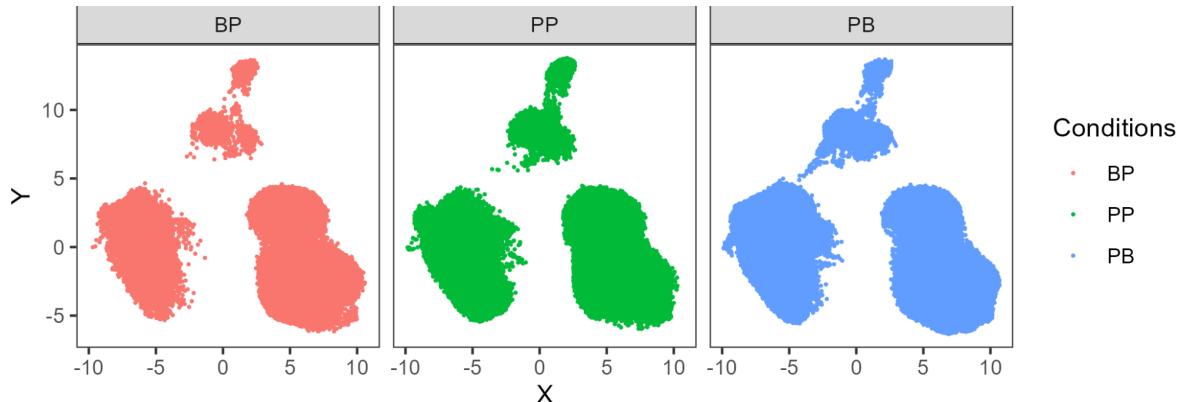
**Figure 8. t-SNE (t-distributed stochastic neighbor embedding) representation using a sample of 100 000 for each condition. Perplexity = 30, maximum iterations = 1000. (BP = Before prime, PP = Post prime, PB = Post boost).¶**

¶

t-SNE representations allow the visualization of high dimensional data. Although it is a sample, it is especially interesting because the data points seem to clusterize from BP to PB condition, from a cluster to another, indicating there may be potentially a phenotypic switch in NK cells, associated with an increase in NK cells by proliferation, probably due to the vaccine response. We can see in **Fig. 8** for example the cloud of points in the upper part of the t-SNE seems to not only aggregate more data as time goes on but that some points seem to shift in the bottom cloud, potentially suggesting a phenotypic switch to another area in the t-SNE.¶The increase in points across timepoints makes sense as cell count and proliferation increases over time.

## UMAP representation¶

¶



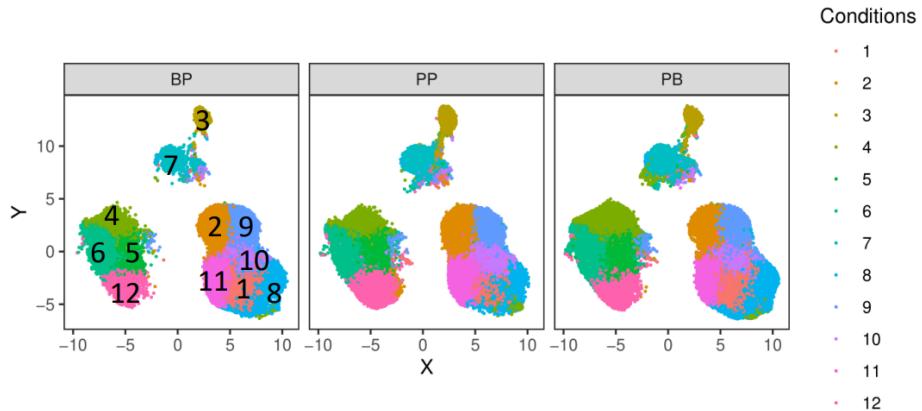
**Figure 9. UMAP (Uniform Manifold Approximation and Projection) representation of the NK cells dataset separated for the three different conditions. Number of neighbors = 15, minimal distance = 0.05. Each dot corresponds to a cell, the color corresponds to the biological condition.¶**

¶

The UMAP representation seems to suggest a great overlap between the three conditions as the points are globally located in the same positions. However what seems apparent is the

appearance of three great clusters. Similarly to t-SNE, the number of points seem to grow between timepoints suggesting an increasing number of cells.¶

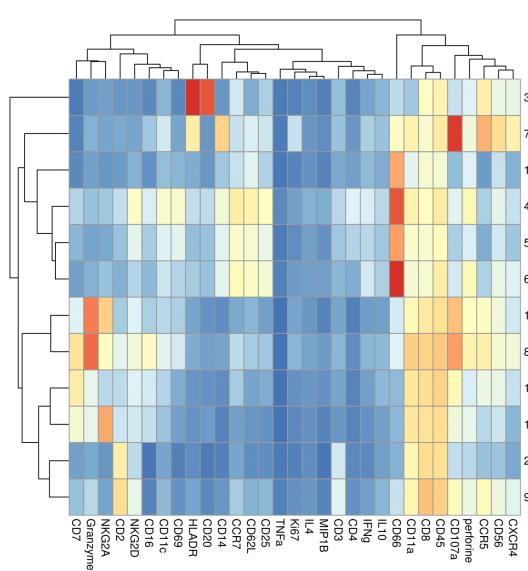
## k-means representation



**Figure 10.** *k-means clustering algorithm on the precedent UMAP representation.* Number of clusters  $k = 12$ . ¶ Each dot corresponds to a cell, the color corresponds to its appartenance to the cluster provided by the k-means algoritm.

