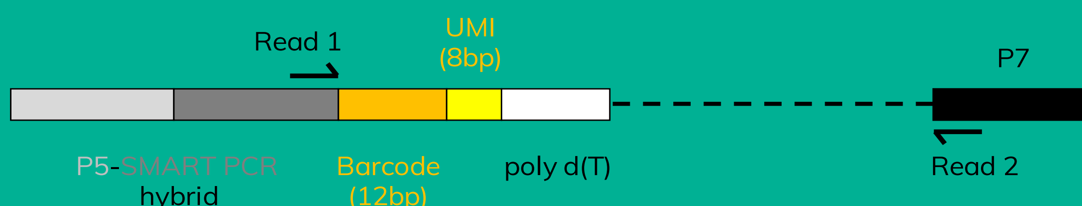
NGS guide for scRNA-Seq libraries (user-sourced reagents)

Guidelines for Next Generation Sequencing with Illumina

Single cell RNA-Seq enables the analysis of thousands of single cells in order to identify and monitor cellular expression patterns. This is a summary of NGS guidelines to help Nadia users through the sequencing process.

Molecular architecture of DNA insert produced

ScRNA-Seq on Nadia culminates in libraries that are compatible with Illumina's single indexed Paired End sequencing. The molecular architecture of a DNA insert following Nextera XT tagmentation is illustrated below:



Number of STAMPs to be analysed

This will be determined by the complexity of the tissue and the objectives of the biological study. The more cells that are sequenced, the more distinguished clusters of cells will be. However, budgetary constraints may dictate a compromise between the number of STAMPs and sequencing depth per STAMP. A lower sequencing depth per STAMP may result in the detection of fewer transcripts and genes per transcriptome. A compromise between quantity of STAMPs profiled and sequencing depth per STAMP may be required. We recommend that the user is guided by prevailing studies in the scientific field of interest.

Multiplexing and recommended read length

The N701 – N712 oligos from the Nextera XT Index Kit can be used to create libraries suitable for multiplexing during sequencing. Note: Please do not use the S5XX oligos and do not use dual indexing during library preparation/sequencing.

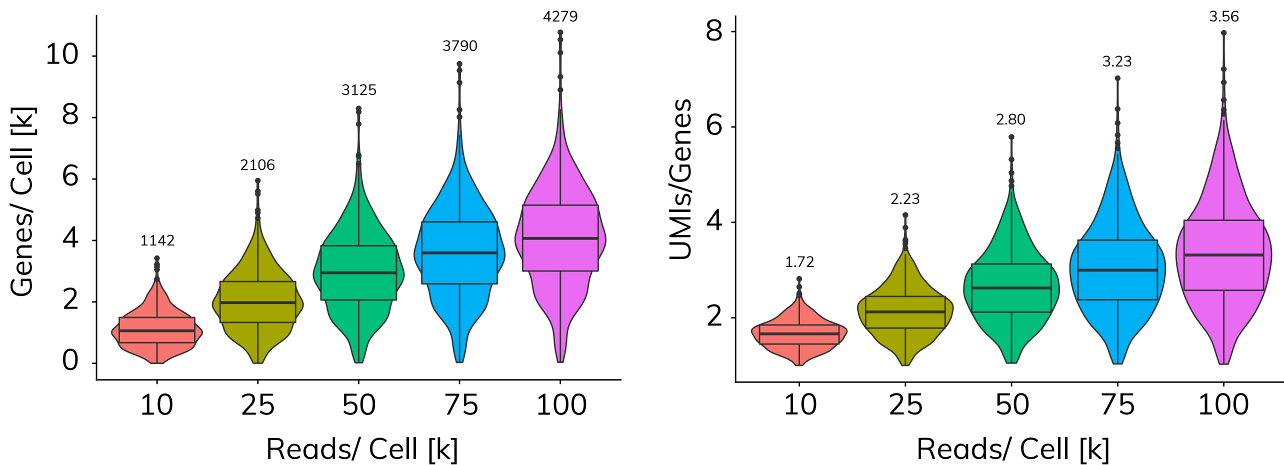
Recommended read lengths with 2x75 bp PE chemistry

Read	Length
Read 1	26 bp
I7 index read	8 bp
Read 2	116 bp

For NGS sequencing please use primer as specified below:
Read 1: "Read1CustomSeqB" (5' GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC)
Read 2: Standard Nextera P7 Sequencing Primer
The inclusion of 10 % PhiX can improve overall read qualities whilst maintaining a read length of 26 bp for Read 1. The use of shorter read lengths may result in Illumina not supporting the diagnosis of problems during sequencing.

