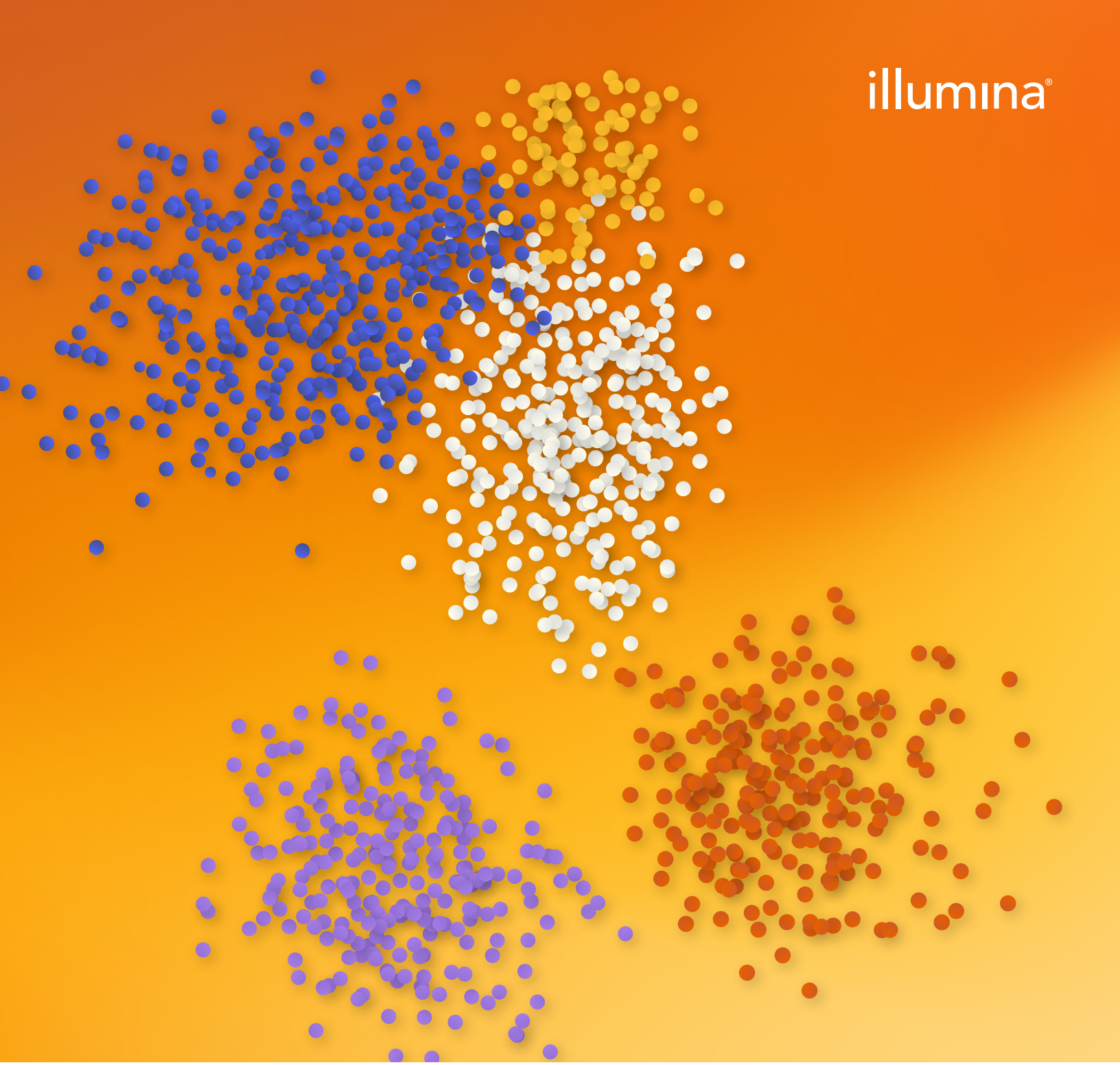


Illumina Single-Cell Sequencing Workflows: Critical Steps and Considerations

Explore every step of single-cell sequencing workflows and learn valuable insights to ensure experimental success.

Table of contents

Single-cell sequencing workflow: a critical guide for beginners

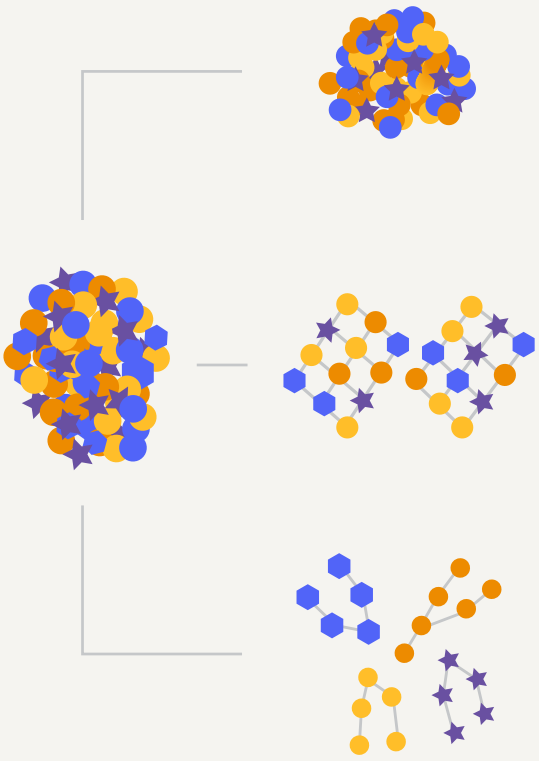
Living tissues are comprised of an extensive variety of cell types, each with a distinct lineage and unique function that contribute to tissue and organ biology. Traditionally, tissue sections are isolated so nucleotides and proteins from the cells of interest can be extracted from a specific anatomical region. The analysis of RNA, DNA, and proteins from these samples can be a powerful tool for examining how tissues or cell populations respond to various stimuli or stressors.

However, the variation between individual cells can be immense, even when examining the same cellular subpopulation. This is especially true of the transcriptome, which is a more reactive and dynamic -ome compared to the relative stability of the genome and epigenome. Furthermore, each cell's lineage, developmental stage, and differentiation pattern can profoundly affect how they respond to other cells and external stimuli within their microenvironment. Together, these attributes ultimately define the biology of the entire organism. Due to these complexities, gaining insights into cellular function through bulk-cell analysis of tissues or cells presents significant challenges, highlighting the need to characterize individual cells.[1]

Quantitative PCR (qPCR) is a valuable and popular molecular method used to analyze single cells, given its wide dynamic range, familiar workflow, and easily accessible instrumentation.[2] However, qPCR interrogates a limited number of targets with known sequences. Furthermore, the PCR workflow is too cumbersome for high-throughput experiments and unsuited to a multiomic approach seeking to combine insights from the transcriptome, genome, epigenome, and proteome.

Conversely, the high accuracy, throughput, and specificity of next-generation sequencing (NGS) technology makes it ideal for bulk, spatial, and single-cell sequencing. Combined with improvements in cost-effectiveness over the last decade, NGS has never been more accessible. No matter the research goal, Illumina next-generation sequencing technology enables scientists and researchers to dive deep into each layer of the central dogma at multiple resolution levels.

Figure 1: Harness the power of Illumina NGS with any analytic approach



Bulk analysis is the preferred approach for understanding the big picture. Bulk analysis excels at providing insights into the entire tissue and can be used as an untargeted approach for new discoveries. When a more focused approach is required, cells from bulk samples can be further divided by size, internal complexity, or known biomarkers with flow cytometry and magnetic beads. Bulk analysis also benefits from a low cost-per-sample, making it significantly more scalable than single-cell or spatial analysis approaches.

Spatial analysis is ideal when the native tissue architecture and microenvironment need to be maintained for the experiment. Spatial analysis captures genomic and transcriptomic information from specific locations within the tissue, providing a realistic look at processes occurring within an organism.[3] Cell-to-cell communication can have significant impacts on cellular expression profiles, and spatial analysis removes single-cell isolation protocols that can alter cell behavior.[4]

Single-cell analysis provides the deepest insights into nuanced distinctions between cells in the same sample. The power of this approach can be further enhanced with a multiomic approach to single-cell sequencing. For example, cellular indexing of transcriptomes and epitopes by sequencing (CITE-Seq) examines RNA transcripts in tandem with surface protein expression at the single-cell resolution to combine the insights of transcriptomics with proteomics. Examining complex organs and tissues at single-cell resolution is critical to advancing our understanding of many diseases and systems.

Why Single-Cell Multiomics?

63%

year-over-year increase in the number of multiomics publications since 2012.[5]

61%

of NGS researchers are incorporating single-cell RNA-Seq into their experiments.[6]

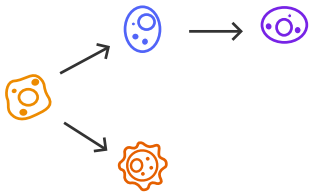
~2.6M

cells sequenced per experiment, an unprecedented throughput volume for the characterization of rare cells or unique microenvironments.[7*]

*Assumes 10K reads per cell with 10x Single Cell Gene Expression Flex on NovaSeqX with 25B flow cell.

Understand the development of complex organs, and their diseases

Model and predict behavior of cells during development



Detect how different cell populations react to external factors

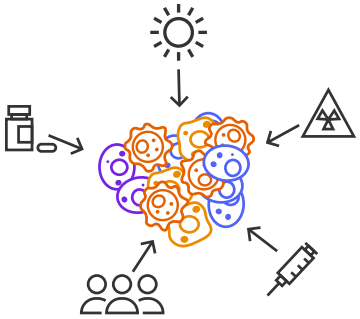


Figure 2: The power of single-cell multiomics

Learn more about taking a multiomics approach to single-cell sequencing.

A well-planned and executed experiment is important to ensure accurate data and draw insightful conclusions.[8]

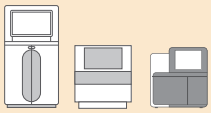
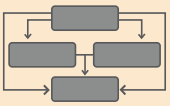
Tissue Dissociation	Generate single-cell libraries	Sequence	Analyze data	Visualize data
				
Cells or nuclei are isolated and suspended	The nucleotides of individual cells are isolated for library preparation	NovaSeq™ X, NovaSeq 6000, NextSeq™ 2000, NextSeq 1000, NextSeq 550	DRAGEN	Illumina Connected Multiomics

Figure 3: The single-cell sequencing workflow.

STEP 1

Tissue preparation

Most single-cell isolation platforms require a viable, monodispersed sample prior to downstream analysis or fixation. The type of tissue, species, and age of the animal can all influence isolating live single cells from tissues. This chapter presents some of the key considerations in the preparation of high-quality single-cell suspensions.

Fixation

Sample fixation will be required when it is not convenient or possible to process samples immediately. When samples are collected at different locations and time points, fixation preserves cells until they can be processed in parallel, which enables collaborations and mitigates technical artifacts that can arise from sample prep and sequencing between samples. Fixation can also increase the power of experiments, facilitating technical or biological replicates of an experiment. In addition, many biologically interesting fixed samples exist in tissue and cell banks for use in single-cell multiomic studies. For these reasons, it's crucial that scientists check the compatibility of single-cell preps with the fixation methodologies.

