

A Complete Workflow from Single Cell Isolation to mRNA Sequencing Analysis

A workflow of BD FACS sorting combined with BD Precise assays

Alice Wang¹, Aaron Middlebrook¹, Kory Pennebaker¹, Christina Chang², Eleen Shum², Christina Fan, Suzanne Weaver²

¹BD Life Sciences, San Jose, CA, USA ²BD Genomics, Menlo Park, CA, USA

When performed on bulk samples, gene expression studies might yield only a limited understanding of complex samples. However, gene expression analyses performed on single cells can offer a powerful method to resolve sample heterogeneity and reveal hidden biology. BD's fluorescence activated cell sorting (BD FACS™) technology can accurately identify and isolate single cells or a small number of cells for use in next generation sequencing (NGS) or transcriptome analyses. BD™ Precise assays enable examination of mRNA sequences for a large number of cells simultaneously at single-cell resolution. In this report, we describe a complete workflow from BD FACS identification and isolation of single cells to BD Precise sequencing library generation, sequencing, and data analysis. The single cells can be isolated using various BD FACS sorting instruments, from our entry-level sorters to our high-end sorters, BD FACSMelody™, BD FACSARIA™ III, BD FACSARIA™ Fusion, BD SORP FACSARIA™ Fusion, and BD Influx™. The index sorting feature of these sorters connects cells' phenotypes to their molecular gene expression profiles. This paper shows a streamlined, fast method using a BD FACS sorter to select and isolate single cells for high-resolution NGS analysis. The combined workflow provides a powerful tool for studying complex biological samples at the resolution of individual cells.

1. Introduction

In recent years, biotechniques have evolved significantly to provide tools for measuring high-throughput, quantitative data. These technologies include fluorescence microscopy, quantitative polymerase chain reaction (qPCR), microarrays, and sequencing. Up until recently, these tools have been applied to bulk samples containing thousands and sometimes millions of cells. While this approach has proven powerful, it has become clear that gene expression is highly heterogeneous from cell to cell in any given biological sample. As a result, gene expression patterns that drive a particular biologic outcome are difficult to detect because the overall biological "noise" in a given sample obscures all but the most dominant gene expression patterns. Indeed, the ability to dissect gene expression heterogeneity with single-cell resolution has become a powerful approach to study complex biological phenomena such as cancer, development, and microbial ecology.^{1,2}

There are a number of ways to isolate single cells for downstream analysis. Historically, single cells or small numbers of cells were isolated and prepared by limiting dilutions, laser capture micro-dissection, microfluidics technologies, or fluorescence-activated cell sorting (FACS). FACS has emerged as a leading platform for this purpose because it is fast, enables high-throughput processing of a heterogeneous mixture of cells, and ensures the delivery of single cells or a small number of cells into a chosen receptacle at a very high purity level. Furthermore, using FACS, the single cells can be selected based on cell surface marker expression, cell size, and granularity. Sorted cells can be used for any downstream application, including next generation sequencing (NGS).²

Appropriately designed single-cell gene expression studies require interrogation of large numbers of single cells for scientific relevance and accurate results. Analysis tools have evolved from

hybridization array based technologies to quantitative real time reverse transcription PCR (qRT-PCR) and most recently RNA-Sequencing (RNA-Seq). Although RNA-Seq has become the gold standard and can be used to determine relative transcript abundance, it is costly and time consuming if each individual cell is sequenced separately. Additionally, the results can be relatively inaccurate due to errors introduced by PCR amplification bias during library construction. Therefore, single-cell RNA-Seq experiments can require considerable expertise for effective assay design, performance, and data analysis.

To address these challenges, BD Genomics has developed a novel product specifically for mRNA sequencing library preparation. The product, the BD™ Precise Library Prep kit, is designed to examine either hundreds of genes or the whole transcriptome using mRNAs from each single cell in a high-throughput manner. Based on the Cellular Research (acquired by BD in 2015) patented Molecular Indexing technology, BD™ Precise assays provide quantitation of target transcripts in a simplified workflow.

