

Tech Note



Single-Cell RNA Sequencing

Single-cell RNA Sequencing (scRNAseq) is an emerging technology that allows the analysis of the whole coding transcriptome of each individual cell in complex samples. With downstream cluster analysis and differential analysis of the transcriptomic profiles, the marker genes of the different cell clusters can be determined, and differences in gene expression patterns can be identified. Thus, different cell types and cellular subgroups, as well as their molecular characteristics, can be identified, quantified, and compared against each other. This makes scRNAseq particularly suitable for heterogeneous and cellularly diverse samples and allows in-depth analysis of the development and progression of complex diseases.

In this Tech Note, we investigate the accuracy and variability of scRNAseq using the 10x Genomics single-cell gene expression workflow and how this technique can be used to identify cell composition shifts in real-world samples. During 10x-based scRNAseq, cell suspensions are processed together with specific gel beads that tag all mRNA molecules of every single cell with cell-specific barcode sequences and unique molecular identifiers. Hence, each mRNA molecule can be traced back to its original cell, which leads to accurate high-resolution data and allows the analysis of the whole transcriptome of individual cells. This enables the discrimination between different cell types in complex samples and detailed gene expression analysis at the single-cell level.

Methods

For the analysis of assay accuracy, we used four spike-in samples generated by mixing PBMCs (Peripheral Blood Mononuclear Cells) collected from a healthy donor with cells of the tumor cell line HEK293T (10%, 5%, 1%, and 0.5% HEK cells) as well as one PBMC control sample. To test the variability of scRNAseq, we used six PBMC aliquots collected from a healthy donor that were processed in parallel in a single batch as well as in different batches. Finally, we applied the 10x technology to samples from a cancer vaccination patient. Here, we analyzed the patients' PBMCs before and after a peptide-based cancer vaccination. All samples were processed using the Chromium Next GEM Single Cell 3' Kit v3.1 according to our optimized 10x cell preparation workflow. cDNA and final libraries were prepared according to our adjusted 10x protocol to maximize yield and were sequenced on the Illumina NovaSeq 6000 sequencing system using paired-end sequencing aiming at a minimum output of 20,000 reads per cell. Data were analyzed using Cell Ranger (10x Genomics) and the R package „Seurat“ (Stuart et al. 2019). Cell clusters were annotated using the expression of marker

genes. For variability determination and analysis of cell type shifts in samples from a cancer vaccination patient, non-PBMCs (erythrocytes, platelets, neutrophils) and damaged cells were excluded, and only PBMCs were considered.

Results

To demonstrate assay accuracy, we spiked HEK293T cells in defined proportions to a PBMC sample collected from a healthy donor to simulate the presence of a rare cell population. In total, 18,168 cells were analyzed in five samples. As shown in Figure 1, HEK293T cells could be clearly distinguished from PBMCs in all HEK cell-containing samples with only 0.02% of PBMCs that were misassigned to the HEK cluster in the PBMC control. On average, the detected HEK cell proportions in the mixtures showed a relative deviation of 28.7% from the expected values (Table 1). Overall, these results indicate that the presence of rare cellular subpopulations in a complex biological sample can be detected down to a relative abundance of at least 0.5%.

Table 1: PBMC and HEK293T proportions detected in spike-in and control samples.

sample type	mixture	proportion of PBMCs [%]	proportion of HEKs [%]
PBMC	-	99.98	0.02
PBMC:HEK	90:10	88.58	11.42
PBMC:HEK	95:5	92.03	7.97
PBMC:HEK	99:1	98.87	1.13
PBMC:HEK	99.5:0.5	99.36	0.64

