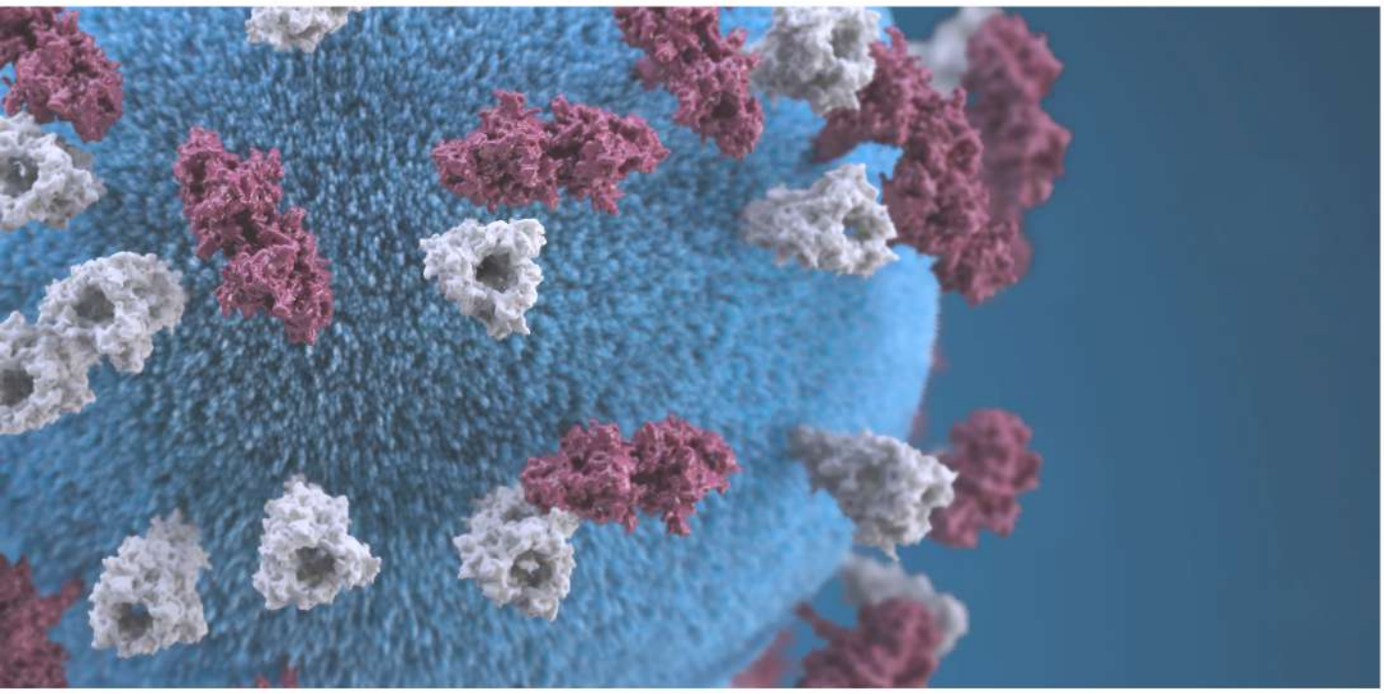
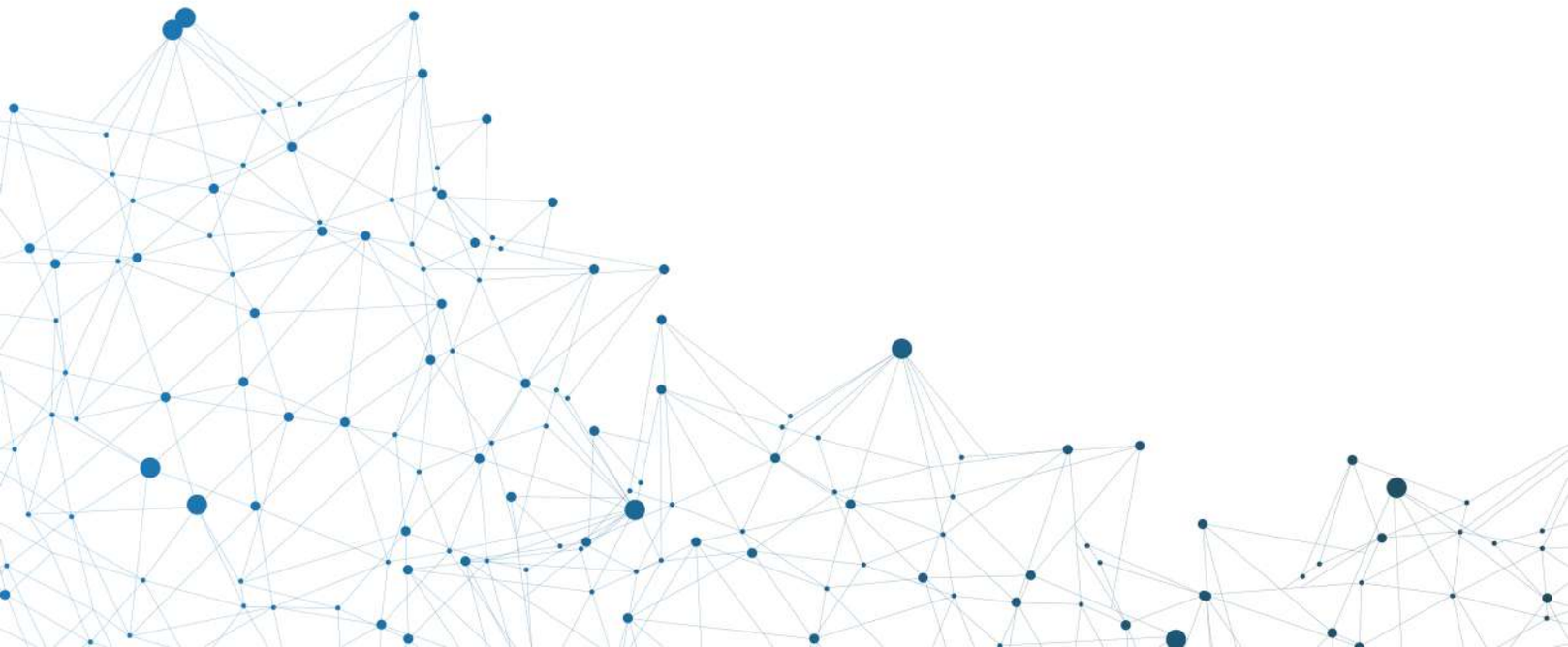
fios GENOMICS



