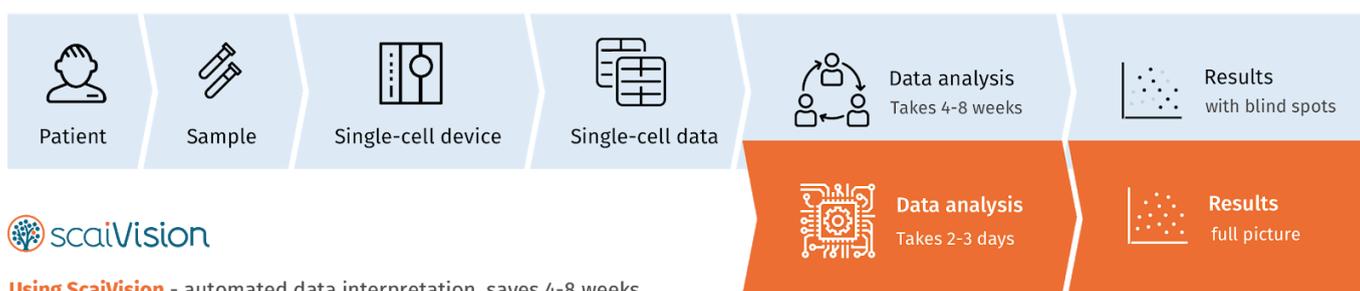




## ScaiVision is an intuitive deep learning software for single-cell data analysis. It helps to discover rare biosignatures in a time-efficient manner.

ScaiVision is a product of Scailyte AG, a Swiss-based leading startup company in the rapidly growing markets of Single-Cell Biology and Artificial Intelligence. We unlock the full potential of single-cell data by providing AI-based technology for the discovery and clinical application of ultra-sensitive Cell-Identity Biomarkers. For more details about us and our work please visit our website: [www.scailyte.com](http://www.scailyte.com)

**Current methods** - requires bioinformatics experts, results in blind spots



## Introduction

The fundamental unit of any living organism is a cell. The human body contains approximately 40 trillion cells forming more than 200 different tissues. For over a century, scientists have been identifying and classifying cells by their size and shape, location, function, and more recently, by their molecular properties. Nevertheless, our knowledge about the individual cell types and cellular states, and consequently also our understanding of the associated biological pathways, has remained rather limited. This is because the majority of commonly used technologies have proven to be unable to capture the full complexity of cellular heterogeneity and the cellular differences between healthy and disease tissues. Standard methods either measure only a few parameters on individual cells or they measure bulk cell populations, averaging the signals from individual cells together, and thus inevitably losing a significant portion of the information within them. This presents a major challenge in our ability to study accurate disease models and to discover and develop efficient treatments. Over the last decade, these issues have motivated scientists to advance single-cell technologies, and today these have become easily accessible and affordable to the biomedical research community.

## Overview of single-cell methods

For decades, flow cytometry has been a primary single-cell method for measuring multiple markers in individual cells<sup>1,2</sup>. It is the most robust single-cell technology but at the same time very limited by the number of features that can be measured simultaneously – typically 3 to 17 markers only<sup>3,4</sup>. A new format of flow cytometry has been developed that leverages the sensitivity of mass spectrometry. This fusion of the two technologies, termed mass cytometry, can provide measurements of over 40 cellular markers simultaneously at single-cell resolution, significantly improving the ability to analyze complex cellular systems and their pathways<sup>5-7</sup>.

At the same time, the development of microfluidic and droplet-based single-cell technologies allows us to investigate thousands of parameters in each individual cell with non-targeted approaches – single-cell genomics<sup>8,9</sup>, transcriptomics<sup>10-12</sup>, and proteomics<sup>13</sup>. These methods are suitable for exploratory studies of cell populations. The most appropriate method to characterize novel cellular populations is by analyzing their transcriptomes using single-cell RNA sequencing. This method can quantify the expression of a large number of genes from individual cells. However, the downside is that the number of cells that can be analyzed per sample is limited, which can be a significant issue when detecting and analyzing rare cell populations. In such cases, the targeted cytometry methods become superior.

Method	Markers		Number of cells	Tissue/samples
	Number	Type		
Flow Cytometry	3-17	Protein	10M+	Liquid
Mass Cytometry (MC)	35-50	Protein	1-10M	Liquid
IMC (Imaging MC)	50-100	Protein, RNA, DNA	100K-1M	Tissues, Formalin-Fixed Paraffin-Embedded (FFPE)
RNA-sequencing	1K-5K	RNA	1K-10K	Liquid

**Table 1.** Comparison of the most commonly used single-cell methods<sup>4</sup>.

## Challenges facing analysis of single-cell data

The advent of single-cell technologies requires the development of new tools in data analysis, as they pose challenges that are not addressable by traditional computational methods<sup>14-16</sup>. These challenges occur at various levels, for example:

- **Stochastic variation:** There is a strong stochastic variation given the usually limited starting source of the material used for such experiments. Technical issues during the wet-lab workflow, such as dropout events and amplification biases, add to the single-cell specific noise. These issues result in technical variation being confounded with biological variation in single-cell datasets.
- **Visualization:** Given the high-dimensionality of single-cell data, the visualization of these datasets poses the next roadblock. This requires stratifying subpopulations of cells whose abundance or behavior is distinct from the rest of the data, which is made challenging by immense issues arising from the biological and technical variations.
- **Relationship between 'cell types':** If one is able to cluster subpopulations within single-cell datasets, it is usually hard to decipher information on the trajectory and relationships between different time points, stages or conditions in datasets that involve these parameters.
- **Modeling:** Finally, current modeling approaches on single-cell datasets are not yet able to fully reveal mechanistic insights into the coordination of gene activities and decode the overall dynamics of the system.