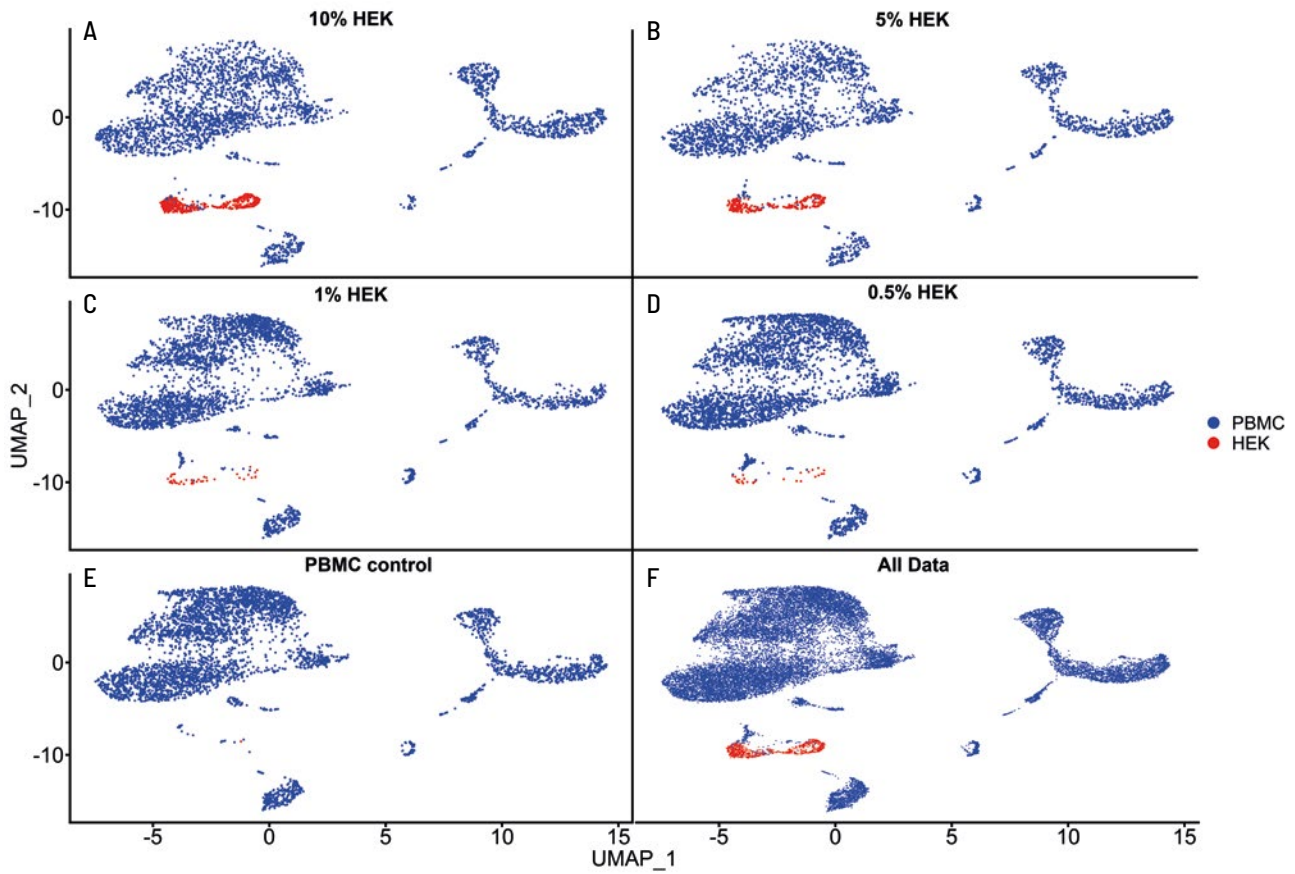


Figure 1: Proportions of PBMC and HEK293T cells detected in spike-in and control samples. A PBMC sample from a healthy donor was spiked with different proportions (10%, 5%, 1%, 0.5%) of the cancer cell line HEK293T. A-E show UMAP plots of the individual samples. F shows a UMAP plot combining all data from all samples. Blue = PBMC cells; Red = HEK293T cells.