Dissociation

The process of single-cell preparation is a significant source of variability in any single-cell study.[9] Clumped cells or a high cell death rate due to the dissociation method a researcher uses can confound data and lead to misinterpretation. Non-adherent cells, such as peripheral blood mononuclear cells, are often more amenable to single-cell processing than adherent cells or cells isolated from tissue. Tissues can vary significantly in extracellular matrix (ECM) composition and cellularity. Thus, dissociation protocols should be optimized for a specific tissue of interest in order to isolate the cell population of interest and maximize the efficiency of the sequencing run.[10] Conventional protocols for tissue dissociation include mechanical dissection, enzymatic ECM breakdown, and combinatorial protocols as discussed in **Table 1** below.

Another important consideration at this stage is whether cell or nuclei isolation is a better fit for your research goals. Using single-cell sequencing with whole cells will yield a higher number of genes compared to single-nuclei sequencing and will also capture a higher proportion of exonic reads.[11] Whole-cell sequencing is also a better fit when the molecule of interest is found in the cytoplasm or cell membrane. However, single-nuclei sequencing is better at preserving relative cell proportions, and is therefore the preferred approach for accurately determining the frequency of different cell types within a population.[12] Single-nuclei sequencing can also be a valuable solution for samples where healthy, intact cells are difficult to isolate (such as brain tissue).

Table 1: Single-cell dissociation can be approached by mechanical, enzymatic, or combinatorial means.

Mechanical
Tissue is mechanically sheared and disrupted through cutting, dicing, and pipetting to isolate single cells. Learn more about mechanical dissociation.

Enzymatic
Tissue is incubated with various enzymes such as collagenase, trypsin, dispase, and elastase, to cleave protein bonds and extracellular matrix. Learn more about enzymatic dissociation

Combinatorial
Tissue is subjected to both mechanical and enzymatic dissociation, sequentially or simultaneously. This method may also include automated systems for more extensive tissue dissociation. See examples and learn more about automated dissociation systems.

Watch Illumina's webinar on preparing samples for single-cell sequencing experiments.

Enrichment

Enrichment of specific cell populations, or removal of unwanted cell populations, including dead cells, is an optional but often critical step in single-cell preparation. This is especially true when working with rare cells. Various methods are available that require optimization for each specific tissue type (**Table 2**). Manual isolation of cells based on size, shape, and density can be achieved through differential or density gradient centrifugation and filtration. For example, mononuclear cells can be isolated from peripheral blood or bone marrow cells by centrifugation through various density gradient media.[13] Various fluorescent dyes are available to label and separate live cells from dead or apoptotic cells using flow cytometry. For enrichment of cell subpopulations or rare cell types, antibody labeling for positive/negative selection can be combined with flow cytometry or magnetic bead-based isolation.

Ultimately, researchers should choose a method driven by a combination of factors, including sample or tissue type, antibody availability, and experimental design.

Table 2: Enrichment methods

Method	Description	Example Protocol/Provider
Centrifugation	Cell populations of interest are enriched based on size, shape or density by centrifugation through a density gradient medium.	Millipore Sigma
Bead-based enrichment	Cell populations of interest are enriched using magnetic bead-conjugated antibodies via positive or negative selection.	Miltenyi Biotec LevitasBio Akadeum
Flow cytometry	Cell populations of interest are enriched by positive or negative selection with fluorophores or fluorochrome-conjugated antibodies.	Beckman Coulter BD BioLegend
Microfluidic cell sorting	Cell populations of interest are enriched using low-pressure microfluidics based on positive and negative selection with fluorophores or fluorochrome-conjugated antibodies.	NanoCellect Celselect Slides
Centrifuge-free cell washing	Centrifuge-free Laminar Wash™ technology can be used to enrich cell populations. Unlike centrifuge-based enrichment methods, centrifuge-free cell washing can reduce cell clumping and helps to maintain cell integrity and viability at a higher percentage.	Curiox
Microdissection	Lasers can be used to isolate single cells from tissue sections. This method can be valuable when you need to preserve the spatial context of the tissue.	See the protocol published by Suarez-Quian CA, et al. 1999

Reagents for separating live and dead cells

The presence of dead cells can, and usually will, introduce artifacts into experiments. To avoid confounding sequencing data, it is crucial that dead cells are eliminated from your samples. There are several methods a researcher could use to separate live cells from dead cells. Researchers routinely use three classes of reagents (depending on the cell type, experimental design, and antibody panel) to analyze and separate dead cells from live cells when viewed under a microscope. Digitonin can also be added to increase cell membrane permeability and enhance live/dead cell differentiation in low quality samples.

The mechanisms for each reagent class as well as the pros and cons of using each dye are summarized in **Table 3** below.

Table 3: Reagents for separating live and dead cells

Reagent	Mechanism	Pros	Cons
Classic DNA dyes (eg, Propidium iodide (PI) and 7-aminoactinomycin (7-AAD)	These are DNA-binding dyes that do not cross an intact cell membrane. Dead cells with non-intact membranes will absorb the dye and fluoresce when the dye is excited.	Easy to use and inexpensive Require minimal incubation periods	Incompatible with intracellular staining
Amine dyes	Amine dyes bind to amine groups of proteins and do not cross an intact cell membrane.	Compatible with intracellular staining A wide selection of these dyes is available	More expensive than other dyes Labeling must be done in the absence of free protein Incubation with your cells may prolong your protocol
Vital dyes	Vital dyes fluoresce as a result of cleavage in response to interacting with metabolically active cells.	Inexpensive and easy to use	Challenging to use with intracellular staining

Quality control

Single-cell sequencing experiments represent a significant investment of time, money, sample material, and resources. Several simple quality control (QC) measures throughout the tissue preparation process can ensure a high-quality experiment before proceeding with cell isolation, library preparation, and sequencing.

Visual inspection

Visual inspection of the cell suspension under a microscope is valuable, as it enables quick identification of debris, cell doublets, and cell aggregates that can complicate downstream steps (**Figure 4**). Importantly, accurate cell counts are critical to achieve target cell throughputs in subsequent single-cell isolation procedures. Cell counts can be determined manually by combining microscopy with a hemocytometer.

Automated counters can also provide accurate cell counts, capture brightfield images of the cell suspension, and generate histograms for more detailed inspection based on cell characteristics such as size, brightness, and circularity. Examples of commercially available automated cell counters include the Countess® II Automated Cell Counter (Thermo Fisher Scientific), the TC20™ Automated Cell Counter (Bio-Rad), and Cellometer™ Auto T4 Bright Field Cell Counter (Nexcelom Bioscience).

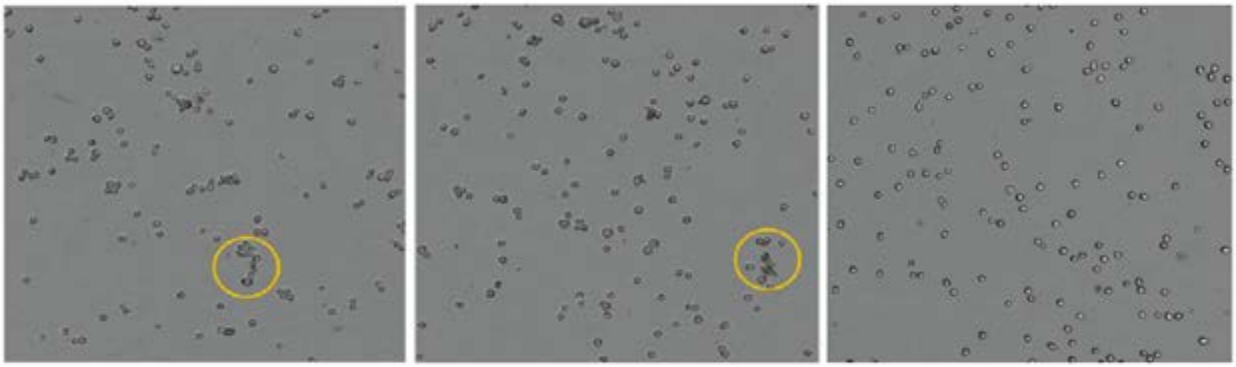


Figure 4: Visual inspection of cell suspensions after tissue dissociation by brightfield microscopy reveals debris, cell doublets, and larger aggregates (yellow circles) present in the samples. A sample with less than 10% doublets is shown on the far right.

Flow cytometry

Flow cytometry is a valuable QC tool, that can be used to simultaneously analyze multiple metrics, including cell size, viability and the presence of doublets or cell aggregates. Additionally, antibody labeling can be included as part of the analysis to evaluate whether cell populations of interest are present.

Cell viability

Dead or damaged cells can release nucleic acids into the cell suspension that remain throughout the subsequent steps possibly impacting results. Cell viability levels above 85% are recommended.

Cell size distribution

Histogram plots can be inspected for presence of multiple peaks indicating cellular fragments (smaller peaks), doublets or aggregates (peak at twice nominal cell size), or large debris (larger peaks).

Cell concentration

The ideal loading concentration for cells depends on the isolation method. Optimal concentration is critical as underloading or overloading can cause issues with single-cell isolation or data quality.

Cell type distribution

Ratios of cell types can be determined using fluorescence-conjugated antibody staining.

Summary

Harnessing the potential of NGS to investigate complex biological systems at the level of individual cells requires that tissues are properly dissociated into monodispersed suspensions of viable cells. A wide selection of methods is available. Specific dissociation protocols should be selected and optimized based on the tissue of interest. Researchers must consider including an enrichment step and key QC metrics to ensure a high yield of single cells while maintaining viability. After a tissue preparation protocol has been optimized, researchers can proceed with confidence to single-cell isolation and library preparation.

STEP 2

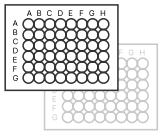
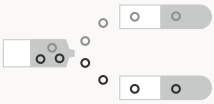
Single-cell isolation and library preparation

Rapid advances in both methods and commercial products available to profile RNA, DNA, and protein in the same cell resulted in single-cell sequencing being named the Method of the Year twice.[14] Ultimately, the method a researcher chooses to capture and isolate single cells will determine library preparation sequencing and downstream analysis. This chapter focuses on available options for single-cell isolation and highlights techniques used for global characterization of isolated cells.

Cell isolation methods and platforms

Cell isolation methods can be distinguished by throughput. Advances in microfluidic technologies have enabled high-throughput single-cell profiling where researchers can examine up to tens of thousands of cells per experiment in a cost-effective manner (**Table 4**).[15] Flow sorting is a commonly used low throughput method that is able to process up to a few thousand cells per experiment.