BD Precise assays combine molecular and sample indexing in 96-well and 384-well formats, enabling researchers to sequence up to 4,000 cells at once into a single Illumina® MiSeq run without new equipment or extensive training. BD™ Precise targeted assays can be designed for specific genes and pathways, enabling increased accuracy and sensitivity for low-expression targets in rare and limited samples. BD™ Precise Whole-Transcriptome Analysis (WTA) assays enable the similarly high accuracy and sensitivity when analyzing the whole transcriptome. Combined with robust design and data analysis pipelines, BD Precise assays deliver a simple solution for both targeted and whole transcriptome gene expression analysis.^{3,4}

Single-cell mRNA sequencing is a powerful method for defining cell-to-cell variation, either within a cell population or among different cell populations. Here, we describe a complete workflow for cost-effective, high-throughput, single-cell mRNA sequencing by combining BD FACS single-cell sorting technology with BD Precise Molecular Indexing technology.

2. Overview of the Complete Workflow

The workflow of single-cell mRNA sequencing using BD FACS cell sorters to isolate the single cells starts from a single-cell suspension. The single-cell suspension can be a natural cell suspension, such as peripheral blood or bone marrow, or a dissociated healthy or cancer tissue sample. The sample can be stained to enable exclusion of dead cells, as well as stained with appropriate antibodies to cell surface and intracellular markers to enable the selection of specific cell types (for example, T cells or cancer cells). The cells are then interrogated on a BD FACS sorting instrument. The cells of interest are isolated into each well of a BD™ Precise 96-well assay plate, with one cell in each well. BD Precise plates are preloaded with a detergent-based single-cell lysis buffer along with RNA stabilizers, molecular indices, and sample-specific barcodes. The plates with sorted cells can be kept frozen at –80°C before further processing.

These samples are then processed according to the BD Precise assay protocol to generate a sequencing library. BD Precise Molecular Indexing technology enables the user to combine all 96 samples from one plate into a single tube for bulk processing and library creation. The library is then sequenced on an Illumina sequencing instrument. Automated data analysis can be performed on the Seven Bridges Genomics Platform using a streamlined computation pipeline, tailor-made for BD Precise generated datasets. Access to this pipeline is included with the purchase of BD Precise assays. A secondary data analysis from the mRNA expression data can be performed to dissect biological significance.

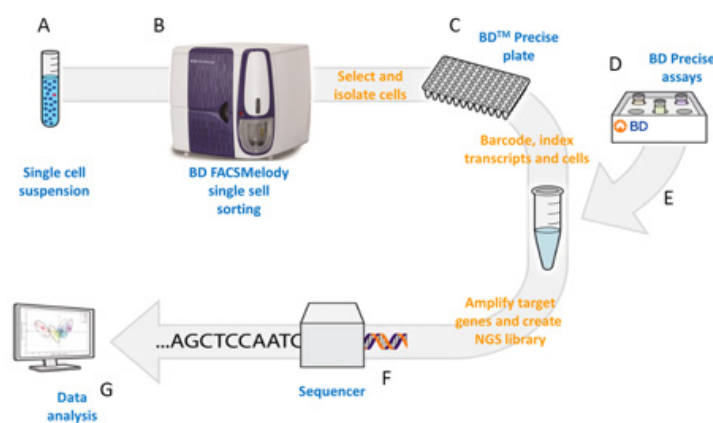


Figure 1. A complete workflow.

- Single cell suspension is prepared and stained with antibodies to surface marker, nuclear, and/or viability dyes.
- Cells are analyzed and sorted using the BD FACS sorter based on desired markers into BD Precise 96-well plates.
- The BD Precise assay protocol is followed to generate a sequencing library.
- After sequencing, data deconvolution is performed with an automated algorithm using a Seven Bridges Genomics pipeline.
- mRNA transcript counts were analyzed to dissect the biological relevance.

3. Isolation of Single Cells by BD FACS Sorters

3.1 Principle of FACS Cell Sorting

FACS is a technology that isolates or sorts out cells of interest from a single-cell suspension. The cells that pass through this instrument are not damaged by the process and can be collected for downstream analysis and/or further processing such as cell culture or transfection. Prior to sorting the cells, samples can be stained with fluorescent antibodies specific to cell surface markers or intracellular marker proteins, or nucleic acid dyes. This makes it possible to sort live cells, as well as different cell types, based on the expression levels of target cell markers. A FACS instrument incorporates the fluidic, electronic, and optical systems of a flow cytometer. When a sample is introduced into a FACS sorter, the fluidic system funnels the cells into a single stream so that they pass one at a time through a laser beam or a series of laser beams. When each cell passes through the laser beam(s), it scatters light and may emit fluorescence light based on cell surface or intracellular fluorescent staining. The scatter light pattern is determined by the cell size, granularity, and complexity.