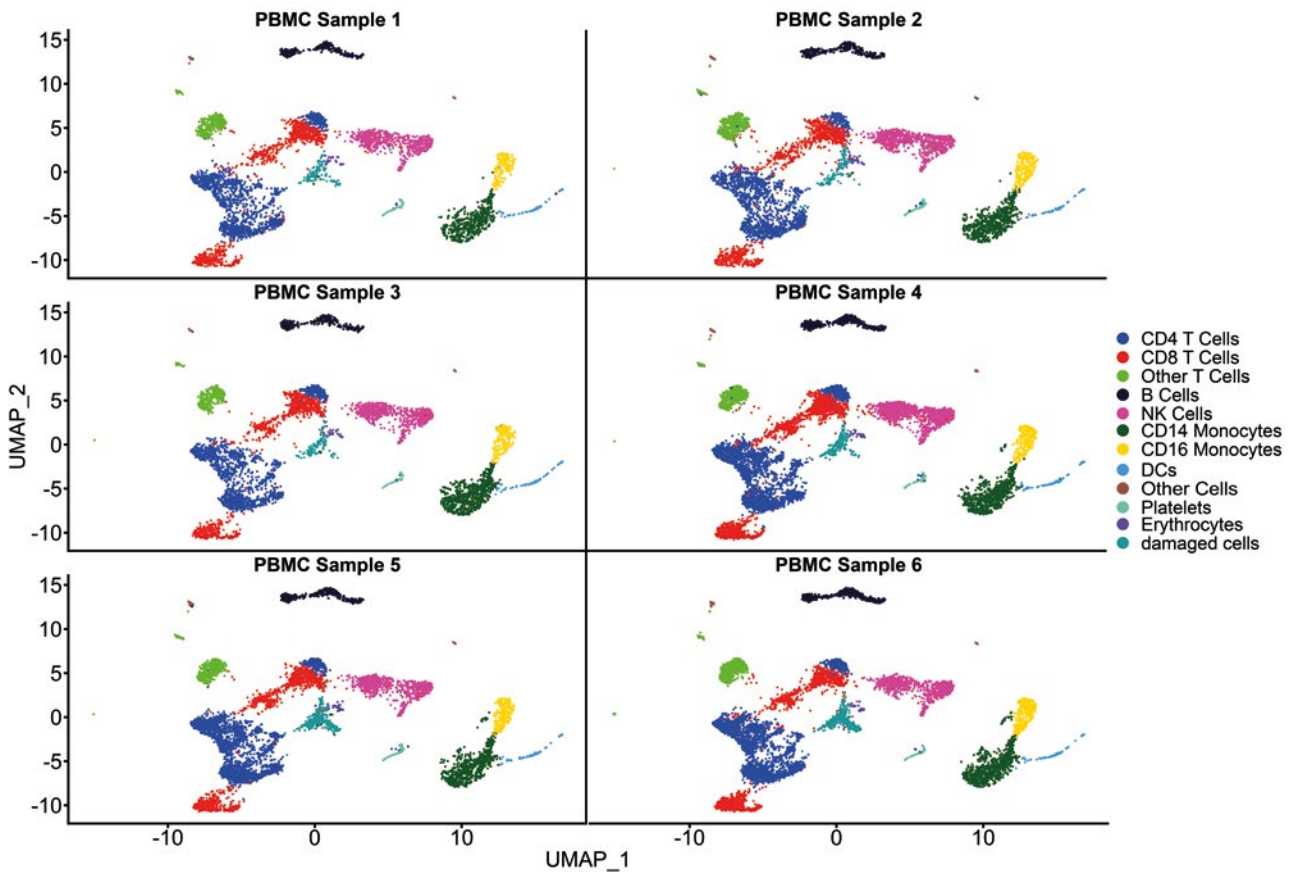


Figure 2: Technical variability in cellular proportions determined in a PBMC sample. Six aliquots of a PBMC sample collected from a healthy donor were processed using our 10x pipeline. Cellular populations (legend) were annotated using cluster analysis and cell type-specific marker genes.