Single-Cell Analysis:

Applications and Resolutions for Research

Contents



1. What is single-cell analysis?

Single-cell sequencing allows the many constituent cell types within a tissue or sample to each be analysed on an individual level. All types of 'omics data can be analysed during single-cell analysis using both high- and low-throughput screening methods. By investigating a sample at such a granular resolution, more subtle fluctuations in expression within a specific cell type can be identified which would otherwise remain hidden when studying a population of mixed cell types.

Multicellular organisms, by nature, have cell to cell variations (Hodzic, 2016). The human body has a wide variety of cell types, where each cell type expresses a complement of genes underlying the distinct function of the cell. Even within a single tissue, individual cells can harbour substantial or subtle genetic differences. Where gene expression measurements are based on a homogenised cell population from a tissue sample, the average expression results can mask cellular heterogeneity. The information from the entire cell population may not account for small but functional changes in individual cells. With the recent advances in single-cell sequencing technologies and bioinformatics approaches, the extent to which cells differ from one another within tissues is becoming clearer. The unprecedented levels of resolution afforded by these technologies may help to provide mechanistic insight into many of the most pertinent biological sciences questions.

Applying single cell approaches can improve our understanding and appreciation that tissues are not a singular entity, but a complex network of distinct single cells that coordinate together to achieve a specific function.

Applications of Single-Cell RNAseq

Identification of gene expression profiles for novel subpopulations

Identification of novel cell subpopulations

Applications of Single-Cell RNAseq

Identification of altered states of cell populations associated with drug response or survival

Comparison of cell type composition of samples across different conditions

2. Research applications

There are multiple methods to extract single-cell data. Multiple scRNA analytical methods have been benchmarked which assists users in making informed decisions regarding the optimal approach for a given experimental setup (Ding et al., 2020).

