Math 612: Single Cell Analysis **2019W Term 1**

Lecture 4: September 17

Lecturer: Geoffrey Schiebinger *Scribe: Laura Stankiewicz*

4.1 Single Cell RNA Sequencing

Single cell RNA sequencing (scRNA seq) provides the transcriptomic expression profiles of individual cells to allow researchers to identify patterns in gene expression throughout the population.

4.1.1 Overview

Input: a population of cells (ex. organism, tissue, gut microbiome ecosystem).

Output: the expression profile of some of the cells.

4.1.2 Steps

Single cell RNA sequencing involves the following procedure.

1. Cells are dissociated to get a suspension.
2. Cells are isolated into individual populations.

   This step used to be performed using FACS. However, an efficiency breakthrough was achieved using microfluidic droplets generated in a microfluidic chip (see Figure 1).

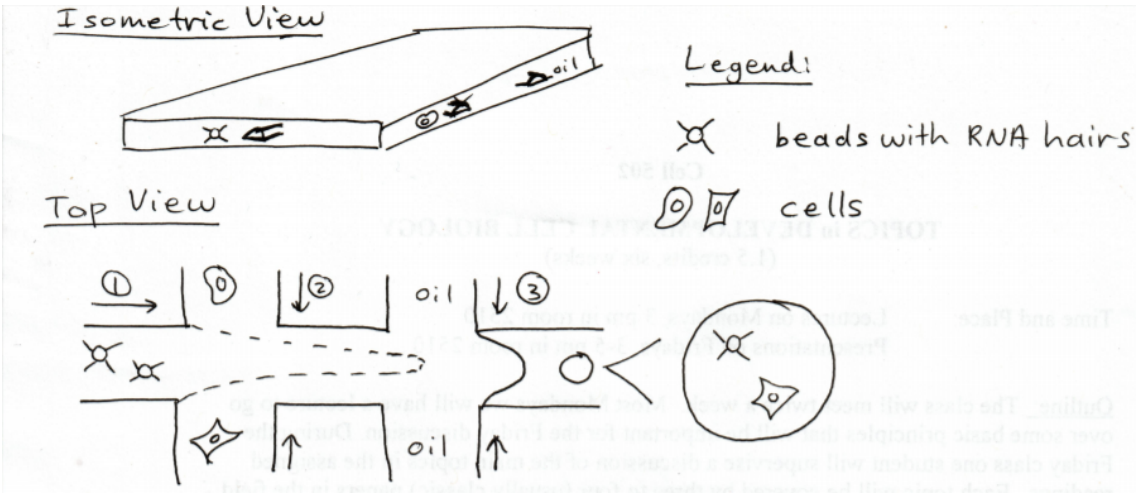

Figure 4.1: Method for generating microdroplets.

Each droplet is assumed to contain 1 bead and 1 cell.

3. Within each droplet cell RNA is captured on bead hairs.

   The lysis buffer contained in each droplet dissolves the cell membrane, allowing the RNA to disperse and bind to the hairs on the beads that contain poly-T tails (see Figure 2). Each bead hair is assigned a cell barcode and a molecule barcode which is used in later analysis to map back to the individual cell and RNA read, respectively. These barcodes are referred to as Uniquie Molecular Identifiers (UMI).

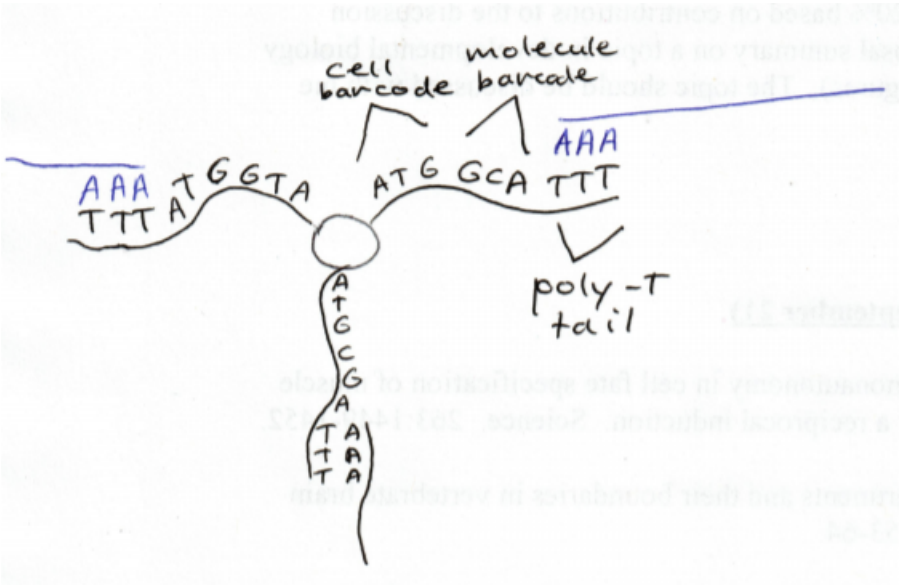

Figure 4.2: RNA binding to bead hairs.