To test for assay variability, a PBMC sample from a healthy donor was analyzed in six replicates yielding 34,919 cells in total. Cell clusters were annotated using marker gene expression in two different resolutions (low and high cell type resolution), and the relative cell cluster proportions, as well as the coefficient of variation (CV), were calculated (Figure 2). The average CV was 12.8% and 16.8% for low and high cell type resolution, respectively. Regarding the CV values of the individual cell clusters, 9 of 9 (low resolution) and 19 of 20 (high resolution) clusters showed CV values below 30% (3.5% - 24.2%; 6.0% - 29.6%) (Table 2). The cluster showing a CV value higher than 30% did not feature a clear marker gene expression profile. Therefore, it is not surprising that cluster assignment for these

cells was inconsistent. In addition, as this cluster consists of very few cells and shows a low relative abundance of 0.5%, little changes like the misassignment of a few cells can have a strong impact on CV values. In contrast, other rare cell types with a low relative abundance of 0.4% - 0.1% (e. g. plasmablast cells, plasmacytoid dendritic cells, and hematopoietic stem and progenitor cells) that featured a clearer marker gene expression profile could be detected with CV values below 30%, confirming and even exceeding the results found in the assay accuracy evaluation. Regarding immune cell type distribution, the proportions of the detected immune cell types were comparable to literature data (Behr-Perst et al. 1999; Böttcher et al. 2019).

Table 2: Average cell type proportions [%] and cell type-specific variability (CV) detected in six PBMC aliquots generated from the same sample of a healthy donor in low (left) and high (right) cell type resolution.

cell type (low resolution)	average proportion [%]	CV [%]	cell type (high resolution)	average proportion [%]	CV [%]
CD4 T Cells	33.6	9.4	CD4 TCM Cells	17.9	8.6
			CD4 naive Cells	9.3	18.7
			CD4 CTL Cells	4.0	20.1
			T regs	2.4	13.7
CD8 T Cells	19.8	5.1	CD8 TEM Cells	9.1	16.5
			CD8 naive Cells	6.5	15.7
			CD8 TCM Cells	4.2	8.2
Other T Cells	7.5	16.8	MAITs	7.0	15.9
			Proliferating T Cells	0.5	43.1
NK Cells	15.2	14.7	NK Cells	15.2	14.7
CD14 Monocytes	11.2	18.3	CD14 Monocytes	11.2	18.3
CD16 Monocytes	4.5	13.2	CD16 Monocytes	4.5	13.2
B Cells	6.1	3.5	mature B Cells	2.5	11.4
			naive B Cells	2.3	6.0
			memory B Cells	1.3	11.6
DCs	1.6	10.3	DCs	1.3	15.4
			pDCs	0.3	18.3
Other Cells	0.5	24.2	Plasmablasts	0.4	29.6
			HSPCs	0.2	19.9

Non-PBMCs (erythrocytes, platelets) and damaged cells were excluded before average cell type proportions and CVs were calculated.
CV = coefficient of variation