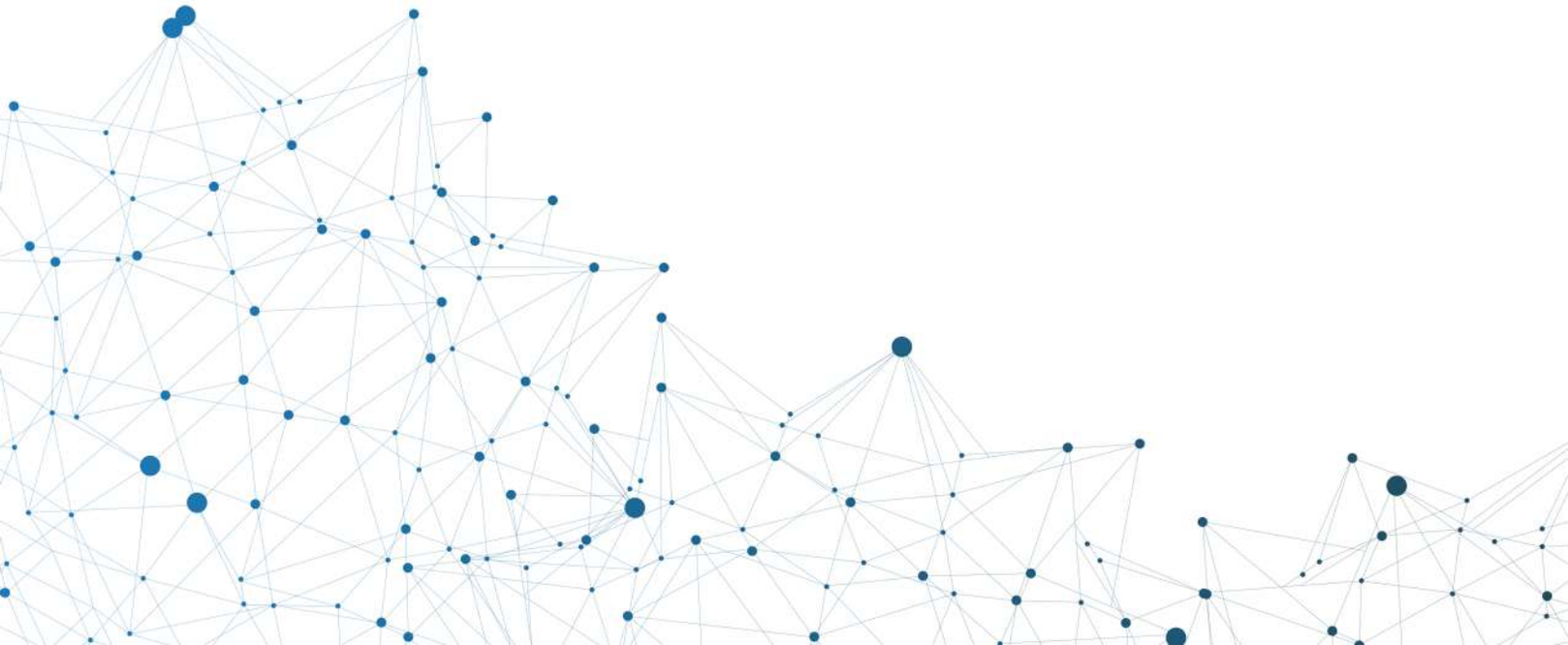
Single-cell genomics lends itself perfectly to the study of changes that occur in a bacterial population, the genetic evolution of cancer, and immune cell subtypes amongst other applications. In bacterial populations, phenotypic differences between bacterial cells can often be overlooked by bulk sequencing techniques. When studying the genetic evolution of cancer, the pseudotime trajectory or divergence of cancer cells can be estimated.

Single-cell epigenomics studies involve the mapping of DNA-methylation across the genome in single cells. This can be used, for example, to better understand mechanisms that regulate gene transcription including environmental influences. Additionally, novel techniques help to study the regulation of transcription by identifying regions of chromatin open and accessible to transcription factors.

Transcriptomics on a single-cell level determines which genes are being expressed in the cell at that particular moment in time and can be used to categorise cell types within the sample and define cell type-specific gene signatures. The transcriptome is often analysed as a proxy for the functional proteome; this is due to the lower costs involved for data generation and the considerable challenges in accurately quantifying cell protein levels as well as post-translational modifications. Gene dynamics studies can determine which gene expression changes accompany phenotypic variations in cell behaviour, allowing for the identity of cells to be established in an unbiased way. Besides cell typing, transcriptomics can be used to study gene expression dynamics from time-series data or RNA splicing to understand how different transcript isoforms are regulated.

Transcriptomics does not provide any information on post-transcriptional regulation processes, which means that there can be discrepancies between the mRNA transcripts that a cell expresses and the protein composition of cells. However, transcriptomics approaches are equipped to capture dynamic changes in gene expression which can happen over far shorter periods of time than equivalent changes in protein expression or modification. Further, current proteomics protocols typically allow for the measurement of only a pre-selected set of proteins.

Finally, single cell metabolomics can be used in combination with single-cell RNA sequencing and single cell proteomics to provide a full picture of the inner working of a cell. It can be used in pharmacodynamics to assess what the by-products of treatments are in a tissue and how they differ between cells within a tissue. Alternatively, it can give a better understanding of ageing, cancer or drug resistance at a molecular level.

