Paper Review

„SCExecute: custom cell barcode-stratified analyses of scRNA-seq data"

Mara Spataru

May 16, 2025

1 INTRODUCTION

RNA sequencing for individual cells (scRNA-seq), which enables the analysis of genetic expression at the cellular level, is one of the most important technological developments in molecular biology over the last decade. With significant applications in disciplines including oncology, neuroscience, immunology, and developmental biology, this technology has fundamentally transformed our understanding of cellular heterogeneity, development trajectories, and transcriptome dynamics. From the first time it was introduced, scRNA-seq has constantly improved, offering thus a greater resolution, an increased sensitivity and lower costs per cell.

When it comes to scRNA-seq, cellular barcodes are essential for recognising and linking RNA sequences to the specific cells from which they originate. These barcodes, which are synthetic nucleotide sequences that are usually 10–16 nucleotides long, are affixed to the RNA molecules while the sequencing library is being prepared. The "barcoding" procedure begins in the first phase of the experiment, when each cell is separated (using techniques like cell sorting or microfluidics) and lysed to liberate the RNA molecules. At this stage each cell (or its content) is associated with a unique barcode, which is then incorporated in all cDNA molecules derived from the respective cell's RNA through processes of reverse transcription and amplification. This molecular labeling later enables "demultiplexing", which is the computational process of separating sequences based on their origin, and the stratification of data according to their cellular source, even if the actual sequencing is performed on an aggregated library that contains genetic material from thousands or tens of thousands of cells. Although there are numerous computational instruments developed to analyse scRNA-seq data, like Seurat, Scanpy, or Monocle, the majority of them are mainly focused on cuantifying the genic expression (the number of transcripts per gene per cell), cell clustering based on expression profiles and the identification of specific markers for cell populations, thus neglecting other potentially relevant transcriptomic characteristics that could be extracted from these data rich in biological information.

Amongst the less explored aspects of scRNA-seq data are the expressed genetic variations, allele-specific expression, and alternative splicing events at the cellular level. Expressed genetic variations (such as single nucleotide polymorphisms or SNVs) represent changes in the nucleotide sequence of transcripts that can influence the function of the resulting proteins or the regulation of gene expression. In the context of single cell sequencing, detecting these variants can highlight the genetic heterogeneity between cells that appear to be similar morphologically or functionally. Allele-specific expression refers to the phenomenon in which only one of the two copies (alleles) of one gene is expressed in a cell; a crucial aspect in understanding genomic imprinting, X-chromosome inactivation, and other epigenetic mechanisms that contribute to phenotypic diversity. Alternative splicing, a process in which the exons of a messenger pre-RNA can be included or excluded in the final mature transcript, is generating multiple protein isoforms with potentially different functions from the same gene. Identification of splicing patterns at the single cell level may reveal cell-specific post-transcriptional regulatory mechanisms relevant to cellular differentiation, development and the pathogenesis of various diseases. All these molecular characteristics have a significant potential in understanding cellular functional heterogeneity, gene expression regulation mechanisms, and post-transcriptional modifications. However, analyzing these characteristics requires specific methods for stratifying sequences based on cell barcodes and individually processing the data corresponding to each cell, an aspect that is not sufficiently supported by existing high-performance sequencing tools, which were initially designed for population-level (bulk) analyses or focus solely on gene expression quantification.

Current methods for specific scRNA-seq analysis, such as STARsolo and CellRanger, integrate the sequence alignment process with the simultaneous demultiplexing of barcodes and the assignment of read counts to genes. Demultiplexing is the computational process through which collectively generated DNA or RNA sequences (in "bulk") are separated and assigned to their individual cellular origins based on unique barcodes. In practice, the demultiplexing algorithms scan each read sequence to identify the cellular barcode, then compare it with a predefined list of valid codes (to correct the possible sequencing errors), and then group all sequences from the same cell. This essential process is transforming aggregated sequencing data in a cell-stratified format, allowing fro cellular-level analysis. The instruments mentioned are mainly designed to generate matrixes of genic expression, where rows represent genes (or other genomic characteristics like exons or transcripts), and the columns represent individual cells, each value from the matrix indicating the number of RNA molecules detected for the respective gene in that cell. Although extremely efficient for their main purpose (quantification of expression and identification of cell types) these tools do not facilitate the extraction and analysis of other transcriptiomic characteristics at the cellular level, such as expressed genetic variability (SNVs, insertions, deletions), alternative splicing patterns, or differential us-

age of polyadenylation sites, aspects that could provide valuable insights into functional heterogeneity and cellular regulatory mechanisms.