Table 4: High-throughput single-cell isolation approaches

	High-Throughput Methods			Low-Throughput Methods
				
	Droplet-based[16,17,18,19]	**Microwell-based**[20,21]	**Instrument-free**[22,23,24] (Geared towards combinatorial indexing)	**Flow sorting**
Description	Uses compartmentalization of individual cells in droplets using a microfluidics device followed by lysis and capture of target DNA and RNA	Microwells containing fabricated arrays are used to capture individual cells	Intact cells or nuclei are tagged via multiple rounds of splitting, pooling, and ligation to generate different barcode combinations	Microdroplets containing single cells will be isolated using electric charge
Advantages	High cell throughput Good data quality Unique molecular identifiers (UMIs) and cell barcodes enable cell and gene-specific identification Most widely published Low cost per cell	Low doublet rate Gentle processing for tricky specimens Supports imaging and short-term cell culture Ideal for adherent cells UMIs enable cell- and gene-specific identification	Very high sample and cell throughput Low doublet rate Lower cost approach to profiling a large number of cells No instrument required	Accurate selection of cell type by size, morphology, internal complexity, and protein expression by antibody labeling
Commercial offerings/ example methods	Illumina Single Cell 3' RNA Prep Mission Bio Tapestri Platform 10x Genomics Chromium Controller	BD Rhapsody™ HT Xpress System BD AbSeq CellMicrosystems CellRaft AIR® System Takara ICELL8® cx Single-Cell System	Illumina Single Cell 3' RNA Prep Parse Biosciences Scale Biosciences	Beckman Coulter BD

Library preparation

Library preparation is the next critical step in the single-cell sequencing workflow. At this stage, adapters are added to make sequences compatible with Illumina sequencers and barcodes are attached to allow identificaiton of samples after sequencing. The cell profiling approach and specific sequencing method chosen are important considerations, as various options are available (**Table 5**). The method you choose will largely depend on the experimental question.

Table 5: Library preparation methods

	Method	Description	Commercial Offerings
Genome	**Targeted Panels**	Various predesigned single-cell targeted DNA sequencing panels enable profiling of hematologic malignancies, solid tumors, copy number variation (CNV), and more	MissionBio Tapestri
	Whole genome amplification or primary template-directed ampification	Uniformly amplify genomic DNA from extremely low input	BioSkryb ResolveDNA Qiagen REPLI-G
Epigenome	**ATAC-Seq**	Assay for transposase-accessible chromatin using sequencing (ATAC-Seq) assesses chromatin accessibility genome-wide by using a transposase to insert sequencing adapters into regions of open chromatin	Scale Single Cell ATAC Pre-Indexing Kit 10x Genomics Chromium Single Cell ATAC
	Hi-C	Hi-C combines chromosome conformation capture (3C) with NGS to enable unbiased identification of chromatin interactions across the epigenome	Hi-C methods as described by Belton et al 2012.
	Methylation Sequencing	Single-cell methylation analysis performs bisulfite conversion on fixed nuclei which are subsequently sequenced to produce a single-cell methylation matrix	Scale Single-Cell Methylation
Transcriptome	**Full-length RNA-Seq**	Switching mechanism at 5’ end of RNA template (SMART) technology enables amplification of full-length cDNA	Takara SMARTer ®
	mRNA	Capture of mRNA enables sequencing of the coding transcriptome	Illumina Single Cell 3' RNA Prep 10x Genomics (3’ WTA) 10x Genomics (5' WTA) 10x Chromium Single Cell Gene Expression Flex Scale Single Cell RNA Sequencing Kit
	Targeted panels	Various predesigned single-cell targeted RNA sequencing panels enable IR, T-cell, breast cancer profiling, and more	BD Rhapsody Single-Cell Analysis
	IR-Seq	Immune repertoire sequencing (IR-Seq) is a targeted sequencing method used to quantify the composition of B- or T-cell antigen receptor repertoires	10x Chromium Single Cell Immune Profiling
Proteome	**AbSeq**	DNA-tagged antibodies enable protein profiling by NGS	BD AbSeq Biolegend TotalSeq™
Multiome	**CITE-Seq**	Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE-Seq) uses oligonucleotide-labeled antibodies to convert protein detection into a quantitative assay by NGS	BioLegend TotalSeq Reagents for Single-Cell Protein and RNA Detection 10x Genomics Single Cell Profiling
	Multiome ATAC + GE	Multiome ATAC and gene expression (GE) enables simultaneous detection of mRNA and chromatin accessibility from the same cell	10x Genomics Single Cell Multiome ATAC + Gene Expression

Quality control (QC) of prepared libraries

To maximize sequencing data quality and output, it is important to accurately measure both the quality and quantity of prepared libraries. We recommend using qPCR to quantitively measure your prepared libraries. Researchers can also depend on Illumina benchtop sequencers, such as the MiSeq™ i100 System, for DNA library QC and measurement or rebalancing of pooled samples before sequencing (**Figure 5**). Using a simple, streamlined workflow on the MiSeq i100 System, libraries can be sequenced to shallow depths and analyzed for quality.[25] Agilent TapeStation, BioAnalyzer, or Fragment Analyzer are additional solutions that provide qualitative nucleic acid measurements like library size and confirm that you have a single product following library preparation for Illumina sequencing workflows.[26]

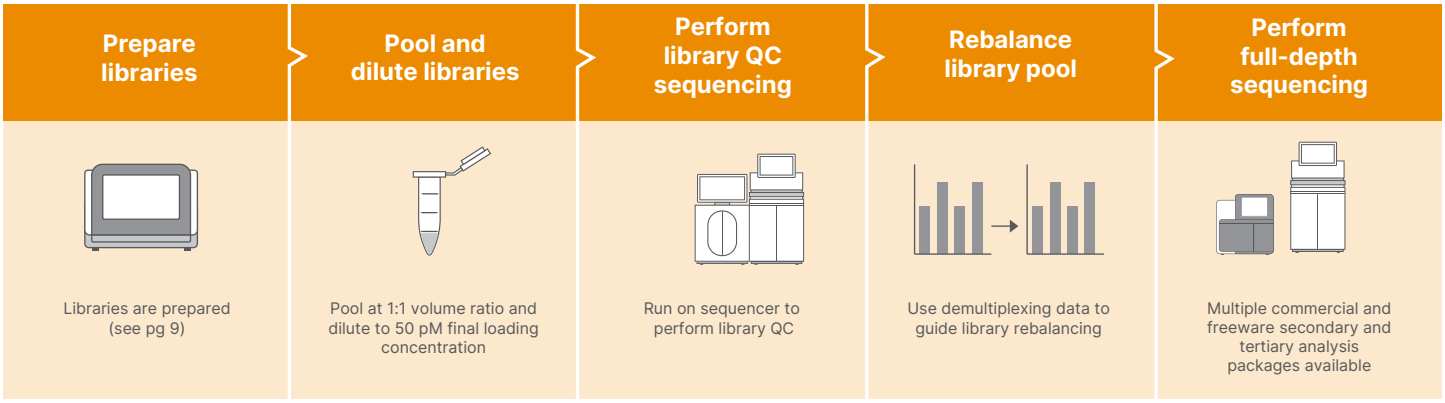


Figure 5: Single-cell library QC with the iSeq 100 System. The iSeq System uses a simple workflow to perform quality control prior to high-depth sequencing.

Quality control in library prep

Before sequencing, it's essential to assess the quality of your cDNA libraries.

Only high-quality libraries should be subject to sequencing to ensure generation of reliable, high-quality data. Quality assessment also ensures the appropriate amount of library cDNA can be loaded for sequencing, which depends on the sequencing platform and flow cell. Three measures, in particular, are crucial for determining cDNA library quality:

Fragment size distribution—The average fragment size for cDNA should be >500 base pairs, although average size may vary depending on cell type and experimental condition.

You can use a Bioanalyzer or TapeStation to check this distribution.

Concentration—cDNA concentration in the prepared library can be quantified using Qubit or similar fluorometric methods.

Purity—Before sequencing, it is essential to ensure minimal adapter dimers or primer artifacts are present in your cDNA library. Adapter dimers contain full-length adapter sequences that are able to bind and cluster on the sequencing flow cell and generate unwanted sequencing data. In contrast, primer dimers do not contain complete adapter sequences, and are not able to bind or cluster on the flow cell, so are not sequenced. During quality checks with the BioAnalyzer or Fragment Analyzer, or on an agarose gel, an unexpected small peak at 120-170 bp indicates the presence of adapter dimers.

Summary

A critical step in the single-cell sequencing workflow is the isolation of individual cells. A wide selection of methods is available. Researchers should select and optimize protocols based on the experimental question and include QC measurement of prepared libraries. After high-quality single-cell libraries have been prepared, you can proceed with confidence to sequencing. If you would like to discuss various single-cell sequencing methods and how they can be integrated with your research, contact your local Illumina expert.

STEP 3

Sequencing

After libraries have been prepared from viable single cells, sequencing is the next crucial step. All Illumina sequencing platforms use sequencing by synthesis (SBS) chemistry. Illumina's SBS chemistry detects single bases as they are incorporated into growing DNA strands in a massively parallel fashion and has been referenced in over 300,000 independent publications. Illumina sequencing systems can deliver data output ranging from 300 kilobases to multiple terabases in a single run, depending on the instrument type and configuration. This chapter presents the Illumina sequencing systems that are appropriate for single-cell studies and discusses important considerations to ensure a successful sequencing run.

Considerations for sequencing

Sequencing coverage/read depth

For single-cell sequencing, read depth is not discussed in the number of reads per base, but in the number of reads per cell. The required sequencing depth for single-cell sequencing will depend on several factors, including sample type, the number of cells that need to be analyzed, experimental objectives, and more. For single-cell RNA-Seq for example, unbiased cell-type classification within a mixed population of distinct cell types can be achieved with as few as 10,000 to 50,000 read pairs per cell (**Table 6**).[27]

Table 6: Some examples of the recommended number of reads per cell calculated by 'reads per modality' x '# of cells' = total sequencing requirement.

	Single-Cell Method	Recommended # Read Pairs Per Cell
Genome	Targeted DNA-Seq	Depends on amplicon panel size
	scWGS	>200M
Epigenome	scATAC-Seq	>25,000
	scMethylations (bisulfite)	1M per cell
Transcriptome	3' gene expression	20,000
	5' gene expression	50,000
	10x FLEX Fixed RNA-Seq	10,000
	IR-Seq	5,000
	Takara SMARTer	1M-2M per cell (>300,000 per cell)
Proteome	Antibody sequencing	100 per antibody per cell

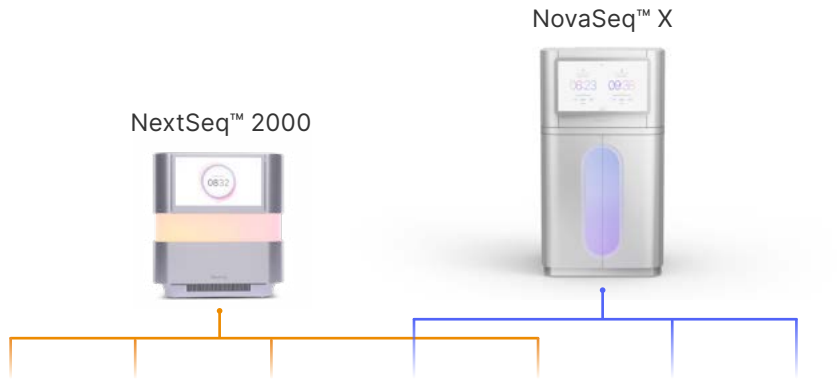


Table 7: Number of cells that can be sequenced on Illumina flow cells (FC) with popular single-cell methods.