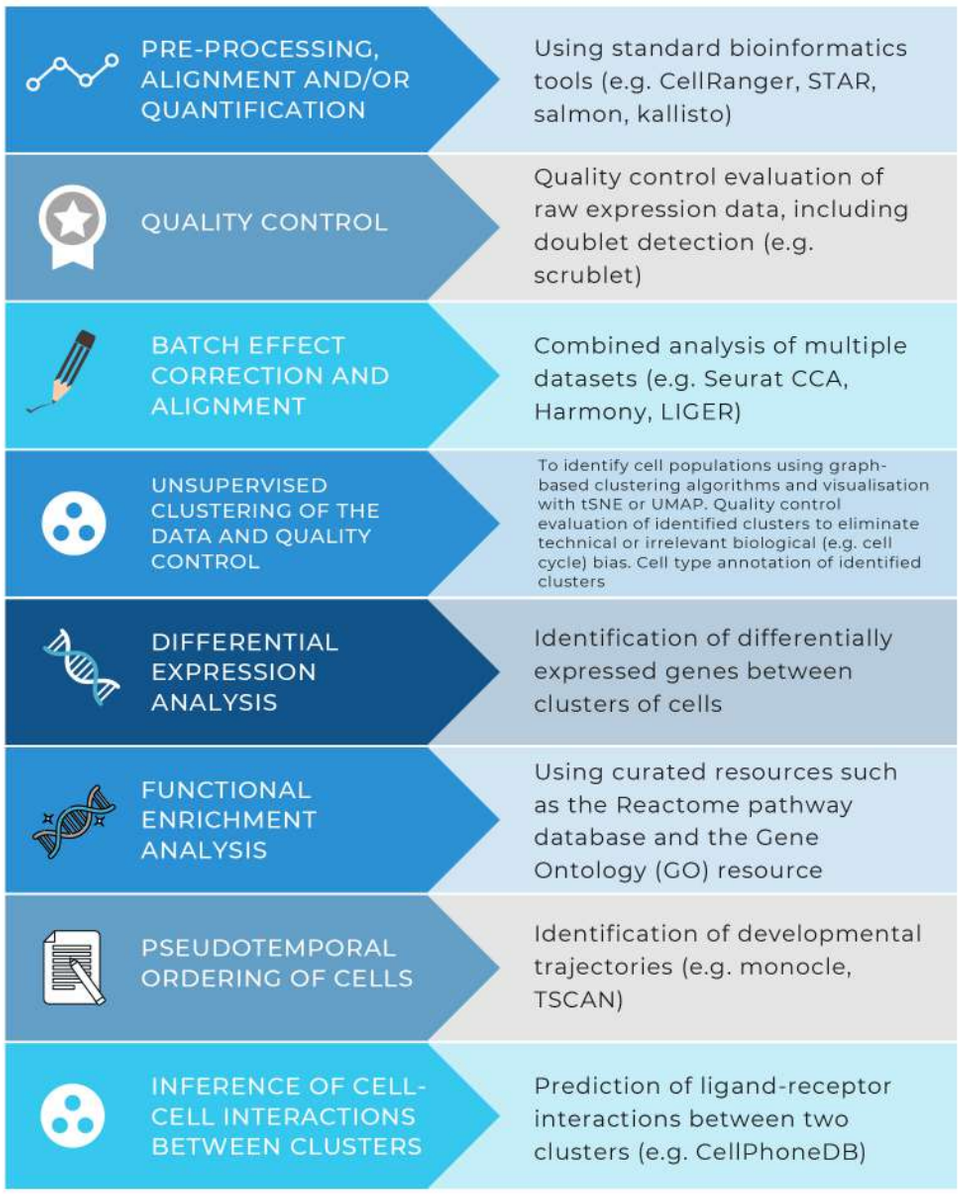


3. Importance in research

Single-cell sequencing marks a shift away from previous methods of data generation as it can bring clarity to complex cell-to-cell variations and interactions. This can include cellular responses involved in immune response, such as antigen-specific T- or B-cells, or in the tumour microenvironment, providing insight into the metabolic and functional state of a cell. This new granularity is important, as it recognises the wealth of data contained within each individual cell rather than grouping large numbers of cells together as a sample.

The main advantage of single-cell analysis is to identify clusters of cells that correspond to distinct cell types or states and to understand the relationship between their gene expression patterns and phenotype. All types of cells can be analysed using a single-cell approach, and even when it is not possible to obtain intact cells from a sample (e.g. post-mortem tissues or highly interconnected neurons), the nucleus can be used instead (single-nuclei analysis). The most common type of data is single-cell RNA-Seq. The workflow for a single-cell RNA-Seq data analysis is similar to bulk RNA-Seq analysis. However, in single-cell data, each cell is its own sample and needs individual library preparation and analysis.

Example of a single-cell analysis pipeline at Fios Genomics



4. Challenges for single-cell research

There are several challenges pertinent to single-cell analysis. These include low sequencing coverage, presence of libraries that represent more than one cell (doublets), and presence of batch effects between data sets that have been generated separately.

