Cells are one of the fundamental units of life. They show an immense complexity and diversity and their identity and function is determined by environmental stimuli, the physical environment, the cell cycle or neighbouring cells. Only recently its possible to investigate the transcriptome of a single cell. Single-cell RNA sequencing (scRNA-seq) was first published by \citet{tang2009mrna}. This allows to address new biological questions like the identification of rare cell populations, measure the frequency of cell types in tissues, characterising differences in similar cell-types, investigating the heterogeneity in cell states or cell lineages.

A typical scRNAseq workflow consists of the isolation of single cells, extraction of RNA, cDNA library preparation, amplification and sequencing of the libraries. A wide variety of scRNA-seq protocols exists, differing in throughput, full transcript or 3'sequencing, costs and automatization.

Small-scale protocols are standard PCR plate based methods or methods in which cell-isolation and library preparation is combined into one protocol. A typical small scale method is the plate-based PCR SMARTseq2. Libraries are full transcript sequenced using a standard Illumina sequencing approach. Typically hundreds of cells are processed and ERCC is used for normalisation.

On the other side of the spectrum Drop-seq is a droplet-based method using microfluidic cell sorting (e.g. FACS), allowing for the processing of thousands of cells. Sequencing is 3'end and uses UMI \citep{wagner2016revealing}. One of the highest throughputs is achieved by 10xChromium, allowing for the sequencing of tens of thousands of cells. The method is again droplet-based, sequencing is 3' end and UMI based.

Differences between bulk experiments are the lower sequencing depth (100000 - 5 million reads per cell), higher variability and more outliers.

scRNA-seq data suffers from technical noise, multiple cells in a library, batch effects and low capture efficiency.

Batch effects occur when different biological conditions are processed in different batches, making the deconvolution between technical noise and biological effect impossible. Whenever possible this should be avoided by an appropriate experimental design that allows for the statistical deconvolution between unwanted and wanted variation. In scRNAseq the single experimental unit is the cell, making this approach not always possible. Different cells in an experiment may need different sample processing or their biological differences affects the downstream analyses.

Doublets occur when multiple cells processed in the same library and is mainly a problem for droplet-based methods. Starting amounts of the library preparation can be as low as ten picogrammes of total RNA. Two main issues are arising to the low starting amount. The massive amplification and low capture efficiency. Low and moderate expressed genes are not captured during the reverse transcription, which leads to dropouts of genes and a zero-inflated gene expression. scRNA-seq data has an excess of zero counts and can be split into systematic, semi-systematic and stochastic zeros \citet{lun2016pooling}. Systematic zeros are silent across all cells. Stochastic zeros are zero counts that were obtained due to sampling. It affects genes with a count distribution near zero.