Light scatter and fluorescence signals are collected and detected, and a decision is made whether the cell is of interest and should be sorted. For FACS sorters, when the stream containing the cells passes through a nozzle, it will be broken into small drops with one cell or no cell in each drop. Once a cell is identified to match the parameters for sorting, the instrument applies an electrical charge to the droplet containing the cell. The droplet is deflected and attracted to the collection device. The droplets that are not charged continue in the stream to a waste container. Flow cytometry cell sorters can collect the cells into collection tubes, trays, and slides, as well as 96-well or 384-well tissue culture and PCR plates.

3.2 General considerations for single cell sorting for sequencing analysis

All BD FACS sorters can prepare single cells for single-cell sequencing analysis. In preparing single cells for sequencing analysis, the cells need to be:

- 1) Delivered individually into each well. Doublets and triplets need to be excluded because they can compromise the data accuracy.
- 2) Viable and maintain integrity. The gene expression profile of a dead or damaged cell will likely be significantly different from a live cell of the same phenotype.
- 3) Free of DNA carryover from previously sorted samples. Contaminating DNA from previous runs can compromise a sample and the resulting data.

To achieve these goals, correct cell sorter operation is required. This includes:

- 1) Cleaning the instrument at regular intervals. Make sure to follow the suggested aseptic cleaning procedure for each instrument.

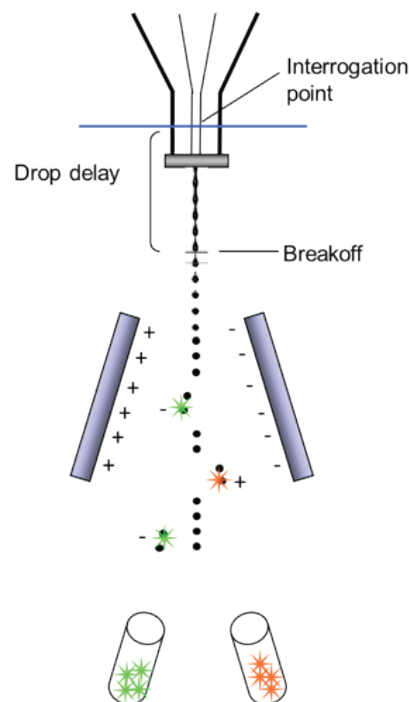


Figure 2. Cell sorting process. This diagram shows how the cells passing through the fluid stream are interrogated and selected based on fluorescence signal or light scatter. The fluid breaks into drops carrying the selected cells, which are electrically charged and collected.

- 2) Applying proper sample and sheath pressure. Lower sheath pressure causes less damage to the cells. However, the stream may not be very stable when the pressure is too low.
- 3) Aligning the deflected drop to the center of the sorted wells. For good well alignment, a small angle for the drop deflection is preferred. The nearly vertical deflection stream will ensure that the sorted drop hits the center of the well of the collection plate.
- 4) Eliminating static electricity buildup to ensure a steady, straight sample stream and deflection stream. For this purpose, lab personnel should wear proper lab clothing; the air flow in the lab should be checked to make sure the air is not too dry; the cells should be sorted into polypropylene containers rather than polystyrene tubes; static-free wipes should be used.
- 5) Selecting the “Single Cell” Sort Mode. This is the Sort Mode for the accurate delivery of one cell into each collection well.

3.3 Sorting Single Cells with BD FACS sorters

BD makes a variety of sorters for sorting single cells to fit the researcher's experimental need, the lab's budget, and the operator's flow sorting experience and preference. The newly developed BD FACSMelody sorter has been designed for ease of use. Our other instruments, from low- to high-parameter range sorters, including the BD FACSaria III, BD FACSaria Fusion, BD SORP FACSaria Fusion, and BD Influx, can sort single cells for mRNA sequencing applications. Go to the BD Biosciences website, www.bdbiosciences.com → INSTRUMENTS → Cell Sorters, for detailed information about the features of the different sorters.

These sorters all have the index sorting feature. This feature provides the phenotype information of every cell sorted into a multi-position sort device, such as a 96-well plate. This information includes light scatter and fluorescence channel MFI (median fluorescence intensity) values. The index sorting feature helps track the cell that is being sorted. It links the cell phenotype information with the cell gene expression profile.