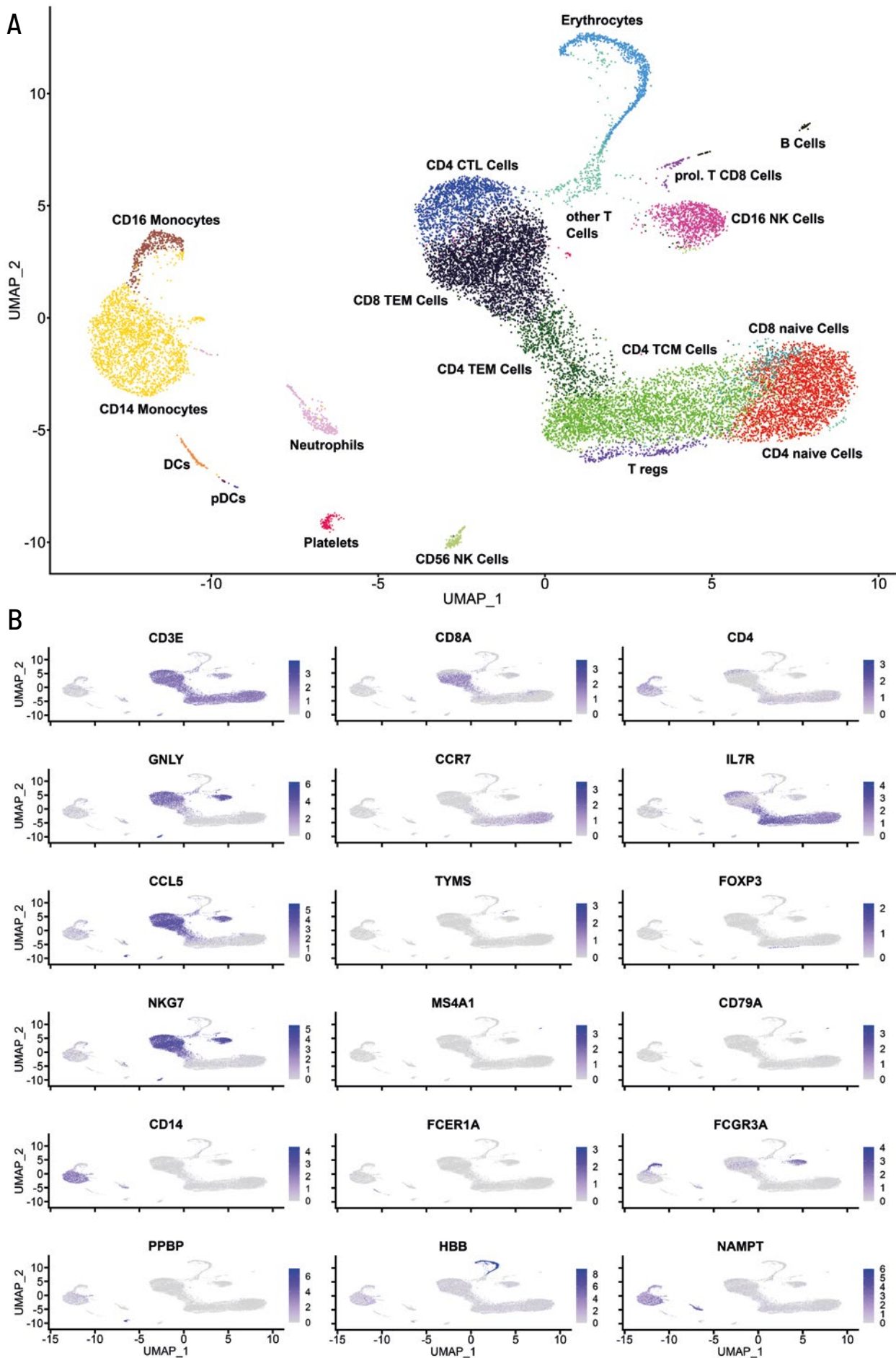


Figure 3: Cellular composition of PBMC samples collected from a cancer vaccination patient. Three PBMC samples were collected from an ovarian cancer patient undergoing individualized anti-cancer peptide vaccination. **A** shows a UMAP plot combining all data from all samples. **B** shows the expression of representative marker genes used for cell type annotation. Increasing expression values are displayed from light to dark blue.

Furthermore, we applied our optimized 10x 3' gene expression workflow to PBMC samples from an ovarian cancer patient that underwent individualized anti-cancer peptide vaccination. Here, we collected PBMC samples before (V1) and after the 7th (V7) and 12th (V12) vaccination and analyzed 17,237 cells in total. The computed cell clusters were annotated using the expression of distinct marker genes as shown in Figure 3B, resulting in an annotated UMAP Plot (Figure 3A). Subsequently, the annotated cell clusters were used to calculate relative cell type abundances and to determine shifts in cell type proportions over the course of vaccination treatments. We found that cell proportions of several PBMC subpopulations varied over time (Figure 4A). To distinguish biological changes from technical noise, we applied a threshold of 2-times the cell type-specific CV determined in the variability experiment that had to be exceeded in both post-vaccination samples as compared to V1. The identified relevant

changes are displayed as fold change compared to the first time point in Figures 4B and 4C. Here, we detected increased proportions of CD4-positive effector and central memory T cells in post-vaccination samples, which is in concordance with simultaneously performed immune monitoring detecting vaccine-induced T cell responses. Interestingly, the patient developed very strong CD4 immune responses against 3 of 10 vaccinated peptides that were not present before treatment. In addition, naive CD4- and CD8-positive as well as other T cell clusters were increased, indicating that the administered cancer vaccination mainly promoted T cell expansion. In contrast, CD14 and CD16 monocytes were decreased over time, probably caused by the applied adjuvant (Sargramostin). Furthermore, B cell proportions decreased over time. All detected biologically relevant changes were in accordance with medical findings and recent publications (Cao et al. 2021).