Recent emerging studies, however, demonstrate the significant potential of scRNA-seq data for the analysis of expressed genetic variants, allele-specific expression, and splicing events at the cellular level. These analyses can offer valuable information about the functional heterogeneity of cells, mechanisms of regulating the genic expression, and post-transcriptional modifications. To design such advanced analyses, the development of some methods that allow the efficient stratification of sequencing data depending on the cellular barcodes and applying specific analytic instruments on subsets corresponding to each cell is needed.

In this context, the analysed article "SCExecute: custom cell barcode-stratified analyses of scRNA-seq data", published in Bioinformatics in January 2023, presents an innovative solution for this problem. The authors Nathan Edwards, Christian Dillard, NM Prashant, Liu Hongyu, Mia Yang, Evgenia Ulianova, and Anelia Horvath have developed a software instrument that facilitates personalised analyses at cellular level for scRNA-seq data, allowing the use of existing tools for high performance sequencing on cell-specific alignment files (scBAM).

SCExecute offers a flexible and efficient approach for performing customized, cell-level analyses on scRNA-seq data, developed by the authors of the study in response to the limitations of existing tools. This software acts as an intermediary between raw sequencing data and various analytical tools, enabling their application at the level of individual cells. From a technical standpoint, the tool extracts alignments corresponding to each cellular barcode from the aggregated alignment files (BAM) generated by single-cell sequencing. These alignments are then saved in separate scBAM (single-cell Binary Alignment Map) files, each containing only the sequences from a single cell. SCExecute then executes user-specified commands or programs on each generated scBAM file, whether they are simple commands, monolithic programs, multi-command shell scripts, or complex shell-based pipelines. This modular architecture allows the use of a wide range of existing analytical tools, originally developed for bulk sequencing data (at the population level), to now be applied to cell-specific data without requiring any modifications to those tools. The program uses an efficient memory and process management system, generating scBAM files in batches and executing the specified commands in parallel, thus optimizing processing time and the use of computational resources. Through this innovative approach, SCExecute opens new possibilities for the analysis of transcriptomic features at the cellular level, including the detection of expressed genetic variants (such as polymorphisms or somatic mutations), the analysis of allele-specific expression (relevant for genomic imprinting and X-chromosome inactivation), and the identification of alternative splicing events (important for proteomic diversity and post-transcriptional regulation).

An essential characteristic of SCExecute is its capacity to function with sequencing

data generated by diverse technological platforms of scRNA-seq, thus ensuring optimal integration into existing workflows. The instrument is compatible with the 10X Genomics Chromium system, which is one of the most widely used technologies for scRNA-seq, known for its ability to process thousands of cells simultaneously at a relatively low cost per cell and high sensitivity. The 10X Genomics platform uses microfluidic droplet technology, which encapsulates individual cells in microscopic droplets together with beads containing unique barcodes, thus allowing identification of the cellular origin of each transcript. In addition to 10X Genomics, SCExecute can efficiently process barcodes generated by various software applications specialized in scRNA-seq analysis. CellRanger, developed by 10X Genomics, represents the standard suite for processing the data that was generated on their platform, performing sequence alignment, demultiplexing, and quantification of gene expression. STARsolo, an extension of the popular STAR aligner, provides an open-source alternative for processing scRNA-seq data with performance comparable to or superior to CellRanger in certain scenarios. UMI-tools, another compatible application, focuses on processing unique molecular tags (UMIs), allowing for the removal of PCR duplicates and more accurate quantification of transcripts. By supporting these diverse formats and applications, SCExecute offers broad compatibility with existing single-cell sequencing workflows, facilitating its integration into already established analytical pipelines without the need for reprocessing primary data or substantial modification of analysis protocols.

The authors demonstrate the utility of SCExecute by applying it to the detection of single-cell-specific expressed single nucleotide variants (sceSNVs) from scRNA-seq data generated by the 10X Genomics Chromium system. For this purpose, they use two popular variant identification tools (GATK and Strelka2) executed in shell scripts along with commands for manipulating BAM files and filtering variants. This approach allows the identification of genetic variants specifically expressed at the cellular level, which would have been difficult or impossible to detect through traditional methods of aggregate data analysis.