		Flow Cell	P1	P2	P3	1.5B	P4	10B	25B
	Reads per cell	Reads per FC	100M	400M	1.2B	1.6B	1.8B	10B	26B
10x Flex (FFPE)	10,000	Number of cells sequenced	10K	40K	120K	160K	180K	1M	2.6M
10x 3'GE	20,000		5K	20K	60K	80K	90K	500K	1.3M
10x ATAC	25,000		4K	16K	48K	64K	72K	400K	1M
10x Multiome	45,000		2.2K	9K	27K	36K	40K	222K	578K
Fluent	15,000		6.6K	27K	80K	107K	120K	667K	1.7M
Parse	20,000		5K	20K	60K	80K	90K	500K	1.3M
Scale	20,000		5K	20K	60K	80K	90K	500K	1.3M

Lower read depth is adequate for identifying major cell types present in a sample. However, this read depth may not be sufficient when more heterogeneous cell populations are studied, and it is unlikely to provide detailed information on gene expression especially within very rare cell populations. In such cases, deeper sequencing is required for improving cell identification and gene detection with low expression. Indeed, there are reports that 500,000 reads per cell are sufficient to detect most genes expressed in a cell. One million reads per cell approaches sequencing saturation, enabling the estimation of the mean and variance of gene expression. Libraries differ in terms of molecular composition. Thus, the diversity of the library is a primary consideration for sequencing depth. As shown in **Figure 6** below, samples from different library workflows require different read depths.[28,29] Ultimately, the required sequencing depth will largely depend on sample type, read type, and experimental objective, and it will need to be optimized for each study.[30,31]

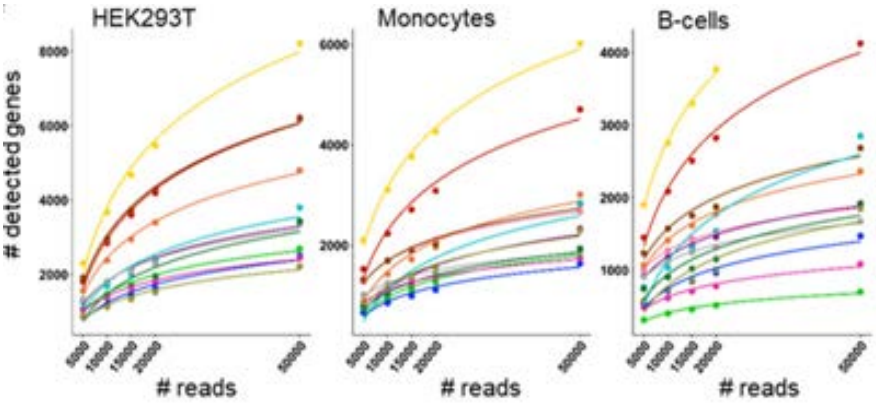


Figure 6: Sequencing depth satutation curves showing the number of genes detected for reads in three different cell types – HEK293T, monocytes, and B-cells. The different colored lines represent different chemistries benchmarked on each cell type.

Compatible sequencing systems

Although all Illumina sequencing systems are capable of sequencing single-cell libraries, the sequencing system chosen for a single-cell sequencing experiment will be determined largely by the research question and scale of the study.

Table 8: Illumina sequencing systems recommended for single-cell sequencing

	 MiSeq i100 Series	 NextSeq 1000 and NextSeq 2000	 NovaSeq 6000 system	 NovaSeq X Series
Description	The MiSeq i100 and MiSeq i100 Plus Systems provide results 4× faster than the MiSeq System, with sequencing run times as fast as four hours.	The NextSeq 1000 and NextSeq 2000 systems offer 2.5x the throughput of the NextSeq 550 system. These sequencers empower deeper investigation, larger studies, and higher resolution.	The NovaSeq 6000 is a scalable, high-throughput Illumina sequencing platform.	The NovaSeq X represents the most powerful Illumina sequencing platform to date.
Ideal application	Compatible with high-throughput single-cell studies.	Compatible with mid to high-throughput single-cell studies.	Compatible with high-throughput single-cell studies.	Ideal for high-throughput single-cell studies.
High throughput?	-	x	x	x

Learn more about Illumina sequencing systems.

Paired-end sequencing

Paired-end sequencing involves sequencing both ends of the DNA fragments in a library and aligning the forward and reverse reads as read pairs. This results in more accurate read alignment. Paired-end sequencing is key for single-cell sequencing because read 1 reads the unique molecular identifier (UMI) while read 2 captures the cDNA.

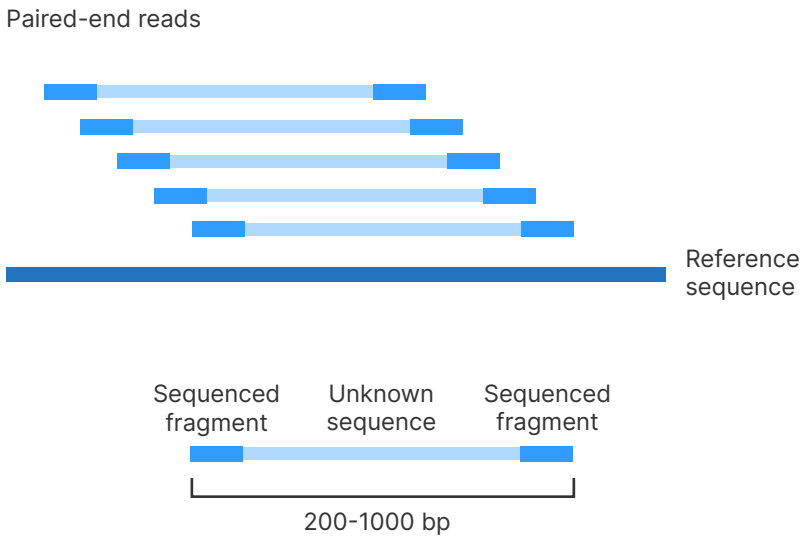


Figure 7: Paired-end sequencing. Paired-end sequencing enables both ends of the DNA fragment to be sequenced.

Assessing the success of single-cell sequencing runs

In Figure 8 below, we show an example of a single-cell RNA-Seq run performed on libraries prepared using 10x Genomics Chromium Single Cell Solution on the Illumina NextSeq 2000, as displayed in BaseSpace™ Sequence Hub (BSSH), the Illumina data analysis environment.

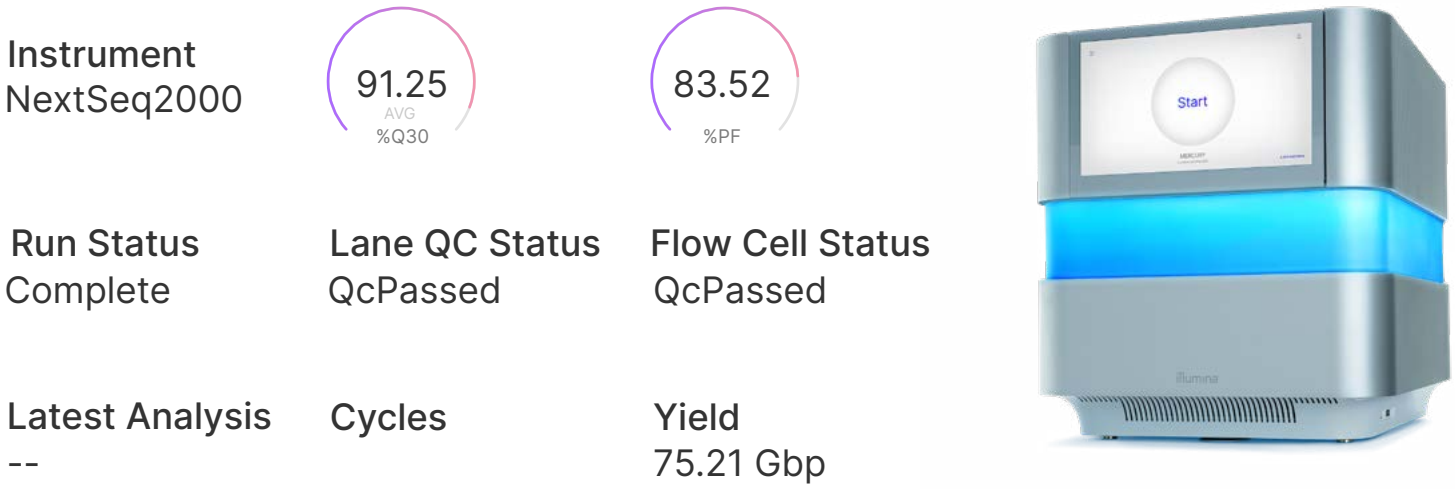


Figure 8: Example single-cell RNA-Seq run performed on the Illumina NextSeq 2000

The initial page in BSSH for your sequencing data provides information on whether your overall sequencing run has met the minimum requirements for a "good run". The Q30 value measures the probability of a base being called correctly, with a miscall occurring 1 out of 1000 times or less. In other words, the probability of the base being called correctly is 99.9%. Under the CHARTS tab on the BSSH page (**Figure 9**), you can inspect additional data on the sequencing run.

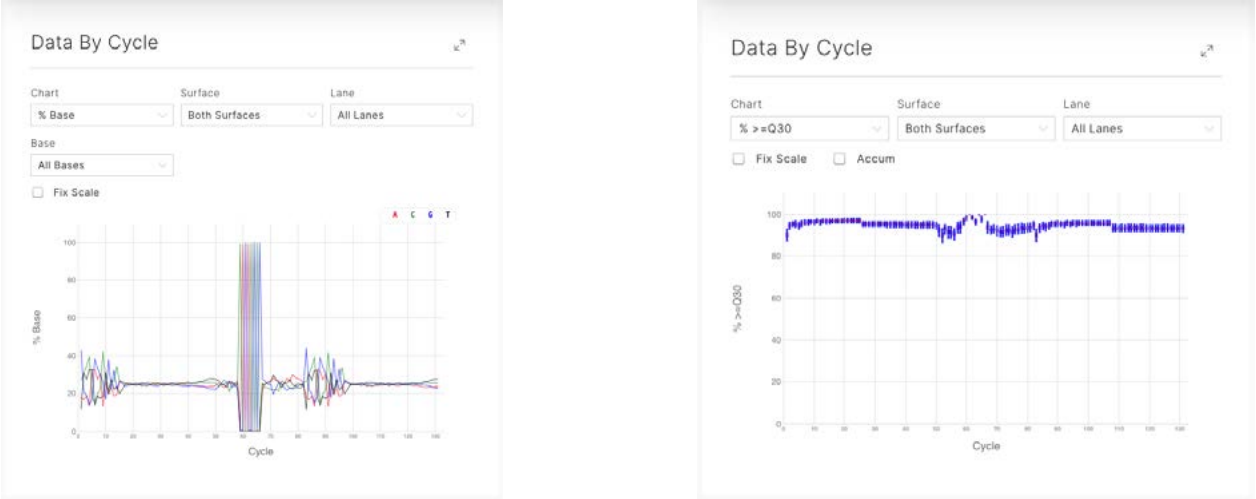


Figure 9: Cycle plots for %base and Q30 scores for a single-cell sequencing run in BSSH.