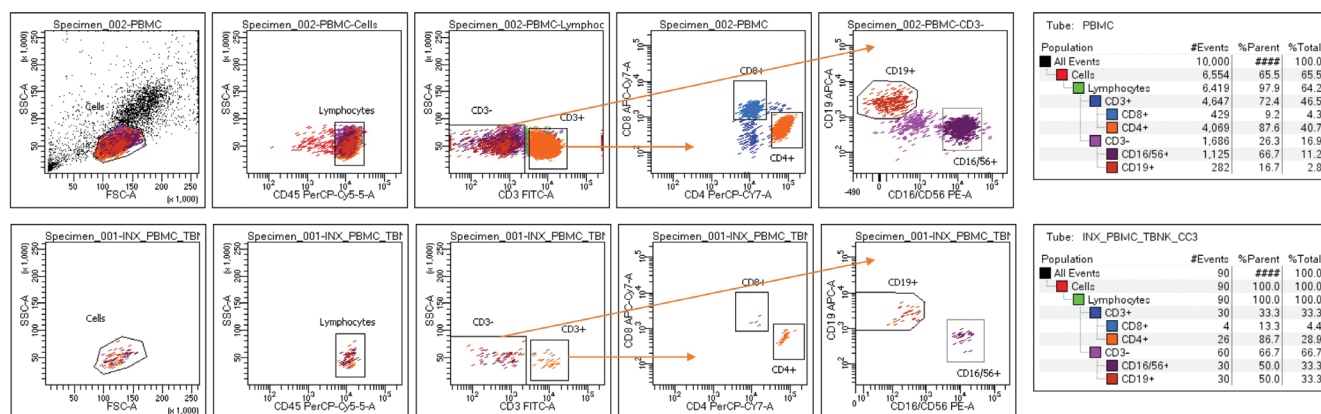


Figure 3. Gating of single cells for NGS. Peripheral blood mononuclear cells (PBMCs) were isolated from human whole blood collected in BD Vacutainer® CPT™ tubes (Cat. No. 362761). Isolated PBMCs were stained with the antibody cocktail containing the cell surface markers CD16 PE, CD56 PE, CD3 FITC, CD45 PerCP-Cy™5.5, CD4 PerCP-Cy™7, CD8 APC-Cy7, and CD19 APC (BD Multitest™ 6-color TBNK Reagent, Cat. No. 644611). Pan T cells were identified by CD45+CD3+ staining; B cells were identified by CD45+CD3-CD19+; and NK cells were identified by CD45+CD3-CD16/CD56+. Thirty cells from each of the T, B, and NK populations were sorted into each BD Precise plate. The upper panel plots show the gating strategy before sorting. The lower panel shows the sorted cells. The information for the cells sorted is captured by the index sorting feature.

4. Generating a Gene Expression Sequencing Library with BD Precise Assays

BD Precise assays are ultra-sensitive RNA sequencing assays designed to examine large numbers of low input mRNA samples in a high-throughput manner. To improve workflow and mitigate amplification bias, these assays append three layers of DNA barcoding to each mRNA molecule.

- 1) A molecule-specific barcode (Molecular Index) is used to identify the molecule of origin of each sequenced read in an amplified sequencing library. This provides a more accurate representation of transcript abundance in the original sample prior to amplification.
- 2) Well-specific barcodes are used to link individual molecules from the same single cell. They also allow for pooling of individual cells to simplify the library preparation protocol.

- 3) Plate-specific barcodes are used to allow content from multiple plates to be run in a single sequencing run.

This technology fits perfectly for single-cell mRNA sequencing analyses to allow high throughput with a simplified workflow and resolution of single cells at an affordable price.

BD has developed a series of BD Precise targeted assays, and a WTA (whole transcriptome analysis) assay to fit the specific needs of researchers.

4.1 BD Precise Targeted Assay Sequencing Library Generation

Precise Targeted assays are designed to interrogate approximately 80-130 genes simultaneously. It can also be designed for up to 240 genes. Pre-designed panels are also available, covering specific biological areas, such as breast cancer, regulatory T cells, prostate cancer, etc.