A k-means clustering shows that among the three big clouds of data points, there exists heterogeneity of clusters inside. In order to understand which clusters belong to each marker and subsequently cell populations, a heatmap representation of the mean of expression between clusters and cell markers has been done (*Fig. 10*). Interestingly, this heatmap seems to show three different groups of genes as shown in the vertical dendrogram, which may account for the three different groups of big clouds displayed on the UMAP. Additionally, it is possible to view the marker expression associated with each number and identify a cell population based on their marker expression. The horizontal dendrogram seems to show two big cluster groups which are further divided in the top and bottom group. Looking at GranzymeB, NK cells population seems to fit cluster 8 and 1, and could possibly fit cluster 10 and 11 as well due to exclusive GranzymeB expression along with NKG2A expression typical of NK cells and CD8.

When we look back at our UMAP (*Fig. 10*), we realize that clusters 1, 8, 10, and 11 co-locate in the right cloud. This could surrogate that they are indeed containing NK cells. We also note that cluster 2 and 9 are also colocated in the same place.



Based on the relative positivity of CD3 in those last clusters in the heatmap (**Fig. 11**) we could hypothesize that they are T cells that have been brought closer to NK cells by the algorithm because of their NKG2D positivity.

The UMAP also shows that the NK cell clusters (1, 8, 10, 11) do not seem to expand, retract or move drastically over time.

**Figure 11.** Mean expression of each marker according to its clusters from the previous k-means clustering.

*Blue to red color gradient is correlated with the intensity of marker expression.*

## References

- Alter, G., Malenfant, J.M., and Altfeld, M. (2004). CD107a as a functional marker for the identification of natural killer cell activity. *J Immunol Methods* *294*, 15–22.
- Cho, M.S., Lee, M.K., Yang, W.I., and Yoon, J.H. (1997). Granzyme B immunoreactivity in T/natural killer cell lymphomas. *Yonsei Med J* *38*, 285–293.
- Gautreau, G., Pejoski, D., Grand, R.L., Cosma, A., Beignon, A.-S., and Tchitchev, N. (2017). SPADEVizR: an R package for visualization, analysis and integration of SPADE results. *Bioinformatics* *33*, 779–781.
- Gomez, C., Najera, J., Krupa, M., and Esteban, M. (2008). The Poxvirus Vectors MVA and NYVAC as Gene Delivery Systems for Vaccination Against Infectious Diseases and Cancer. *Curr Gene Ther* *8*, 97–120.
- National Institute of Health (2016). HIV Treatment: The Basics | NIH.
- Price, P.J.R., Torres-Domínguez, L.E., Brandmüller, C., Sutter, G., and Lehmann, M.H. (2013). Modified Vaccinia virus Ankara: Innate immune activation and induction of cellular signalling. *Vaccine* *31*, 4231–4234.
- Qiu, P., Simonds, E.F., Bendall, S.C., Gibbs, K.D., Bruggner, R.V., Linderman, M.D., Sachs, K., Nolan, G.P., and Plevritis, S.K. (2011). Extracting a Cellular Hierarchy from High-dimensional Cytometry Data with SPADE. *Nat Biotechnol* *29*, 886–891.
- Rerks-Ngarm, S., Pitissutthithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., Premsri, N., Namwat, C., Souza, M. de, Adams, E., et al. (2009). Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand. *New Engl J Medicine* *361*, 2209–2220.
- Sandström, E., Nilsson, C., Hejdeman, B., Bråve, A., Bratt, G., Robb, M., Cox, J., VanCott, T., Marovich, M., Stout, R., et al. (2008). Broad Immunogenicity of a Multigene, Multiclade HIV-1 DNA Vaccine Boosted with Heterologous HIV-1 Recombinant Modified Vaccinia Virus Ankara. *J Infect Dis* *198*, 1482–1490.
- Sutter, G., and Staib, C. (2003). Vaccinia Vectors as Candidate Vaccines: The Development of Modified Vaccinia Virus Ankara for Antigen Delivery. *Curr Drug Target -Infectious Disord* *3*, 263–271.
- Vargas-Inchaustegui, D.A., Ying, O., Demberg, T., and Robert-Guroff, M. (2016). Evaluation of Functional NK Cell Responses in Vaccinated and SIV-Infected Rhesus Macaques. *Front Immunol* *7*, 340.
- Vasan, S., Schlesinger, S.J., Chen, Z., Hurley, A., Lombardo, A., Than, S., Adesanya, P., Bunce, C., Boaz, M., Boyle, R., et al. (2010). Phase 1 Safety and Immunogenicity Evaluation of ADMVA, a Multigenic, Modified Vaccinia Ankara-HIV-1 B'/C Candidate Vaccine. *Plos One* *5*, e8816.
- Webster, R.L., and Johnson, R.P. (2005). Delineation of multiple subpopulations of natural killer cells in rhesus macaques. *Immunology* *115*, 206–214.
- Wilk, A.J., and Blish, C.A. (2018). Diversification of human NK cells: Lessons from deep profiling. *J Leukocyte Biol* *103*, 629–641.
- World Health Organization (2021). HIV/AIDS.