In Figure 9 above, the cycle plots for % base (left) and the Q30 score (right) are shown. While there is fluctuation in the % base reading throughout the sequencing run, this is to be expected due to the variation in base composition of the libraries.

As can be observed, the Illumina sequencer can handle this variation in base composition and maintains a high Q30 throughout the run (**Figure 9**, right image).

Note: Q30 scores are based on PhiX (from a bacteriophage), the first DNA genome to be sequenced by Fred Sanger. Due to its small and well-defined sequence, PhiX has been commonly used as a control for Illumina sequencing runs.

Need to check to see if your run looks good? Check out our blog post to find out.

Instrument control software

Instrument control software is preinstalled on all Illumina sequencing systems. Control software guides users through the steps to load the flow cell and reagents and provides an overview of quality statistics for monitoring a sequencing run as it progresses. The software can also perform image analysis, base calling, and base call quality automatically.

Summary

Illumina sequencing systems offer high data accuracy with flexible throughput to deliver a proven NGS solution for single-cell sequencing studies, regardless of scale. It is important to consider read depth before committing to a sequencing run. This will help to balance cost with sequencing parameters best suited to meet your experimental objectives. Researchers should consider sequencing metrics such as % ≥ Q30 scores after sequencing to ensure a successful sequencing experiment. After you obtain high-quality reliable sequencing data, you can proceed with data analysis, visualization and interpretation.

STEP 4

Data analysis, visualization, and interpretation

After single-cell sequencing is complete, you can now perform data analysis. Generally, the analysis pipeline for single-cell sequencing experiments involves three phases: primary analysis or base calling, secondary analysis, and tertiary analysis. There is no one correct way to carry out an analysis pipeline for single-cell sequencing experiments. Many approaches and software programs are available for each step in the pipeline. The research objective, single-cell isolation platform, and general lab considerations will largely determine the specific pipeline used. This chapter outlines the steps involved in single-cell sequencing analysis and some of the available tools.

Primary analysis: file conversion

Illumina BCL Convert

The Illumina BCL Convert stand-alone app can be downloaded onto your local computer, and converts BCL files produced by Illumina sequencers to FASTQ files. BCL conversion is available on both the NextSeq 2000 and NovaSeq X sequencers to generate FASTQ files.

The CellRanger software from 10x Genomics (CellRanger mkfastq) demultiplexes BCL files generated by Illumina's sequencers into FASTQ files and has additional features that are specific to 10x Genomics libraries.

***.bcl file format**

Illumina sequencing systems generate raw data files in binary base call (BCL) format. This sequencing file format contains both the base call and the quality of that base call for each cluster on a per-cycle basis. While the BCL format is efficient for the sequencing system, it requires conversion to FASTQ format for use with user-developed or third-party data analysis tools.

***.fastq file format**

FASTQ is a text-based sequencing data file format that stores both raw sequence data and quality scores. FASTQ files have become the standard format for storing NGS data from Illumina sequencing systems and can be used as input for a wide variety of secondary data analysis solutions.

Learn more about best practices for single-cell analysis across modalities.

Secondary analysis: demultiplexing, alignment, and QC

Read mapping in alignment to a reference genome is often the first step in data analysis. Various software applications are available for alignment, including the DRAGEN™ Single-Cell App. Sequencing data can be instantly transferred, stored, and analyzed securely onboard, on–premise (server) or in the cloud, all powered by DRAGEN. There is a large collection of apps that feature intuitive push-button user interfaces that are designed for use without the need for bioinformatics expertise.

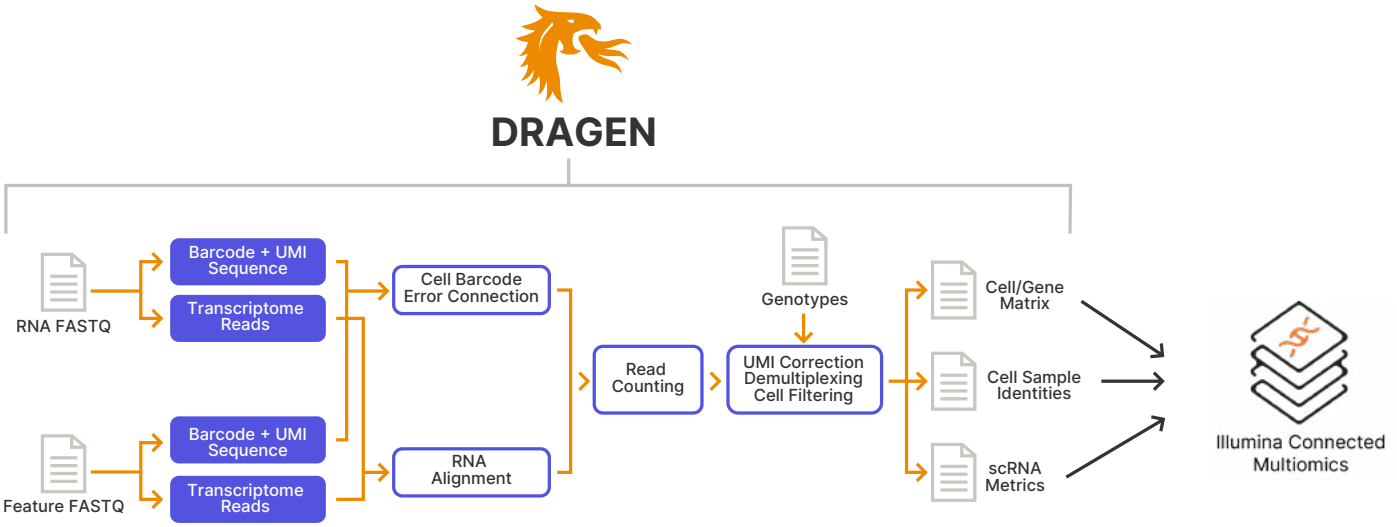


Figure 10: The BCL to Illumina Connected Multiomics workflow

Table 9: Secondary analysis BSSH app

DRAGEN App	Description
DRAGEN Single-Cell App	The DRAGEN Single Cell RNA Pipeline is a fast and scalable pipeline that can process a wide range of single-cell RNA-Seq datasets from reads to cell-by-gene expression matrices. Additionally, it is endowed with several useful extensions that allow processing of multiplexed datasets consisting of several samples (eg, using genotype demultiplexing or cell-hashing) and counting the expression of cell-surface proteins.

Expected library size and number of expressed genes

All cell types have an expected library size and a typical number of expressed genes. Cells that fall outside the expected range (too low or too high), may be either low-quality "cells" that need to be excluded from downstream analysis, or unusual "cells of interest" that warrant further investigation before inclusion in downstream analysis (**Figure 11**). For instance, technical artifacts such as dead or dying cells, or multiple-cell aggregates may be labeled as cells when in fact, they should be excluded.

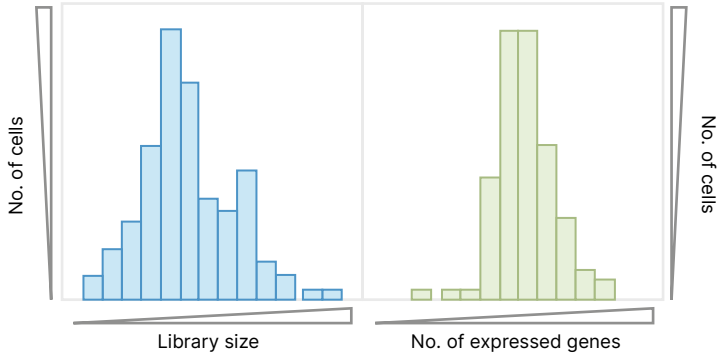


Figure 11: Filtering by library size or number of expressed genes. Cell distribution plots by library size and number of expressed genes. Cells outside of the expected range for each cell type may be poor-quality or unusual cells of interest.

Proportion of reads aligning to mitochondria and ribosomes

Another QC metric is the proportion of reads that mapped to genes in the mitochondrial genome or reads that map to ribosomal RNAs (**Figure 12**). High mitochondrial and ribosomal proportions are indicative of poor-quality cells, most likely because of increased apoptosis. These cells may be excluded from downstream analyses.

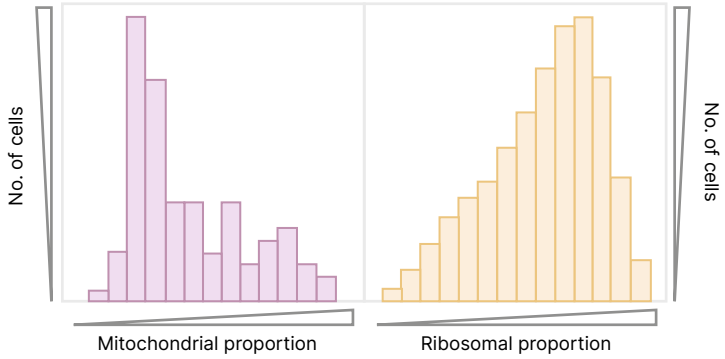


Figure 12: Filtering by mitochondrial or ribosomal proportion. Cell distribution plots by proportion of reads mapping to mitochondrial genome and ribosomes. Cells with high proportions of either are likely poor quality.

Knee plot

Plotting genic UMI counts in descending order against cell barcodes enables statistical identification of “true” cells and exclusion of noncellular barcodes (**Figure 13**). Cell barcodes above the threshold (left of the knee) have genic UMI that represent true cells, while those below the threshold (to the right of the knee) have genic UMI counts below what is expected for that cell.

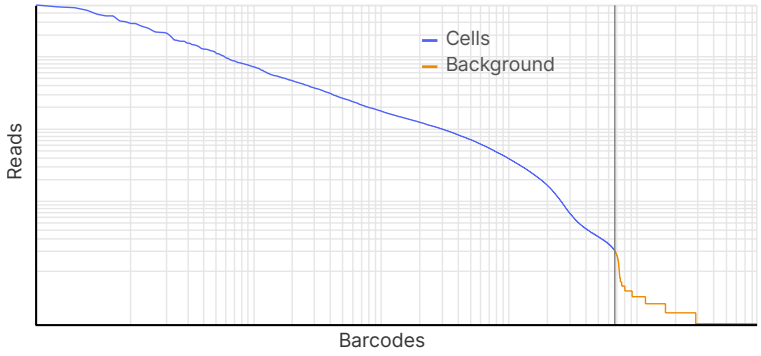


Figure 13: Filtering out noncellular barcodes. Cell barcodes to the left of the threshold (vertical red line) have genic UMI counts in the thousands, representing true cells. Cell barcodes to the right of the threshold have genic UMI counts of 1-100, typically below what is expected for live, intact cells, representing empty beads.