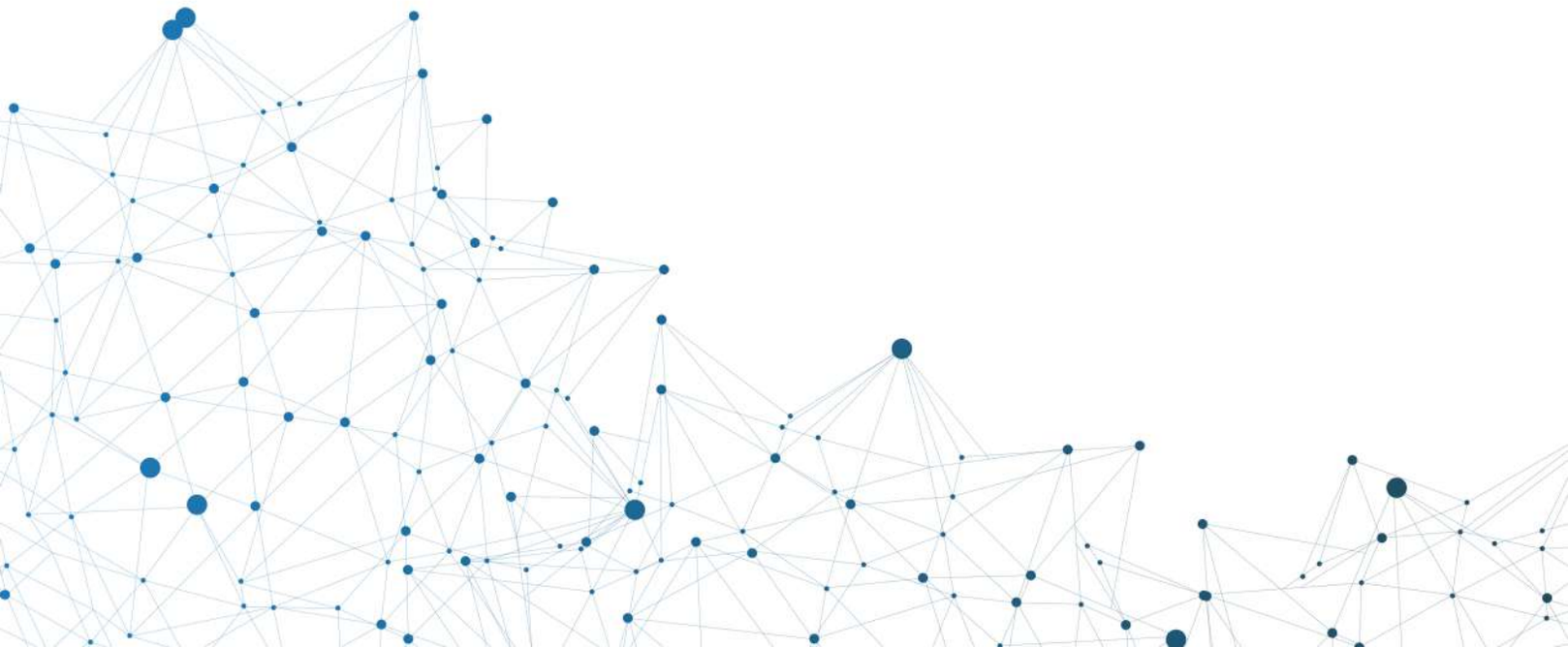
Compared to traditional bulk experiments, the number of reads per sample generated from single-cell experiments is considerably lower. This can cause issues in the measured gene expression as for many genes, especially those with low expression, no corresponding reads are captured, known as drop-outs. Thus, a high proportion of genes in single-cell RNA-sequencing data appear not to be expressed. Additionally, the stochasticity of transcript capturing results in high variation in the measured gene expression between two cells, even those of the same type. This high sparsity and overdispersion of single-cell RNA-seq data need to be addressed with appropriate statistical models. Filtering of low-quality cells (with significantly lower depth libraries) and use of appropriate normalisation methods are crucial pre-processing steps for obtaining accurate results. This quality control step should also address the effects of irrelevant biological variation, such as differences in proportions of cells at different cell cycle stages.

Care needs to be taken at both the experimental design and the analysis stages as multiple cells could be inadvertently captured in a single barcoded bead (known as doublets), giving rise to libraries which represent more than one cell, possibly of different cell types. Doublets can have implications in clustering and differential gene expression analysis. Since the percentage of doublets increases with an increased number of cells processed, this has to be carefully considered during experimental design. However, as doublets exist in any data set, especially from bead-based methods, these need to be filtered during quality control evaluation of the sequencing data.

An issue with single-cell research is the varying levels of resolution that are generated depending on the platform chosen. Integration of multiple single-cell datasets that have been processed independently can cause challenges, especially when batch effects or other technical artefacts are apparent. This is a particular issue when there is a need to compare datasets. Technical artefacts, cell quality, and batch effects all need to be accounted for at the analysis stage as these can mask biological variation in the data. Furthermore, integration of different sources of single-cell data such as proteomic, genomic, and metabolomic data also requires appropriate methods that consider technical artefacts but also dependencies between different measurements (for example, protein levels are not independent of gene expression).