The study reviewed is notable for its potential to greatly expand the information that can be extracted from existing scRNA-seq data. By facilitating personalized analyses at the cellular level, SCExecute allows researchers to more efficiently exploit single-cell sequencing data, uncovering subtle transcriptomic features that could have significant implications for understanding cell biology, tissue heterogeneity, and molecular mechanisms involved in diverse biological and pathological processes.

In the context of the rapid evolution of sequencing technologies and the increasing volume of scRNA-seq data generated globally, tools such as SCExecute are becoming essential for maximizing the information extracted from these valuable data sets. The exponential growth in single-cell sequencing experiments across diverse biological disciplines has created an urgent need for computational methods that can efficiently process

and analyze this wealth of data. SCExecute addresses this need by providing a scalable solution that enables researchers to acquire deeper insights from existing datasets without requiring specialized computational expertise. By combining the flexibility in processing cellular barcodes with the capacity to execute diverse analytical instruments on data specific for each cell, SCExecute significantly contributes to advancing the field of single-cell genomics and opening new research directions in systems biology and personalized medicine. This tool represents a critical bridge between the growing technical capabilities of sequencing platforms and the biological understanding that can be derived from the resulting data, thereby accelerating the translation of genomic information into clinically relevant insights

2 RESULTS

SCExecute is an innovative software tool that facilitates customized analyses of single-cell RNA sequencing (scRNA-seq) data at the individual cell level. Evaluating the results presented in the reviewed article, several significant contributions can be identified that justify the importance and utility of this development in the field of bioinformatics.

The main concept of SCExecute is the execution of user-defined commands on BAM stratified files depending on the cellular barcode (scBAM), extracted from aggregated single-cell sequencing data. In essence, the instruments works like an intermediary between raw sequencing data and multiple preexisting analytical instruments, allowing their application at the individual cell level. This strategy offers the research the ability to perform customized analyses on data specific to each cell without requiring the development of completely new tools.

The SCExecute implementation is inspired by classic Linux utilities for managing files and executing commands, following the Unix philosophy of creating modular tools that can be combined in powerful ways. The program employs an intelligent approach in which it alternately generates file groups for individual cells and executes user-specified commands on these files, rather than attempting to process all cells simultaneously. This batch processing strategy creates technical advantages at multiple levels of the computational pipeline. The mode of operation elegantly circumvents common operating system limitations. Problems such as restrictions on the number of files that can be opened simultaneously or insufficient memory allocation typically plague large-scale bioinformatic analyses. The developers incorporated advanced memory management techniques that minimize overhead while maximizing throughput, resulting in remarkably efficient resource utilization. For the user, this translates to significantly better performance on modern multi-core computing systems, providing more fine control over how many analyses run in parallel and how many cells are processed at each stage. This technical architecture ensures that SCExecute can scale effectively with increasing dataset sizes

while maintaining computational efficiency and robustness.

An important aspect of SCExecute is the ease with which it can work with different data formats generated by various sequencing platforms. The program can recognize and extract cell barcodes from the resulting sequencing files, regardless of the program that was used for the initial data processing (CellRanger, STARsolo or UMItools). This broad compatibility means that SCExecute can be easily integrated into existing workflows in research laboratories, without requiring major changes to established protocols.

A valuable feature of SCExecute is the ability to restrict analysis to specific genomic regions of interest. This allows researchers to focus on specific genes or segments of the genome, saving time and computational resources when the entire genome is not required to be analyzed. This option is particularly useful in studies focused on specific metabolic pathways or biological processes, allowing for deeper analysis in relevant regions.

SCExecute also provides the option to automatically index the generated scBAM files before executing user-specified commands. Indexing facilitates rapid access to information from alignment files and is required for many genomic analysis tools. This automation eliminates a manual step that would otherwise be required in the workflow.

To demonstrate the practical utility of SCExecute, the authors analyzed 10 public single-cell sequencing datasets from diverse tissue types and cell lines, including neuroblastoma tumors, prostate cancer, and MCF7 cells (a breast cancer cell line commonly used in research). After aligning the sequences to the reference genome, they used SCExecute to identify genetic variants at the level of each cell. This analysis led to the discovery of approximately 70,000 genetic variants expressed in at least two cells per sample. Interestingly, between 11% and 36% of these variants were not previously known in standard genetic databases, suggesting the potential of SCExecute to contribute to the discovery of new genetic variants with possible biological significance.