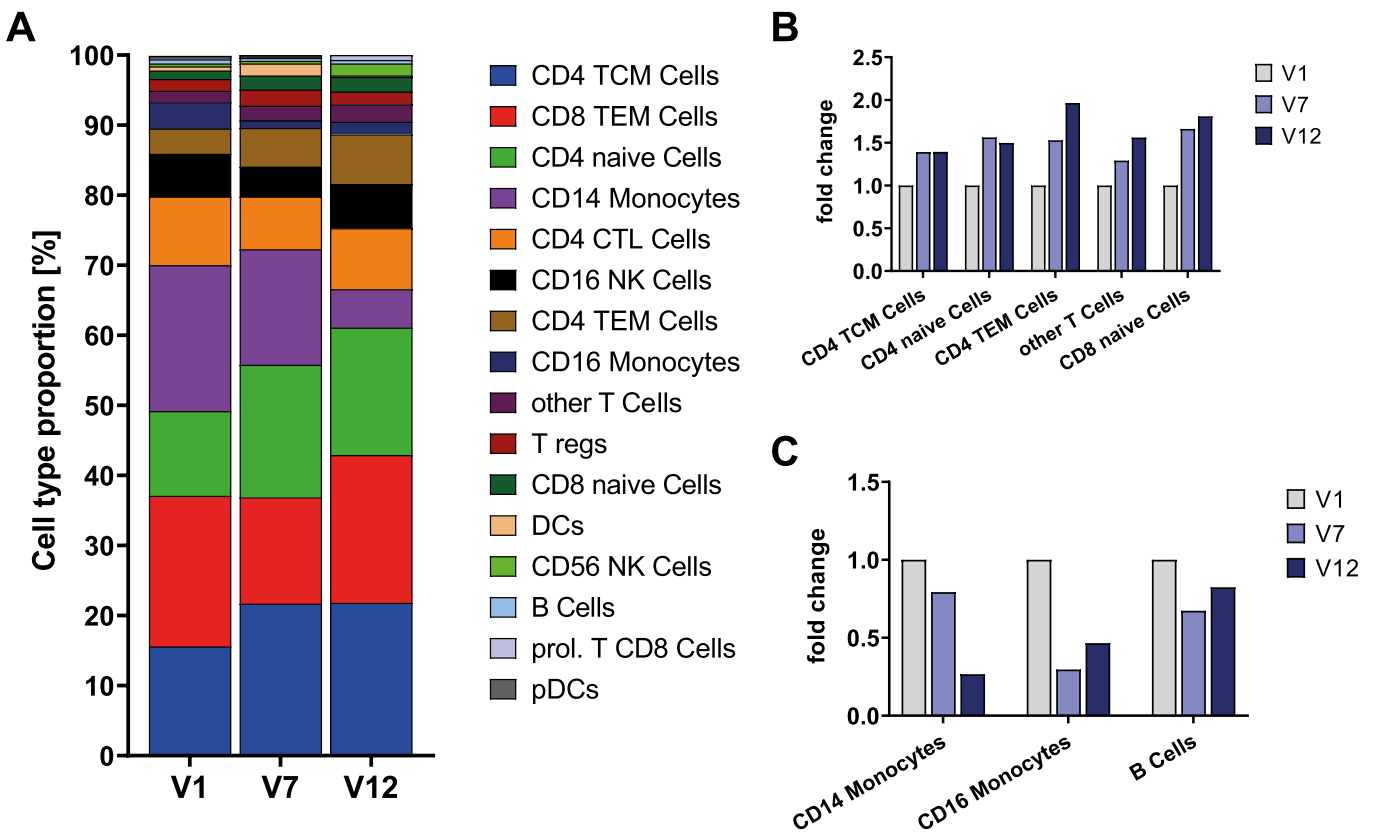


Figure 4: Shifts in immune cell type proportions determined over time in samples collected from a cancer vaccination patient. PBMC samples of a cancer vaccination patient before (V1) and after vaccination (V7, V12) were analyzed, and cell populations were identified using cluster analysis and cell type-specific marker genes. Non-PBMCs (erythrocytes, platelets, neutrophils) were excluded before cell type proportions were calculated. **A** shows a stacked bar plot of the detected immune cell types. **B** shows cell types that were increased after vaccination. **C** shows cell types that were decreased after vaccination. To distinguish biological changes from technical noise, we used a threshold of 2x the cell-specific CV determined using PBMC samples collected from a healthy donor (Table 2). The changes determined for the cell types shown in **B** and **C** exceeded this threshold and are thus considered biologically relevant.

Conclusion

Our analytical evaluation revealed that scRNAseq using the 10x technology is a robust tool to detect rare cell populations in complex samples down to a relative abundance of 0.5%. Our optimized scRNAseq workflow showed a relatively low technical variability with average CV values of 12.8% for low and 16.8% for high cell type resolution. Furthermore, it was very well suited to determine cell type proportion shifts in samples collected over time. In conclusion, these results emphasize the great potential of scRNAseq for the discovery of new biomarkers or in-depth analysis of therapeutic mechanisms by analyzing differential gene expression or changes in cellular diversity on a cell-by-cell basis.

References

Behr-Perst, S. I.; Munk, M. E.; Schaberg, T.; Ulrichs, T.; Schulz, R. J.; Kaufmann, S. H. (1999): Phenotypically activated gammadelta T lymphocytes in the peripheral blood of patients with tuberculosis. In: *The Journal of infectious diseases* 180 (1), S. 141-149. DOI: 10.1086/314844.