There is a challenge in how best to integrate spatial information from cells to find cell types or functions. By isolating single cells, this removes them from their native environment and the contextual information about the cell's spatial environment and dynamic position can be lost.

With the rapid improvement and the decreasing costs of experimental protocols, single-cell analysis is becoming a routine method for studying cell biology. This leads to an ever-increasing amount of reference single cell data sets and tissue atlases to explore and map generated data against.



5. Resolutions

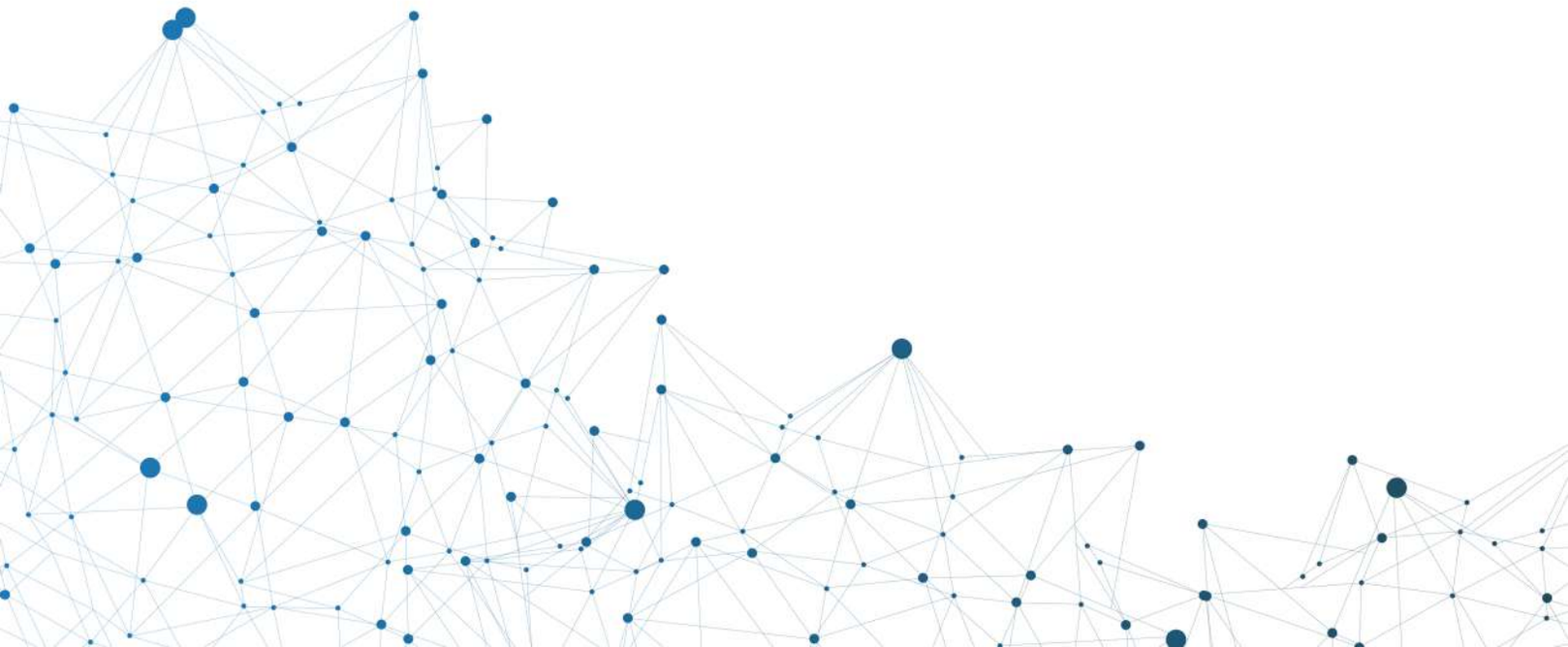
As with many challenges for emerging research, there are methods being developed to resolve or mitigate them. The first opportunity to minimise single-cell analysis issues is at the experimental design stage. Choosing a well-suited method for your research question will allow for a balance between throughput and transcript coverage. Additionally, this can reduce the number of low-quality cells and doublets as well as the impact of batch effects, allowing for better integration of multiple data sets.

The second opportunity to minimise single-cell analysis issues is at the quality control stage. This stage is crucial for single-cell analysis, as this can monitor gene expression distribution to eliminate cells with abnormally high levels of expression as well as highlighting issues surrounding the cell cycle stage which could affect clustering. Similar cells cluster together due to their gene expression and subtypes may cluster separately within the larger clusters, which allows for uncovering of previously unappreciated heterogeneity within cell populations. Further, batch effects or technical bias in the data can be accounted for in order to ensure that these do not impair downstream analyses or biological interpretation of the results.