SCExecute's flexibility is also demonstrated by the ability to specify templates for the names of the generated scBAM files, working directories, and names for standard outputs. This customization capability allows for efficient organization of results and effortless integration into complex workflows, tailored to the specific needs of different research projects.

The performance tests show that SCExecute is significantly more efficient tan other existing approaches for similar analyses. Compared to the standard method using the samtools program (a commonly used toolkit for manipulating sequencing data), SCExecute reduces execution time by 35-66%. Samtools is a collection of programs used for manipulating alignment files in SAM/BAM format, but it is not optimized for processing single-cell sequencing data. By comparison, SCExecute is specifically designed for this type of data and manages computing resources more efficiently.

Another important aspect highlighted in the results is the computational efficience of SCExecute. For a dataset with approximately 1800 cells, the total time for analysing was

approximately 12 hours on a system with 8 processors, with minimal overhead attributable to the program itself (less than 5%). This means that the vast majority of time is dedicated to the actual analysis of the data, not the technical aspects of file management or barcode extraction.

The obtained results have been visualised using modern techniques of dimensionality reduction, which allow the graphical representation of relationships between cells in two-dimensional space. These visualizations allowed the identification of patterns of distribution of genetic variants in different cell types and highlighted interesting relationships between the genetic profiles of cells and their functional characteristics. For example, the authors could observe how some genetic variants are present only in a cellular subpopulation, suggesting possible connections to the specific functions of these cells.

Besides identifying genetic variants, SCExecute has also proven useful in allele-specific expression studies, allowing the analysis of allelic balance at the single cell level. This application is particularly valuable for the study of genes with monoallelic expression or allelic imbalance, phenomena important in processes such as X chromosome inactivation or genomic imprinting.

An important feature of SCExecute is its complementarity with other popular scRNA-seq data analysis tools, such as Seurat or Scanpy. This allows for the correlation of genetic variation information with identified cell types and developmental trajectories, providing a more complete picture of cell biology.

Scalability is another strength of SCExecute, with results demonstrating that the program can efficiently process datasets of varying sizes, from a few hundred to tens of thousands of cells. The system automatically adapts to the available computational resources, optimizing performance based on the hardware configuration used.

A major benefit of SCExecute is its ability to highlight transcriptomic features that would be difficult to detect using standard methods. The program allows the identification of genetic variants, alternative splicing events, or expression patterns that occur only in a small number of cells. This capability is essential for a complete understanding of cellular diversity and the molecular mechanisms involved in normal or pathological biological processes.

In conclusion, the results presented in the article demonstrate that SCExecute represents a valuable tool for researchers working with single-cell sequencing data, addressing a critical gap in the bioinformatics methods that are now available to the scientific community. The program facilitates the extraction of complex biological information, going beyond the simple measurement of gene expression levels to reveal subtler aspects of cellular heterogeneity and molecular function. The authors have demonstrated through multiple use cases that SCExecute can effectively identify genetic variants, alternative splicing events, and allele-specific expression patterns that would otherwise remain hidden in aggregate analyses. By combining computational efficiency with flexibility in data

analysis, SCExecute opens up new possibilities for fully exploiting the wealth of information contained in modern sequencing data. The ability to apply existing analytical tools to cell-specific data without modification represents a significant advance in the accessibility of sophisticated analyses, democratizing access to cutting-edge computational methods across the biomedical research community. The demonstration of substantial performance improvements over conventional approaches further establishes SCExecute as an important contribution to the bioinformatics ecosystem.

3 CONCLUSIONS

The SCExecute tool, described in the reviewed article, represents a valuable contribution to the field of bioinformatics and single-cell RNA sequencing (scRNA-seq) data analysis. By facilitating personalized analyses at the cellular level, SCExecute addresses a real need in the scientific community, providing an efficient solution for studying transcriptomic features beyond simple gene expression quantification.

3.1 Impact and relevance of the study

The emergence and rapid development of single-cell RNA sequencing technologies over the past decade has generated huge volumes of data with tremendous information potential. However, most existing analytical tools focus almost exclusively on quantifying gene expression and identifying cell types, leaving other valuable aspects of the cellular transcriptome unexplored. In this context, SCExecute brings a significant contribution by creating a bridge between scRNA-seq data and the numerous bioinformatic traditional instruments developed to analyse mass sequencing.

Published in the journal Boinformatics in January 2023, the study already has a visible impact on the scientific community, being cited in multiple subsequent papers exploring various aspects of cellular heterogeneity and genetic variability at the single cell level. This rapid interest demonstrates the utility of the tool and its potential to open new research directions in cellular-resolution genomics.