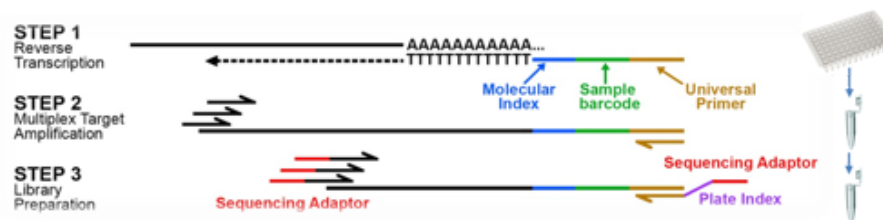


Figure 4. BD Precise targeted assay workflow. Each well of the BD Precise targeted plate is preloaded with lysis buffer, dNTPs, and an oligo dT primer containing sample and molecular indices along with a universal priming site. When cells are sorted into the wells on the plate, the cells are lysed and mRNA contents are released. During cDNA synthesis, individual transcripts are tagged with a unique molecular index to provide a more accurate representation of transcript abundance in the original sample prior to amplification. Following cDNA synthesis, the sample indices allow all of the wells to be pooled together into a single tube to simplify sample handling. Two rounds of PCR amplification are done to amplify the specific genes of interest and to add sequencing adaptor and plate-specific indices. Each targeted panel allows users to interrogate approximately 100 genes simultaneously. The amplified library is run on an Illumina NGS instrument.

4.2 BD Precise WTA Assay Sequencing Library Generation

BD Precise whole transcriptome analysis assay is also available for gene expression discovery studies. In these assays, any transcript with a poly A tail will be interrogated.

WTA by second-strand ligation



Library preparation

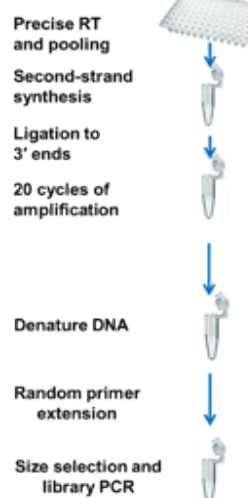
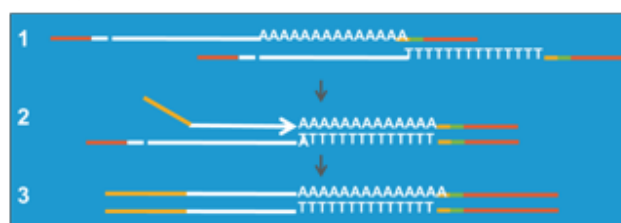


Figure 5. BD Precise WTA assay workflow. Cell lysis and cDNA synthesis are identical to the BD Precise targeted assay. Following cDNA synthesis, an optimized second-strand reaction is performed on the pooled samples, followed by adapter ligation to the 3' end of the first-strand cDNA which, along with the RT primer's universal sequence, is used as a primer-binding site for whole transcriptome amplification. Library preparation is performed by random priming from the 3' ends of the amplified transcriptome, followed by size selection and PCR amplification, which adds Illumina flow cell binding sequences and optional plate indices.

5. Data Analysis

5.1 BD Precise Assay Sequencing Data Analysis Pipeline



Figure 6. Data processing steps. The BD Precise assay sequencing data analysis pipeline is provided via Seven Bridges Genomics, a secure online cloud environment. The pipeline takes FASTQ files from the sequencing and collapses molecular indices using a proprietary algorithm. The number of molecules per gene per cell is reported in a simple, easy-to-read Microsoft® Excel® (or tabular) file, which can then be imported into statistical software packages such as R for secondary data analysis.