6. Future of single-cell research

Single-cell analysis allows for information to be drawn from individual cells rather than from the combination of cells of diverse composition within a cell population. As each cell has a distinct lineage, viewing the variations between individual cells of a tissue gives greater granularity into how cells react or change in response to environmental factors.

The use of single-cell analysis will help to better understand the nature and complexity of different diseases, with the potential to unlock more effective therapeutics for a wide range of disease indications. The challenges for single-cell analysis must not be overlooked, particularly the true heterogeneity in cell populations which can now be fully detected, assessed, and appreciated. The introduction and maturing of new computational tools will give greater flexibility to future research. Additionally, advanced tools encompassing both high- and low-throughput methods continue to be developed, enhancing integration of different 'omic datasets.



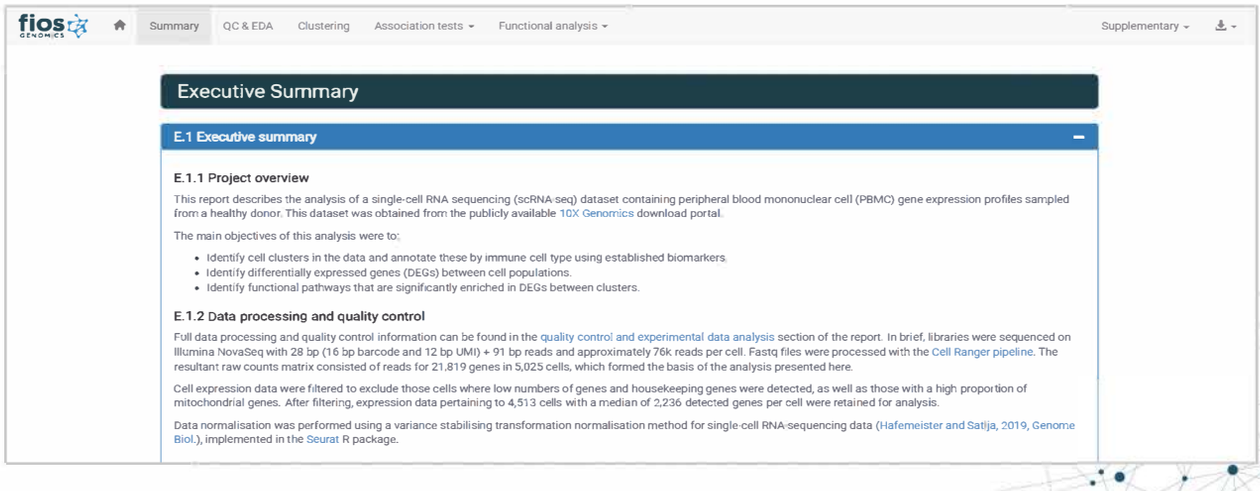
7. Fios Genomics' expertise

Fios Genomics has experience with single-cell analysis related to different therapeutic areas and disease indications. We have expertise in the analysis of complex single-cell RNA-Seq using standard workflows to address the inherent statistical challenges such as low library size, drop-outs, and batch effects.

Every time our clients work with Fios, they benefit from:

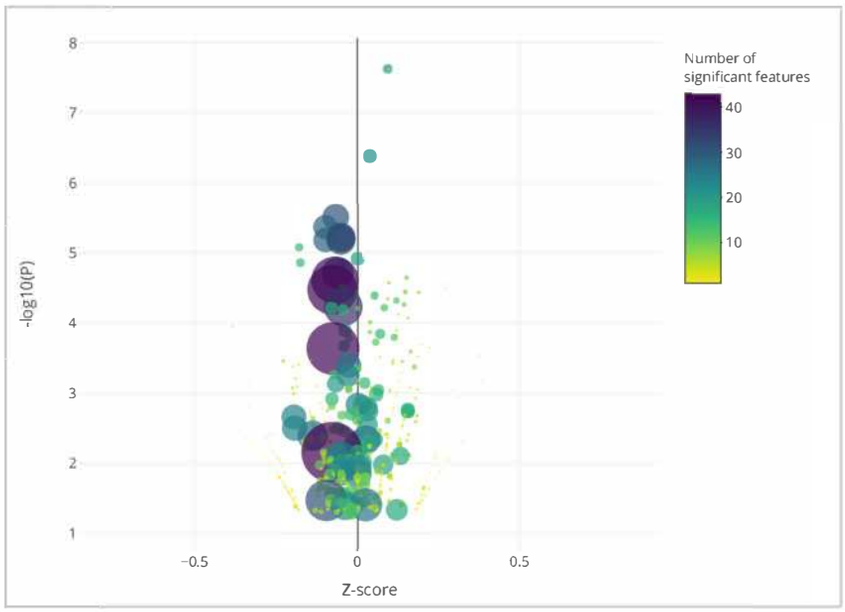
A DEDICATED ANALYST	Backed by an experienced team to curate all data, identify the most appropriate statistical approach to take, and provide biological interpretation of results
AN INTERACTIVE DATA ANALYSIS REPORT	Internally peer-reviewed, including all analysis, methods and outcome
POST-REPORT FOLLOW UPS	Upon receipt of our data analysis report, we arrange a teleconference so that our lead analyst can talk through the results