The scientific relevance of SCExecute is highlighted by its ability to facilitate the detection of cell-specifically expressed genetic variants (sceSNVs), many of which were previously unreported in standard databases such as dbSNP. The discovery of these variants may have important implications for the understanding of cellular heterogeneity, cellular development and differentiation processes, as well as the molecular mechanisms involved in various pathologies, especially those characterized by genetic mosaicism.

3.2 Innovative approach and the advantages of the implementation

A particularly valuable feature of SCExecute is its innovative way of working, which allows efficient stratification of sequencing data based on cellular barcodes and the application of existing analytical tools to cell-specific files. This approach eliminates the need to develop entirely new tools for analyzing scRNA-seq data and leverages the rich ecosystem of bioinformatics software already available.

SCExecute's modular architecture, inspired by the design principles of classic file manipulation tools, offers remarkable flexibility and easy integration into existing workflows. Support for various cell barcode formats and compatibility with popular platforms such as 10X Genomics Chromium, as well as processing tools such as CellRanger, STARsolo and UMI-tools, ensure the broad applicability of SCExecute in research laboratories.

Efficient management of computational resources is another important advantage of the implementation. By generating and processing scBAM files in batches, SCExecute optimizes the use of available memory and processors, significantly reducing analysis time compared to alternative approaches. This computational efficiency is essential in the context of the large volumes of data generated by modern single-cell sequencing, which can include thousands or tens of thousands of cells per experiment.

The ability to restrict analysis to specific genomic regions constitutes a substantial practical benefit, allowing researchers to focus on genes or sequences of interest and save valuable computational resources. This functionality is very useful in studies focused on certain metabolic pathways, candidate genes or regulatory regions.

3.3 Its applicability in bioinformatic research

SCExecute opens new perspectives in diverse fields of biomedicine, offering the possibility to study transcriptomic characteristics that would otherwise remain undetected in traditional analyses. I believe that SCExecute could have valuable applications in fields like oncology, neurobiology, or immunology, where analysing the transcriptomic differences at single-cell level is essential for understanding the cellular heterogeneity and the molecular mechanisms involved.

Through its characteristics, I believe that SCExecute could highly contribute to more studies in the following directions:

Developmental Biology: Study of dynamic transcriptomic changes during embryonic development and cellular differentiation, including alternate splicing events and variability in specific allelic expression, which can influence cellular developmental trajectories.

Regenerative medicine: Detailed analysis of stem and progenitor cell populations, including identification of subpopulations with different regenerative potential based on

their genetic and transcriptomic profiles.

These potential applications demonstrate the versatility and possible impact of SCExecute in advancing scientific knowledge and improving clinical approaches in various fields of medicine.

3.4 Limitations and directions for improvement

Despite its valuable contributions, SCExecute also has some limitations that could be addressed in future versions. An obvious limitation is its dependence on the quality of the initial alignments and the barcode demultiplexing process. Errors in these preliminary steps can affect the accuracy of subsequent analyses performed with SCExecute.

Another challenge is the management of cells with low sequencing coverage, which can generate false negatives in the detection of genetic variants or splicing events. Implementation of adaptive statistical methods, which take into account the variable sequencing depth between cells, could improve the sensitivity and specificity of the analyses.

The current command-line interface may be a barrier for users without bioinformatics or programming experience. The development of a graphical interface or web-based implementation could expand the accessibility of the tool to a wider research community.

Scalability is another potential area for improvement. Although SCExecute is efficient for the datasets analyzed in the current study, the continued increase in the size and complexity of single-cell sequencing data may require further optimizations to efficiently handle very large datasets, possibly by implementing distributed computing strategies.

Direct integration with popular scRNA-seq data analysis platforms, such as Seurat and Scanpy, would facilitate smoother workflows and allow immediate correlation of results obtained through SCExecute with other cellular features identified by these tools.

3.5 Development perspectives and scientific contribution

SCExecute opens up multiple promising directions for future research. In the short term, its application to identify genetic variants expressed at the cellular level in diverse biological and pathological contexts could generate valuable information about cellular heterogeneity and the molecular mechanisms involved.

In the mid-term, expanding functionalities to include specific analyses of alternative splicing events, differential usage of polyadenylation sites, or post-transcriptional modifications could further leverage the wealth of information contained in scRNA-seq data.