Evaluating doublets

Number of genes per cell for any given cell type there is a typical expected number of expressed genes. Historically, this has been used to detect and exclude doublets from downstream analysis. However, while using gene numbers per cell can be useful for single-cell sequencing experiments from a homogeneous cell population, eg, cultured cell lines, it can be problematic with complex heterogeneous tissues. Indeed, while most viable single cells may fall in a natural distribution around an expected number of expressed genes, (n), cells observed outside that distribution, eg, with roughly twice that number, (2n), may represent cells of interest that warrant further investigation and characterization. An example of this would be circulating cancer cells in a blood sample (**Figure 14**). Ultimately, given the lack of credible computational methods for detecting doublets, researchers should optimize doublet rates by experimental design.[33]

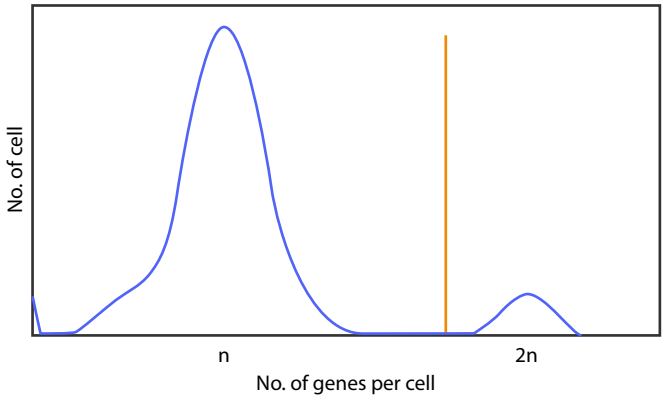


Figure 14: Representation of distribution plot of cells by gene content. Cells to the right of the threshold (vertical red line) have twice the expected number of genes per cell and are likely doublets.

Cross-species analysis

Cross-species analysis, also known as “barnyard plots”, represent the percentage of doublet cells in droplets or microwells across a given experiment. An effective way to determine cellular crosstalk is by mixing cells from two different species in one sample at a 1:1 ratio. When analyzing sample types that include cells from two different species, any cells detected with UMIs from both species represent doublets (**Figure 15**).

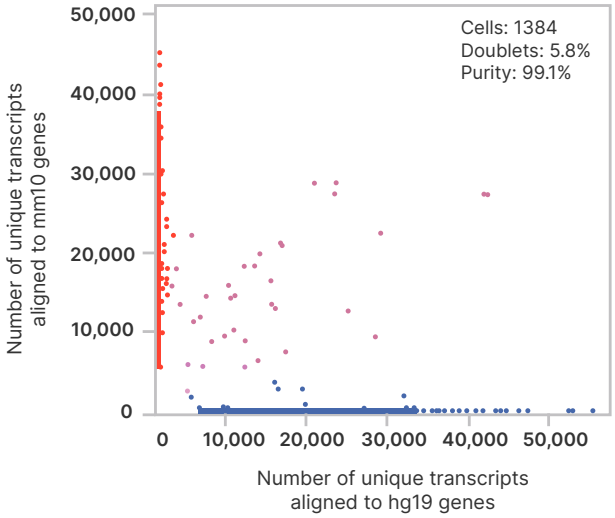


Figure 15: Doublet exclusion by species-specific UMIs. In a cell mixing experiment with cells from two different species, detection of cells with UMIs mapping to both species (purple dots) represent doublets. Sequences were aligned to both a mouse reference genome (mm10 genes, vertical axis), and a human reference genome (hg19 genes, horizontal axis).

Tertiary analysis with Illumina Connected Multiomics: data visualization and interpretation

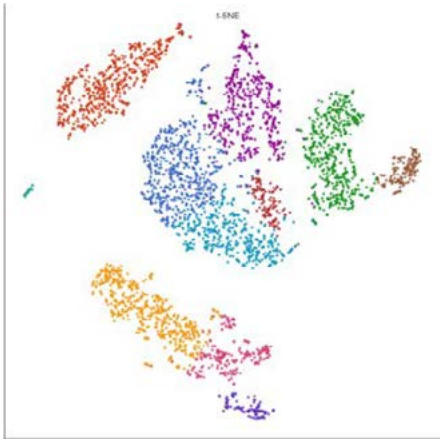
After reads have been aligned to a reference genome and secondary analysis has been performed, including data QC to remove noncellular barcodes and/or poor-quality cells, a good quality data set can be visualized and explored to gain insights into the biology of the cells being studied. Below, we outline a few of the many ways that Illumina Connected Multiomics can assist with data analysis and visualization.

UMAP

Uniform Manifold Approximation and Projection (UMAP) is a dimensional reduction technique. UMAP aims to preserve the essential high-dimensional structure and present it in a low-dimensional representation. UMAP is particularly useful for visually identifying groups of similar samples or cells in large high-dimensional data sets such as single cell RNA-Seq.

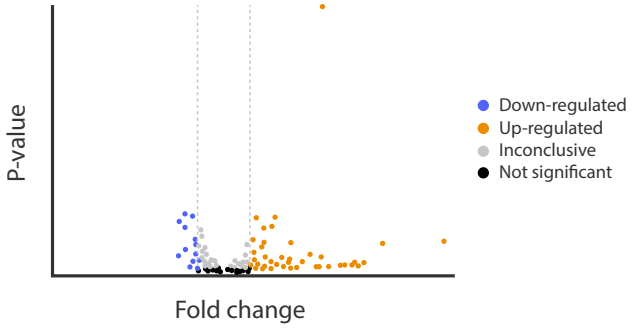
T-SNE

t-Distributed Stochastic Neighbor Embedding (t-SNE) is a dimensional reduction technique. This graph aims to preserve the essential high-dimensional structure and present it in a low-dimensional representation. t-SNE is particularly useful for visually identifying groups of similar samples or cells in large high-dimensional data sets such as single cell RNA-Seq.



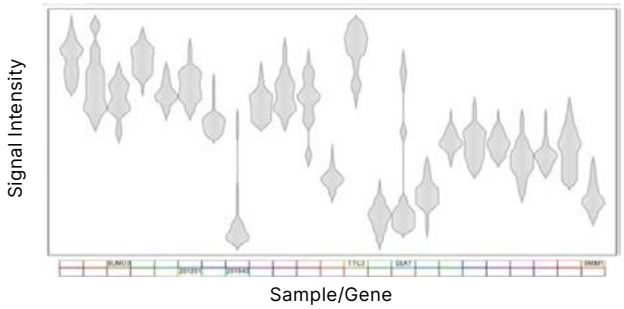
Volcano Plot

The volcano plot is a special 2-D scatter plot used to visualize significance and the magnitude of changes in features (eg, genes or transcripts) within a given comparison. By convention, the X-axis represents the fold change between the two groups on a log2 scale. On the other hand, the Y-axis shows negative log10 of the p-values from the statistical test of the comparison.



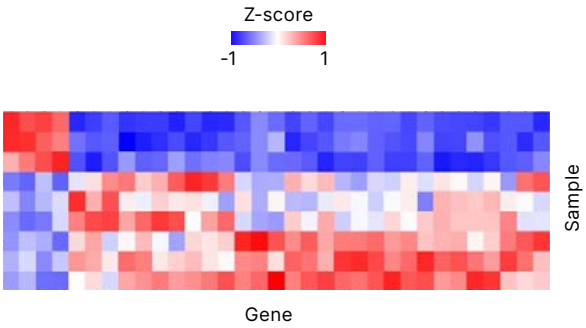
Violin plot

The Violin plot in Illumina Connected Multiomics is similar to the Profile Trellis plot in that it displays probe(set)/gene intensity values across samples and genes. However, the Violin plot has additional options not shared by the Profile Trellis plot.



Heat map

Heatmaps can be used to explore gene expression trends across samples. They are a useful tool in visualising the results of the hierarchical clustering of a gene list. They can be employed to visually distinguish between different samples and different treatments across a dataset.



While there are many software options for tertiary analysis, they can often require prior expertise and the use of multiple programs to gain a complete picture. **Tables 10** and **11** outline some of the open-source and commercial offerings available.

Table 10: Open-source tertiary analysis software.

Software	Provider	Description
Seurat	Satija Lab	Seurat is an R package designed for QC, analysis and exploration of single-cell RNA-Seq data. Seurat enables users to identify and interpret sources of heterogeneity from single-cell transcriptomic measurements, and to integrate diverse types of single-cell data.
t-SNE	Van der Maaten Lab	t-SNE is a computational technique that visualizes high dimensional data by giving each data point a location in a two or three-dimensional map. t-SNE is commonly used to visualize subpopulations with single-cell sequencing data.
UMAP	Git Hub	UMAP is an algorithm for analysis of high dimensional data. It is an alternative to t-SNE offering faster computation times.
Monocle 3	Trapnell Lab	Monocle is an R-based single cell RNA-Seq analysis software designed to determine cell development trajectory. Monocle is ideal for experiments where there are known beginning and terminal cell states.
Human Cell Atlas	Broad Institute	The Human Cell Atlas is a consortium effort that curates a data coordination platform intended to provide four key components: intake services for data submission, synchronized data storage across multiple clouds, standardized secondary analysis pipelines and portals for data access, tertiary analysis, and visualization.

Table 11: Commercially available tertiary analysis software.

Software	Provider	Description
Illumina Connected Multiomics	Illumina	Illumina Connected Multiomics is a powerful cloud-based multiomic analysis solution that enables researchers of all skill levels to explore and analyze biological data at scale. An intuitive interface and interactive visualizations enable researchers of all skill levels to reveal insights from large, complex datasets and get publication-ready figures.
SeqGeq	BD Biosciences	SeqGeq is a desktop application with an easy-to-use interface for advanced data analysis, exploration, and visualization of single-cell gene expression data. SeqGeq offers powerful data reduction and population identification tools.
CytoBank Platform	CytoBank, Inc/Beckman Coulter	Cytobank is a cloud-based platform designed for analysis and visualization of multiple single-cell data sets simultaneously.
Loupe Cell Browser	10x Genomics	The Loupe Cell Browser is designed to enable users to quickly and interactively find significant genes, cell types, and substructure within single-cell data.
Tapestri Insights	MissionBio	Tapestri Insights provides sequence import, data analysis, and visualization for single-cell DNA analysis.

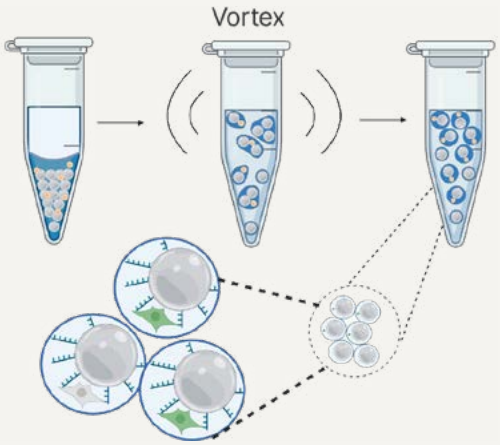
Summary

The precise analysis pipeline used for a single-cell sequencing experiment is variable and can be customized based on the study's research objectives. Generally, this pipeline includes primary, secondary, and tertiary phases. Sequences are aligned, genetic components are characterized, and data are visualized and explored, respectively, in each phase. If you would like to discuss various single-cell sequencing analysis options and how they can be integrated with your research, contact your local Illumina expert.