Given the number of steps and considerations involved in analyzing and understanding single-cell datasets, the required time investment is enormous. Advanced data analysis methods and expert computational scientists are necessary to systematically extract quantitative information from single-cell datasets, in order to facilitate addressing biomedical problems through single-cell technologies.

## Our Solution

# ScaiVision: unlocking the potential of single-cell data for pharma and research.

Scailyte has developed ScaiVision, an Artificial Intelligence (AI) based software that offers best-in-class performance for automated pattern-recognition and interpretation of single-cell data. The performance and clinically-relevant application of the core algorithm of our software have been demonstrated in a scientific publication by our co-founder Prof. Claassen<sup>17</sup>.

ScaiVision utilizes *CellCnn*, which is a representation learning approach using convolutional neural networks to detect condition-associated cell subsets from high-dimensional single-cell data. This algorithm builds on developments in artificial intelligence and utilizes deep learning to identify conditions (e.g. disease or drug response) associated with cell populations. In simple words, the network is built to learn the relationship between input data (e.g. labeled single cell datasets) and the associated label (e.g. phenotype). When trained, the network can predict the label from the new samples. ScaiVision outperforms any other currently available analysis methods<sup>17</sup>.

ScaiVision is available as a cloud-based software for biomedical research and pharmaceutical discovery. Its key features address current bioinformatic pipeline limitations for single-cell data :

- a. Ultra-rapid analysis of high-dimensional single-cell data;
- b. Intuitive and user-friendly graphic interface (no computational/bioinformatic background needed);
- c. Automated data interpretation and pattern recognition;
- d. Bioinformatic support to complement and supervise the machine learning functionality.



Single-cell data is collected from a cytometer and labelled depending on the different phenotypes present in the samples

After running ScaiVision, we obtain single-cell biosignatures, which highlight phenotype-related cells (e.g. disease related)

**Figure. 1** General application of ScaiVision: a user-friendly AI-powered software that does not require any bioinformatic skills, accelerates analysis of single-cell data from weeks to days, and discovers novel phenotype-specific biosignatures.

## Outlook

Our ScaiVision software is suitable for analysis of various biological and clinical single-cell data. We are constantly improving our software and adding additional features based on our customers' valuable feedback. Currently, ScaiVision is an ideal solution for the analysis of mass cytometry datasets, and we have ongoing developments for analysis of more complex single-cell datasets such as RNA-sequencing and IMC.

## Citations

1. Fulwyler, M. J. Electronic Separation of Biological Cells by Volume. *Science* **150**, 910–911 (1965).
2. Herzenberg, L. A. *et al.* The History and Future of the Fluorescence Activated Cell Sorter and Flow Cytometry: A View from Stanford. *Clin. Chem.* **48**, 1819–1827 (2002).
3. Perfetto, S. P., Chattopadhyay, P. K. & Roederer, M. Seventeen-colour flow cytometry: unravelling the immune system. *Nat. Rev.* **4**, 1160–1163 (2004).
4. Heath, J. R., Ribas, A. & Mischel, P. S. Single-cell analysis tools for drug discovery and development. *Nat. Rev. Drug Discov.* **15**, 204–216 (2016).
5. Bandura, D. R. *et al.* Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. *Anal. Chem.* **81**, 6813–6822 (2009).
6. Bendall, S. C. *et al.* Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. *Science* **332**, 687–695 (2011).
7. Ornatsky, O. *et al.* Highly multiparametric analysis by mass cytometry. *J. Immunol. Methods* **361**, 1–20 (2010).
8. Xu, X. *et al.* Single-Cell Exome Sequencing Reveals Single-Nucleotide Mutation Characteristics of a Kidney Tumor. *Cell* **148**, 886–895 (2012).
9. Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-Wide Detection of Single Nucleotide and Copy Number Variations of a Single Human Cell. *Science* **338**, 1622–1626 (2012).
10. Klein, A. M. *et al.* Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* **161**, 1187–1201 (2015).
11. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
12. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
13. Budnik, B., Levy, E., Harmange, G. & Slavov, N. Mass-spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *bioRxiv* (2017). doi:<http://dx.doi.org/10.1101/102681>
14. Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145 (2015).
15. Saeys, Y., Gassen, S. Van & Lambrecht, B. N. Computational flow cytometry: helping to make sense of high-dimensional immunology data. *Nat. Rev.* **16**, 449–462 (2016).
16. Spitzer, M. H. & Nolan, G. P. Mass Cytometry: Single Cells, Many Features. *Cell* **165**, 780–791 (2016).
17. Arvaniti, E. & Claassen, M. Sensitive detection of rare disease-Associated cell subsets via representation learning. *Nat. Commun.* **8**, 1–10 (2017).