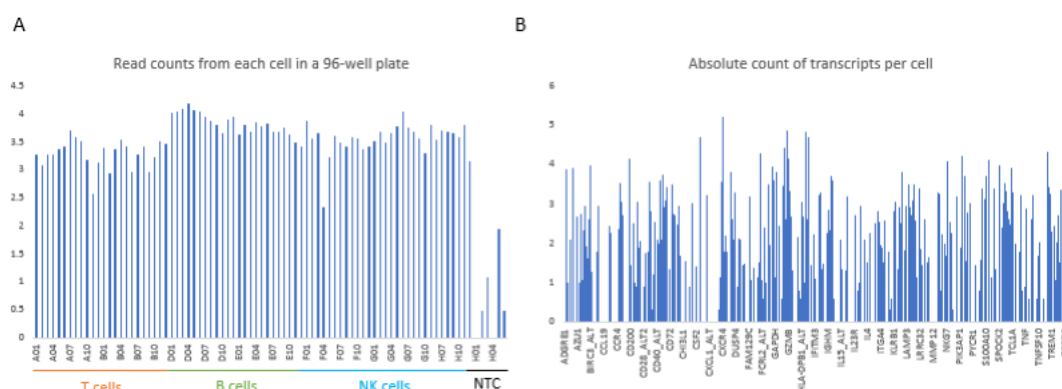


Figure 7. Number of reads and transcripts obtained from single cells. For the single T, B, and NK cells sorted from the experiment shown in Figure 3, mRNA expression was measured by the BD Precise targeted assay using a panel targeting 100+ genes expressed by immune cells in human peripheral blood. Genes included those that encode surface proteins, cytokines, and transcription factors. **A.** Read counts from each cell in a 96-well plate (log scale). Individual cells were sorted into a BD Precise plate, featuring plate indexing that allows multiple plates (up to 4,000 cells) to be pooled into a single MiSeq run if desired. **B.** The average transcript count across 96 single cells from the same plate shown in Panel A after the reads have been collapsed into absolute molecule counts. Data illustrates absolute mRNA quantitation in single cells across a large measurement range. This information can be obtained from primary data analysis. NTC: no template control.

5.2. Secondary Data Analysis

Once the transcript counts are obtained from the automated Seven Bridges Genomics data analysis pipeline, the sequencing data can be further analyzed by various methods for biological relevance.

A mixing cell experiment was conducted to demonstrate several different ways of data analysis (Figure 7). A breast cancer cell line, T47D cells, and a T-leukemia cell line, Jurkat cells, were stained with their specific cell surface marker EpCAM or CD3, and mixed at a 50%:50% ratio. Dead cells were eliminated by propidium iodide (PI) staining. Following Precise WTA assay, the gene expression level was detected for each cell line.

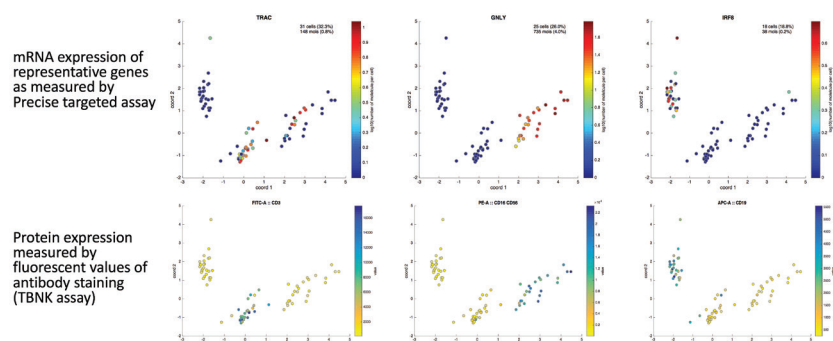


Figure 8. Linking protein expression and mRNA expression using index sorting and BD Precise assay. Each cell in the TBNK sorting experiment was plotted on a 2-D space according to principal component analysis of the detected mRNA gene expression profiles (PC1 and PC2 are shown). Three populations were observed based on gene expression profiles, representing T, B, and NK cells. Top: each cell is colored by detected transcript level from the BD Precise assay. Markers shown: TRAC (TCR alpha constant region), GNLY (granulysin, secreted protein), IRF8 (transcription factor, intracellular). Bottom: each cell is colored by surface protein level as measured by FACS: CD3, CD16/56, and CD19, respectively. The cell type specific surface protein expression agrees largely with the cell types identified by gene expression. The index sorting feature enables the linking of protein expression with mRNA expression

6. Conclusions

We demonstrated the power of a streamlined workflow from single-cell preparation with various BD FACS cell sorters, to single-cell targeted or whole transcriptome sequencing analysis with BD Precise assays. The combined workflow provides investigators a complete solution for high-throughput single-cell gene expression sequencing analysis. The index sorting feature of BD FACS sorters connects protein expression information and gene expression information. Overall, these powerful tools will allow researchers to dissect otherwise complicated biological systems for a better understanding at the single-cell level.

7. References:

1. Gawad C, Koh W, Quake, SR. Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics*. 2016;17:175-188.
2. Baslan T, Kendall J, Rodgers L, et al. Genome-wide copy number analysis of single cells. *Nature Protocols*. 2012;7:1024-1041.
3. Fu GK, Xu W, Wilhelmy J, et al. Molecular indexing enables quantitative targeted RNA sequencing and reveals poor efficiencies in standard library preparations. *PNAS*. 2014;111:1891-1896.
4. Fu GK, Hu J, Wang P, Fodor SPA. Counting individual DNA molecules by the stochastic attachment of diverse labels. *PNAS*. 2011;108:9026-9031.

Acknowledgments

The authors thank Pierce Norton, Kathleen Shelton, and Amy Tam for their feedback and suggestions.