Simplifying sequencing with Illumina Single Cell Prep

The key to the new, straightforward Illumina Single Cell Prep lies with our pre-templated instant partitions (PIPs) isolation technique. Unlike microfluidic methods, where water-in-oil droplets are created sequentially, Illumina Single Cell Prep uses a vortex to gently capture all cells simultaneously in PIPs. Each PIP contains templated particles with unique barcodes for the identification of cells after sequencing, and simultaneous capture helps maintain the integrity of your sample by synchronizing capture time points across all cells.[2] Illumina Single Cell Prep also eliminates the need for expensive microfluidics instrumentation, wasteful consumables, and complex purchasing calculations. Illumina Single Cell Prep is reagent-based, so the number of droplets easily scales with the quantity of template particles added and kits are pre-portioned to support as few or as many cells as you need to sequence.

Using vortexing alone, Illumina Single Cell Prep isolates single cells with barcoded template particles



There are many benefits to this method, including:

Value: For the same price, you can process 5X more cells using Illumina Single Cell Prep than the leading single cell alternative[3]

Scalability: Kit sizes scale with your experiment, providing a 75% capture rate and exceptionally low multiplet rate (2–8%)[2]

Flexibility: Compatible across species, with many different cell types, and from both fixed and frozen samples[2]

Getting Started

Using Illumina Single Cell Prep requires minimal investment in a one-time purchase of familiar lab equipment engineered to enable consistency and confidence in your results. The Illumina Single Cell Prep Starter Kit requires **no service contract** and includes:

Illumina Vortex

This vortexer ensures the consistent mixing required for PIPseq and reduces variability between locations and runs.

Tube Adapter

Adapters mount to the Illumina Vortex to enable consistent mixing at the appropriate angles. Adapters are specific to the kit size and desired cell capture number.

Dry Bath

The digital dry bath precisely brings samples to the necessary temperatures during library prep protocols.

Figure 2: Simple, benchtop devices included in the starter kit.

A closer look at the Illumina Single Cell Prep workflow

Illumina Single Cell Prep brings revolutionary Single Cell Prep isolation chemistry and utilizes familiar, straight-forward lab techniques. While Illumina Single Cell Prep is compatible with any sample of properly-tagged nucleotides, the following example uses 3' RNA Gene Expression as a model workflow.

First, determine the appropriate kit for your sample size. Then, after just two minutes of vortexing, cells will be captured in PIPs and ready to be lysed.[4] Poly(T)-coated beads will capture mRNA for barcode attachment and cDNA synthesis. Illumina Single Cell Prep libraries use standard Illumina paired-end constructs and are compatible with any Illumina sequencer. The entire workflow is designed with informatics in mind, and Illumina offers easy on-cloud storage alongside class-leading analysis and visualization with Partek.

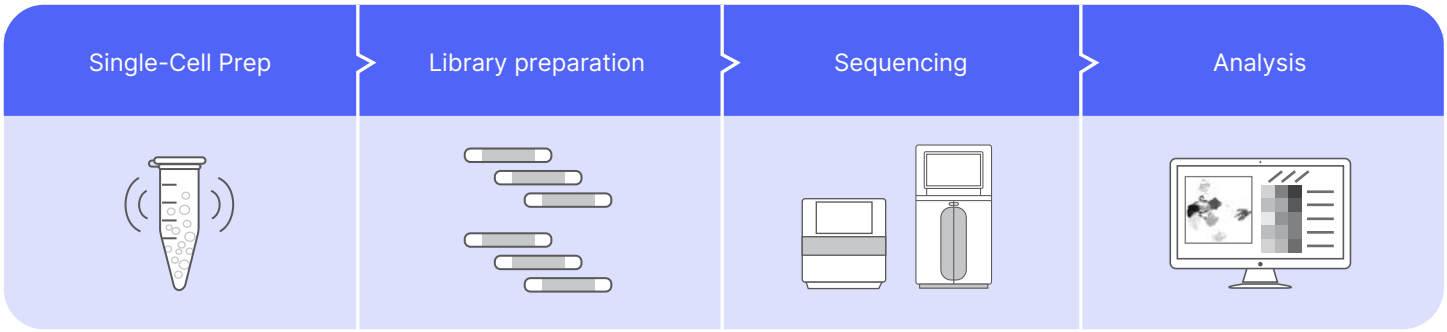


Figure 3: Illumina Single Cell Prep follows a straightforward workflow that uses familiar lab tools.

Illumina Single Cell Prep removes common barriers to single-cell sequencing

Illumina Single Cell Prep

- **Low-cost equipment with a small lab footprint**
- **No service contract required**
- **Kit sizes enable projects from 100 cells to over 1 million cells, so you only pay for what you need**
- **For the same price, you can process 5X more cells using Illumina Single Cell Prep than the leading single cell alternative[2]**
- **Simple workflow with flexible stopping points**
- **Uses familiar lab tools**

The leading single cell alternative

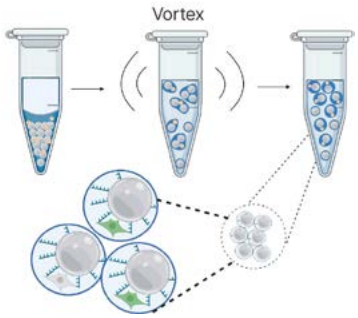
- **Equipment costs are 7X higher than Illumina Single Cell Prep[5]**
- With the leading single cell alternative, a user must buy eight lanes with 10,000 cell capacity per lane, for a total of 80,000 cells[6]
- Large batch sizes and rigid workflows increase the likelihood of unused capacity, increasing the cost per cell
- Requires specialized equipment

> There are a lot of researchers who will now be able to actually complete projects that they've wanted to do for years and have never been able to do because it's just not in the budget.

Jamie Padilla
University of New Mexico

From cells to insights

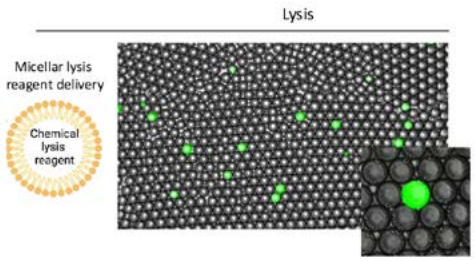
The process of taking your sample from cells to library takes 15 hours total and only six hours of hands-on time, with flexible stopping points throughout the protocol.[4] With Illumina Single Cell Prep, single-cell sequencing has never been easier.



STEP 1 Creating PIPs

The first step is to add cells to the barcoded template particles provided with Illumina Single Cell Prep kits. Illumina Single Cell Prep kits sizes enable projects from 100 cells to over 1 million cells.

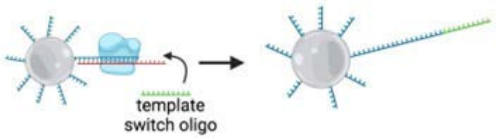
The consistent vortexing of the Illumina Vortex will emulsify the water-in-oil mixture to bring individual cells together with template particles at a reliably high capture rate. At the end of this step, the reaction is stable and at a safe stopping point, allowing breaks if necessary.



STEP 2 Cell lysis

The PIP emulsion is mixed with a chemical lysis reagent that is transferred into the single-cell droplets by micellar molecular transport. The emulsion is then incubated in the dry bath to facilitate cell lysis.

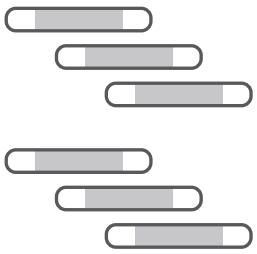
As shown by the image on the left, PIPs contain the contents of lysed cells, which helps reduce RNA background noise in samples. mRNA is captured by poly(T)-coated, barcoded beads within the PIPs.



STEP 3 Reverse transcription

After mRNA has been captured on templated beads, the PIPs are broken by chemical disruption. The hybridized mRNA particles are then mixed with a novel reverse transcriptase reaction mixture, which has been designed for improved transcription and template switching efficiency. This chemistry results in improved improved assay sensitivity.

The final result is a barcoded cDNA copy of mRNA transcripts from each cell, which are now ready for library prep.



STEP 4 Library prep

To begin library prep, cDNA is amplified with five cycles of PCR, regardless of sample size. Amplified DNA is isolated and randomly fragmented by endonucleases to generate unique cut sites that are used for PIPseq's novel molecular counting method.[2] These fragments then undergo end-repair, ligation, and amplification with standard Illumina paired-end adapters for next-generation sequencing.

Fast, accurate, and cost-effective single-cell sequencing with Illumina

Once samples are prepared, libraries can be loaded onto flow cells and sequenced. While libraries are compatible with any Illumina sequencer, the NovaSeq™ X and NextSeq™ 2000 are recommended for their throughput and cost. **With the 25B flow cell on the NovaSeq X, sequencing is 62% cheaper, twice as fast, and three times as accurate.*** For smaller studies, the MiSeq i100 is perfect for rapid, cost-effective benchtop sequencing.

*compared to S4 Flow Cell on NextSeq 6000

Illumina Single Cell Prep is compatible with Illumina sequencers

Figure 4: Approximate cell throughput per run and per flow cell assuming 20,000 reads per cell.

More cells, more discoveries, same budget

With 3' RNA sequencing, Illumina Single Cell Prep (ISCP) can process 5X as many cells for the same price as the leading single cell alternative (LSCA), allowing researchers to identify twice as many unique cell types.[3] This increased discovery power allows you to get a more complete picture of your tissue of interest and may even capture cell types missed with other isolation methods.