Sequencing depth

The required sequencing depth for a particular experiment will depend upon sample type or the experimental questions that need addressing. RNA-rich samples such as cell lines may not reach full sequencing saturation at lower sequencing depths such as 50k reads per cell, however this might be enough to distinguish sub-populations during clustering. If you are interested in identifying smaller subpopulation of cells or as many genes as possible (or lowly expressed marker genes) higher NGS-depth might be required. Furthermore, non-nuclear RNA should be taken into account such as RNA from chloroplasts or mitochondria as well as rRNA. If higher amounts of non-mRNA are expected such as in plants or cardiomyocytes it is advised to increase the sequencing-depth to account for reads going towards those RNAs.
Finding the optimal sequencing depth for one's experiment can be an iterative process and may require multiple rounds of optimization. Below we have given an indication of sequencing saturation by plotting down sampled sequencing depths as average number of reads per cell. The recommended sequencing depth on Illumina machines is 25k read pairs/cell.



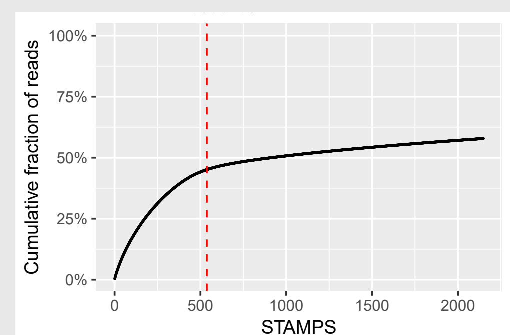
Computational Tools and Data Analysis

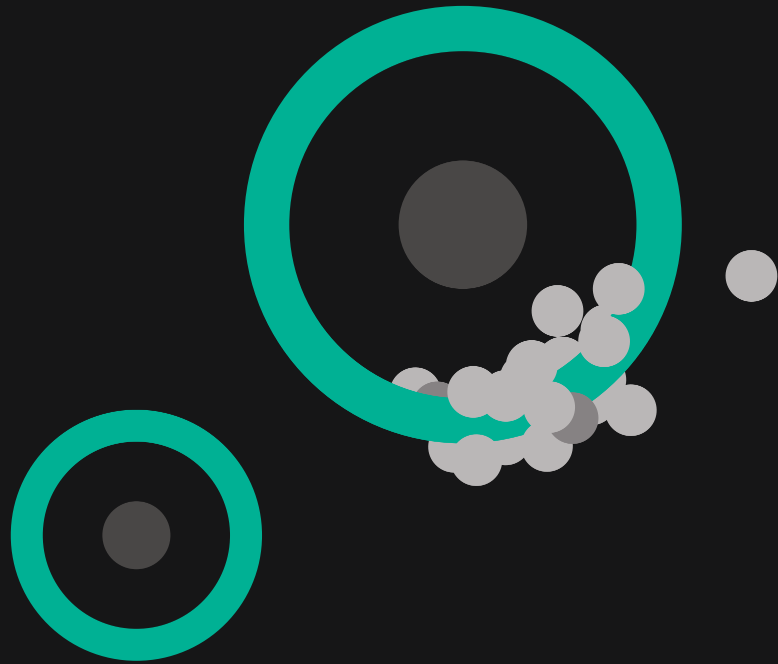
Computational tools

For data analysis, dropSeqPipe a computational pipeline originally developed by the McCarroll Lab at Harvard can be used. Additional guidance on usage of this tool can be obtained here.
Alternatively, data analysis can be handled through Dolomite Bio's Bioinformatics Service. Contact your local Dolomite Bio sales representative to find out more.

Data Analysis: Use of the knee-plot to identify high quality STAMPs

The knee-plot allows the user to identify a subset of STAMPs that will be used for analysis. Use of the knee-plot is highly recommended during data analysis and is instrumental in producing good data.
The knee-plot (shown on the right) arranges barcodes in descending number of sequenced reads associated with each individual barcode. It allows the user to identify a cut-off threshold (dashed red line) and shortlist data-rich barcodes that represent high-quality STAMPs. Beyond the point along the plot at which the cut-off is placed, barcodes are associated with decreasing numbers of reads that approach zero.





dolomite-bio.com

Get in touch

Email
info@dolomite-bio.com

Join us on

UK Head Office
(Europe, S.E. Asia, Australasia, China, Middle East, Africa)
t: +44 (0)1763 252 102

North America Office
t: +1 (617) 848 1211

Japan Office
t: +81 45 263 8211

Asia Regional Office
t: +84 93 55 60 80