Böttcher, Chotima; Fernández-Zapata, Camila; Schlickeiser, Stephan; Kunkel, Desiree; Schulz, Axel R.; Mei, Henrik E. et al. (2019): Multi-parameter immune profiling of peripheral blood mononuclear cells by multiplexed single-cell mass cytometry in patients with early multiple sclerosis. In: *Scientific reports* 9 (1), S. 19471. DOI: 10.1038/s41598-019-55852-x.

Cao, Qiqi; Wu, Shipo; Xiao, Chuanle; Chen, Shuzhen; Chi, Xiangyang; Cui, Xiuliang et al. (2021): Integrated single-cell analysis revealed immune dynamics during Ad5-nCoV immunization. In: *Cell discovery* 7 (1), S. 64. DOI: 10.1038/s41421-021-00300-2.

Stuart, Tim; Butler, Andrew; Hoffman, Paul; Hafemeister, Christoph; Papalexi, Efthymia; Mauck, William M. et al. (2019): Comprehensive Integration of Single-Cell Data. In: *Cell* 177 (7), 1888-1902.e21. DOI: 10.1016/j.cell.2019.05.031.



About Us

CeGaT was founded in 2009 in Tübingen, Germany. Our scientists are specialized in next-generation sequencing (NGS) for genetic diagnostics, and we also provide a variety of sequencing services for research purposes and pharma solutions. Our sequencing service portfolio is complemented by analyses suited for microbiome, immunology, and translational oncology studies.

Our dedicated project management team of scientists and bioinformaticians works closely with you to develop the best strategy to realize your project. Depending on its scope, we select the most suitable library preparation and sequencing conditions on our Illumina platforms.

We would be pleased to provide you with our award-winning service. Contact us today to start planning your next project.

CeGaT GmbH
Research & Pharma Solutions
Paul-Ehrlich-Str. 23
72076 Tübingen
Germany

Phone: +49 707156544-333
Fax: +49 707156544-56
Email: rps@cegat.com
Web: www.cegat.com



Accredited by DAKkS according to
DIN EN ISO/IEC 17025:2018



CLIA CERTIFIED ID: 99D2130225