In the long term, integrating cellular-level genetic variability data with other types of omics data, such as proteomics or single-cell metabolomics, could lead to a deeper understanding of the complex relationships between genotype, gene expression, and cellular phenotype.

SCExecute's scientific contribution is reflected in its potential to transform the way the research community leverages single-cell sequencing data. By facilitating personalized analyses at the cellular level, the tool allows for the exploration of aspects of cell biology that would otherwise remain hidden in traditional aggregate analyses.

3.6 Personal conclusions

In my opinion, SCExecute represents an excellent example of an innovative approach in bioinformatics, which responds to a real need in the scientific community and opens up new research possibilities. The conceptual elegance of the tool lies in its fundamental simplicity, that of stratifying data according to cellular barcodes and applying existing analytical tools, combined with a robust and efficient technical implementation.

I particularly appreciate SCExecute's flexibility and ability to integrate into existing workflows, leveraging the rich ecosystem of bioinformatics tools already available. This pragmatic approach allows researchers to apply validated methodologies to single-cell sequencing data without the need to develop entirely new tools.

The results presented in the article, in particular the identification of a significant number of previously unreported genetic variants, demonstrate the potential of SCExecute to contribute to original scientific discoveries. Its application in various fields of biomedical research could generate valuable information about cellular heterogeneity and molecular mechanisms involved in normal and pathological biological processes.

I believe that SCExecute represents an important step in the evolution of single-cell sequencing data analysis methods, contributing to the transition from simple quantifications of gene expression to much more detailed molecular characterizations at the cellular level. This evolution is essential for fully realizing the potential of single-cell sequencing technologies and for advancing our understanding of cell biology in all its complexity.

In conclusion, SCExecute perfectly illustrates how innovations in computational methodologies can catalyze significant advances in the biomedical sciences, demonstrating once again the essential role of bioinformatics in the modern genomics era. By combining sound software design principles with a deep understanding of the specific needs of the single-cell sequencing field, the authors have created a tool that has the potential to accelerate scientific discovery and contribute to the advancement of personalized medicine.

4 REFERINȚE

References

[1] Edwards, N., Dillard, C., Prashant, N. M., Hongyu, L., Yang, M., Ulianova, E., & Horvath, A. (2023). *SCExecute: Custom cell barcode-stratified analyses of scRNA-seq*

data. Bioinformatics, 39(1), btac768. https://doi.org/10.1093/bioinformatics/btac768

[2] Zhao, S., Zhang, B., & Auerbach, R. K. (2020). *RNA-seq analysis: Methods, applications and challenges*. Frontiers Media SA.

[3] Kebschull, J. M., & Zador, A. M. (2018). *Cellular barcoding: Lineage tracing, screening and beyond*. Nature Methods, 15(11), 871–879. https://doi.org/10.1038/s41592-018-0185-x

[4] Roth, C. M. (2002). *Quantifying gene expression*. Current Issues in Molecular Biology, 4(3), 93–100. https://doi.org/10.21775/cimb.004.093

[5] Francesconi, M., & Lehner, B. (2014). *The effects of genetic variation on gene expression dynamics during development*. Nature, 505, 208–212. https://doi.org/10.1038/nature12772

[6] Buckland, P. R. (2004). *Allele-specific gene expression differences in humans*. Human Molecular Genetics, 13(suppl_2), R255–R260. https://doi.org/10.1093/hmg/ddh227

[7] Marasco, L. E., & Kornblihtt, A. R. (2023). *The physiology of alternative splicing*. Nature Reviews Molecular Cell Biology, 24(4), 242–254. https://doi.org/10.1038/s41580-022-00545-z

[8] Wang, X., He, Y., Zhang, Q., Ren, X., & Zhang, Z. (2021). *Direct comparative analyses of 10X Genomics Chromium and Smart-Seq2*. Genomics, Proteomics & Bioinformatics, 19(2), 253–266. https://doi.org/10.1016/j.gpb.2020.02.005

[9] Kaminow, B., Yunusov, D., & Dobin, A. (2021). *STARsolo: Accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data*. bioRxiv. https://doi.org/10.1101/2021.05.05.442755

[10] Kim, S., Scheffler, K., Halpern, A. L., Bekritsky, M. A., Noh, E., Källberg, M., ... Saunders, C. T. (2018). *Strelka2: fast and accurate calling of germline and somatic variants*. Nature Methods, 15(8), 591–594. https://doi.org/10.1038/s41592-018-0051-x