Access our data analysis sample reports

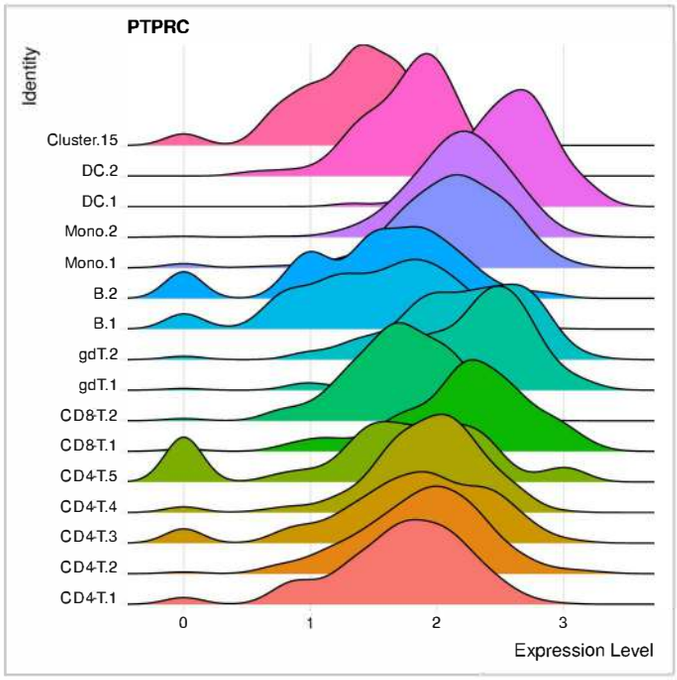


Excerpt from Fios Genomics' data analysis report showcasing single-cell analytical capabilities

Fios Genomics offers a wide range of analysis solutions, including all types of 'omics data. For more information about how Fios can help with your 'omics data analysis needs, visit our **services page**.



Bubble plot from Fios Genomics' data analysis report



Expression of PTPRC (CD45) from Fios Genomics' data analysis report

References

1. Andor et al. (2020) Joint single cell DNA-seq and RNA-seq of gastric cancer cell lines reveals rules of in vitro evolution. NAR Genomics and Bioinformatics, 2(2): lqaa016. [https://academic.oup.com/nargab/article/2/2/lqaa016/5805307]

2. Ding et al. (2020) Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nature Biotechnology, (38): pp. 737–746. [https://www.nature.com/articles/s41587-020-0465-8]

3. Fios Genomics (2020) Single-Cell Example Report

4. Fios Genomics (2018) Single-Cell Analysis [https://www.fiosgenomics.com/single-cell-analysis/]

5. Hodzic (2016) Single-cell analysis: Advances and future perspectives. Bosnian Journal of Basic Medical Sciences, 16(4): pp. 313-314. [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5136769/]

6. Lähnemann et al. (2020) Eleven grand challenges in single-cell data science. Genome Biology, 21: 31. [https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-1926-6]

7. Macosko et al. (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell, 161(5): pp. 1202-1214. [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4481139/]

8. Micheli et al. (2020) A reference single-cell transcriptomic atlas of human skeletal muscle tissue reveals bifurcated muscle stem cell populations. Skeletal Muscle. 10: 19. [https://skeletalmusclejournal.biomedcentral.com/articles/10.1186/s13395-020-00236-3]

9. Nguyen et al. (2018) Experimental Considerations for Single-Cell RNA Sequencing Approaches. Frontiers in Cell and Developmental Biology. 6: 108. [https://www.frontiersin.org/articles/10.3389/fcell.2018.00108/full]

10. Picelli et al. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nature Methods, 10(11): pp. 1096-1098. [https://pubmed.ncbi.nlm.nih.gov/24056875/]

11. Sierant and Choi (2018) Single-Cell Sequencing in Cancer: Recent Applications to Immunogenomics and Multi-omics Tools. Genomics & Informatics, 16(4): e17. [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6440661/]

12. Yuan et al. (2017) Challenges and emerging directions in single-cell analysis. Genome Biology, 18: 84. [https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1218-y]