A. 19 distinct cell types identified from 20K mouse brain nuclei using LSCA

B. **42 distinct cell types** identified from >120K cells of the same sample prep using Illumina Single Cell Prep

C. Cell type detection is highly comparable between LSCA and Illumina Single Cell Prep

Figure 5: Matched samples of peripheral blood mononuclear cells (PBMCs) processed through Illumina Single Cell Prep (B), compared to the leading single cell alternative (A). Illumina Single Cell Prep produced highly comparable UMAP clustering and improved overall performance (C).[3]

The Illumina Single Cell Prep revolution is here

Illumina Single Cell Prep offers easy, cost-effective sequencing at any sample size

Example kit formats

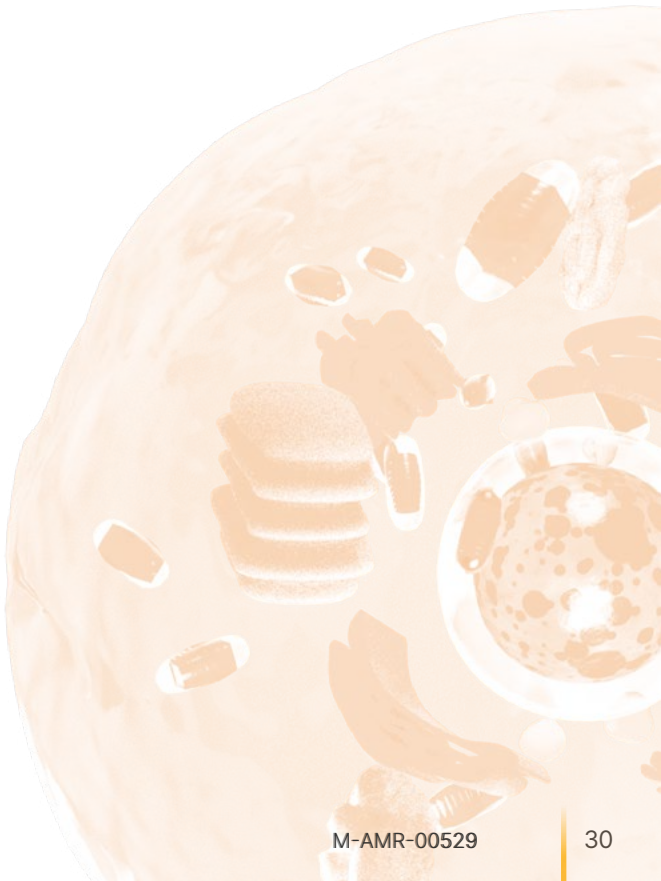
Number of cells per reaction	Up to 2,000 cells	Up to 20,000 cells	Up to 100,000 cells
Illumina Single Cell Prep in action	A UCSF study used Illumina Single Cell Prep on a small number of co-cultured macrophages and fibroblasts to reveal important crosstalk between these cells that is necessary for injury-associated fibrosis.[7]	Another study out of UCSF used Illumina Single Cell Prep on mixed phenotype acute leukemia samples to examine the diversity of cells present within chemotherapy-resistant subsets. PIPseq revealed cell profiles that were not detected by standard immunotyping.	On the other side of the country, a study based out of NYU used Illumina Single Cell Prep for a large scale single-cell study (100,000 cells/sample) to examine inflammatory genes within the brains of multiple mouse species and genotypes.

Overall Summary

Over the past decade there has been significant advancement in single-cell characterization and study, with development of new technologies for cell isolation and new methods and applications for single-cell sequencing. These advances have stimulated the launch of numerous, accessible commercial solutions for every step of the single-cell sequencing workflow, from tissue preparation through data analysis. With increasing options for single-cell isolation and interrogation, there has been a remarkable diversification of experimental protocols, each with inherent strengths and weaknesses. Researchers, therefore, face decisions about cell throughput, sequencing depth, required transcript length, whether epigenetic- or protein-level measurements should be included, and other concerns when designing an experiment.

To fully harness the potential of single-cell sequencing to elucidate complex biological systems, careful experimental design and optimization of every step of the workflow is critical.

Researchers must have clearly defined biological objectives and a rational experimental design to make informed decisions about the optimal approach for their research question. Here, we have outlined every step of the single-cell sequencing workflow and discussed important considerations and potential challenges for each, presented commercial offerings, and offered advice for designing and executing a successful single-cell study. Illumina is committed to harnessing the power of NGS for single-cell sequencing to build a deeper understanding of cellular and molecular biology, complex diseases, and environmental impacts on human health.



Glossary

flow cell: A glass slide with one, two, four, or eight physically separated lanes, depending on the instrument platform. Each lane is coated with a lawn of surface-bound, adapter-complimentary oligos. A single library or a pool of up to 96 multiplexed libraries can be run per lane, depending on application parameters.

index/barcode/tag: A unique DNA sequence ligated to fragments within a sequencing library for downstream *in silico* sorting and identification.

multiplexing: A technique to increase throughput of sequencing systems where large numbers of libraries with unique indexes can be pooled together, loaded into one lane of a sequencing flow cell, and sequenced in the same run. Reads are later identified and sorted via bioinformatic software in a process called demultiplexing.

next-generation sequencing (NGS): A non-Sanger-based high-throughput DNA sequencing technology. Compared to Sanger sequencing, NGS platforms sequence as many as billions of DNA strands in parallel, yielding substantially more throughput and minimizing the need for the fragment-cloning methods that are often used in Sanger sequencing of genomes.

patterned flow cell: A flow cell that contains billions of nanowells at fixed locations, providing even cluster spacing and uniform cluster size to deliver extremely high cluster densities.

paired-end sequencing: A process of sequencing from both ends of a DNA fragment in the same run and aligning the forward and reverse reads as read pairs.

percent passing filter (%PF): Percent passing filter (%PF) is an important sequencing QC metric that refers to the number of clusters that have passed a filter and will be retained for downstream analysis.

percent ≥Q30: Q30 is a quality score in which one base call in 1000 is predicted to be incorrect. Percent ≥Q30 refers to the percentage of bases that have a quality score of Q30 or above.

quality score (Q-score): A prediction of the probability of an error in base calling.

quantitative polymerase chain reaction (qPCR): An application that enables the measurement of nucleic acid quantities in samples. The nucleic acid of interest is amplified with the polymerase enzyme. The level of the amplified product accumulation during PCR cycles is measured in real time. These data are used to infer starting nucleic acid quantities.

read depth: See "sequencing coverage". Alternatively, in single-cell sequencing read depth is discussed not in the number of reads per base, but in the number of reads per cell.

sequencing by synthesis (SBS): SBS technology uses four fluorescently labeled nucleotides to sequence the tens of millions of clusters on the flow cell surface in parallel. During each sequencing cycle, a single labeled dNTP is added to the nucleic acid chain.

polymerization: after dNTP incorporation, the fluorescent dye is identified through laser excitation and imaging, then enzymatically cleaved to allow the next round of incorporation. Base calls are made directly from signal intensity measurements during each cycle.

sequencing coverage: The average number of sequenced bases that align to each base of the reference DNA. For example, a whole genome sequenced at 30× coverage means that, on average, each base in the genome was sequenced 30 times. Sequencing coverage can also be referred to as "read depth".

References

1. Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat Rev Genet. 2013;14(9):618-630. doi:10.1038/nrg3542
2. Jones GM, Busby E, Garson JA, et al. Digital PCR dynamic range is approaching that of real-time quantitative PCR. Biomol Detect Quantif. 2016; 10:31-33.
3. Lein E, Borm LE, Linnarsson S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science*. 2017;358(6359):64-69.
4. O'Flanagan CH, Campbell KR, Zhang AW, et al. Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. *Genome Biol*. 2019;20(1):210.
5. Digital Science. Dimensions [Software] available from app.dimensions.ai. Accessed May 20, 2021, under license agreement.
6. Percepta Associates, Inc. CMB Market Research for Illumina. 2023
7. NovaSeq X Series Enables Broader, Deeper Sequencing. https://www.illumina.com/systems/sequencing-platforms/novaseq-x-plus/applications/broad-sequencing.html. Accessed March 12, 2024.
8. Nguyen QH, Pervolarakis N, Nee K, Kessenbrock K. Experimental considerations for single-cell RNA sequencing approaches. *Front Cell Dev Biol*. 2018; 6:108.
9. Tung PY, Bilscak JD, Hiao CJ, et al. Batch effects and the effective design of single-cell gene expression studies. *Sci Rep*. 2017; 7:39921.
10. Herzenberg LA, Parks D, Sahaf B, et al. The history and future of fluorescence activated cell sorter and flow cytometry: a view from Stanford. *Clin Chem*. 2002;48(10):1819-1827.
11. Wu H, Kirita Y, Donnelly EL, Humphreys BD. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. J Am Soc Nephrol. 2019;30(1):23-32. doi:10.1681/ASN.2018090912
12. Bakken TE, Hodge RD, Miller JA, et al. Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. PLoS ONE. 2018;13(12):e0209648. doi:10.1371/journal.pone.0209648
13. Sigma-Aldrich. Centrifugation. *Biofiles*. 2011;6(5):4-14.
14. Method of the Year 2019: Single-cell multimodal omics. *Nat Methods*. 2020;17(1):1.
15. Valihrach L, Androvic, P, Kubista M. Platforms for Single-Cell Collection and Analysis. *Int J Mol Sci*.2018;19(3): DOI.10.3390.
16. Macosko EZ, Basu A, Satija R, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. 2015:161(5):1202-1214.
17. Klein AM, Mazutis L, Akartuna I, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*. 2015:161(5):1187-1201.
18. Zheng GX, Terry JM, Belgrader P, et al. Massively parallel digital transcriptional profiling for single cells. *Nat Commun*. 2017; 8:14049.
19. Pellegrino M, Sciambi A, Yates JL, Mast JD, Silver C, Eastburn DJ. RNA-Seq following PCR-based sorting reveals rare cell transcriptional signatures. *BMC Genomics*. 2016; 17:361.
20. Retig JR, Folch A. Large-scale single-cell trapping and imaging using microwell arrays. *Anal Chem*. 2005;77(17):5628-5634.
21. Han X, Wang R, Zhou Y, et al. Mapping the mouse cell atlas by microwell-seq. *Cell*. 2018;172(5):1091–1107.
22. Cao C, Lemaire LA, Wang W, et al. Comprehensive single-cell transcriptome lineages of a proto- vertebrate. *Nat*. 2019; 571:349-354.
23. Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nat*. 2019;566(7745):496-502.
24. Srivatsan SR, McFaline-Figueroa JL, Ramani V, et al. Massively mutliplex chemical transcriptomics at single-cell resolution. *Science*. 2020;367(6473):45-51.
25. Illumina. QC and rebalancing of single-cell gene expression libraries using the iSeq™ 100 System. 2020. Accessed April 2024.
26. Illumina. Scalable Nucleic Acid Quality Assessments for Illumina Next-Generation Sequencing Library Prep. 2017. Accessed April 2024.
27. Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med*. 2017;9:75.
28. Illumina. Is it possible to pool different library types in the same sequencing run? https://knowledge.illumina.com/library-preparation/general/library-preparation-general-reference_material-list/000003284. 2020. Accessed April 2024.
29. Zhang X, Li T, Liu F, et al. Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. *Mol Cell*. 2019;73(1):130-142.e5.
30. Streets AM, Huang Y. How deep is enough in single-cell RNA-seq? *Nat Biotechnol*. 2014;32(10):1005–1006.
31. Rizzetto S, Eltahla AA, Lin P, et al. Impact of sequencing depth and read length on single cell RNA sequencing data of T cells. *Sci Rep*. 2017;7(1):12781.
32. Illumina. (2013) bcl2fastq Conversion User Guide. Accessed April 2024.
33. Lafzi A, Moutinho C, Picelli S, Heyn H. Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nat Protoc*. 2018;13(12):2742–2757.

The time to get started is now

Here's why you should partner with Illumina as you embark on your single-cell studies:

Our NGS reagents, equipment, and protocols have been validated in hundreds of high-impact peer-reviewed journals.
https://pubmed.ncbi.nlm.nih.gov/?term=illumina+ngs

Illumina has the broadest range of applications to enable NGS analyses on one device and is built to support emerging changes in the field.

Illumina supports over 10,000 labs across 115 countries. Our 24/7 technical support professionals are ready to help you with any project.

Reach out to Illumina Single Cell Prep experts to learn more

Phone: +1 800 809-4566 ext 1 (Toll Free)
Email: insidesales@illumina.com
To order: https://ilmn.ly/single-cell-prep

illumina

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com