4. RNA is converted into cDNA to increase stability.

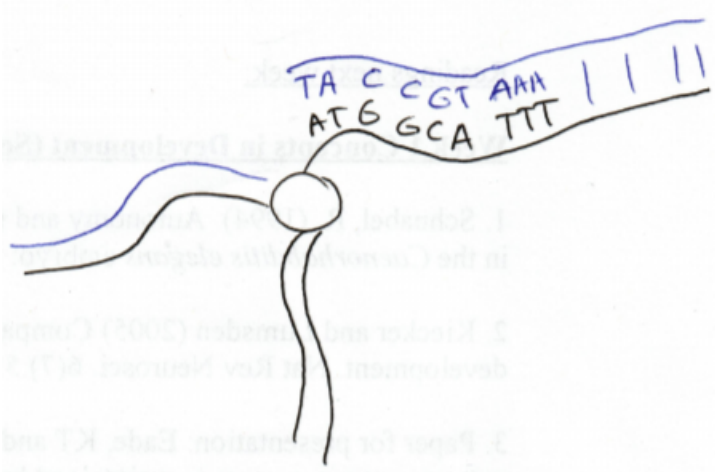

Figure 4.3: Bead structure after cDNA polymerization.

5. The cDNA library is amplified and then sequenced.

   To estimate frequency of RNA in each cell, the RNA must first be amplified (see Figure 4). Amplification via polymerize chain reaction (PCR) doubles the concentration of DNA each cycle. However, due

to differential binding strength of GC and AT base pairs the probability, p, of amplification changes per gene. This introduces amplification bias. A correction factor is introduced to account for amplification bias.

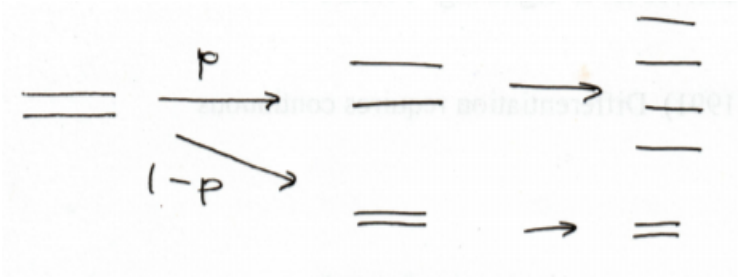

Figure 4.4: Depiction of amplification bias in PCR.

After amplification, the UMI allows the user to generate an expression matrix, E. The index e_{ij} represents the number of unique molecular identifiers for genes i in cell j (see Figure 5). The matrix begins with all 0s and for each read the corresponding index is increased by 1.

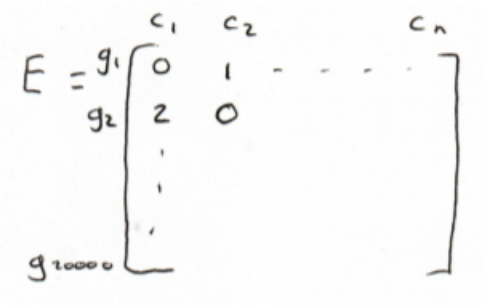

Figure 4.5: The expression matrix output. g represents gene number. c represents cell number.

4.2 Additional Measurement Technologies

Additional measurement technologies used in the fields of developmental biology and stem cell engineering, among others, are discussed below.

4.2.1 Split Pool Indexing

This measurement technology does not isolate individual cells in droplets. Instead, it uses the boundary of the cell itself. For each round:

- The cells are split into n number of pools.
- A specific barcode, B, is added to each pool that binds RNA within the cells.
- The pools are mixed together.

- The process is repeated with new barcodes.

This process is repeated several times to generate a large number of unique barcodes. The method is effective for experiments with less than 100 pools.

4.2.2 Spatial Transcriptomics

In this method, thin slices of tissue are placed on a glass slide with a 256x256 matrix of RNA hairs, similar to the hairs on the single cell sequencing beads (see Figure 6). This technology is currently not at single cell resolution.

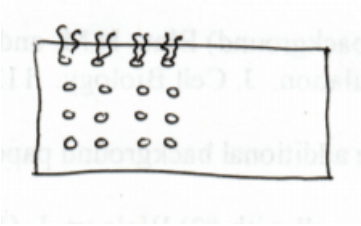

Figure 4.6: Glass slide set up used in spatial transcriptomics.

4.2.3 Measuring Protein Expression

Measuring protein expression allows the user to directly identify functional pathways currently employed by the cell. This may be preferential to avoid assumptions regarding mRNA transcription within the cell.

4.2.3.1 Ex. CyToF

In this method the cells in the tissue are initially stained with antibodies attached to heavy metal ions (see Figure 7).

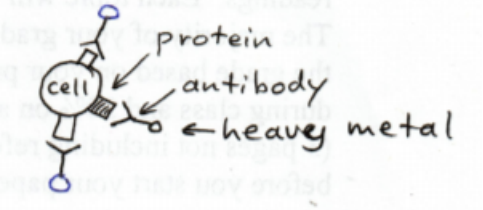

Figure 4.7: CyTof cell staining diagram.

The tissue is then vapourized and the detached metal ions are accelerated through a magnetic field. The resulting differential force causes the ions to hit the detector in different positions based on the charge and mass. This technology is also not currently at single cell resolution.

4.2.4 Measuring Chromatin State

Used to measure the openness, 3S configuration, and co-localization of DNA within the nucleus.

4.2.4.1 Ex. ATAC-Seq

This technique involves sending an enzyme into the nucleus to cut segments of the DNA. However, the enzyme is unable to cut tightly packed sections (see Figure 8, on next page).

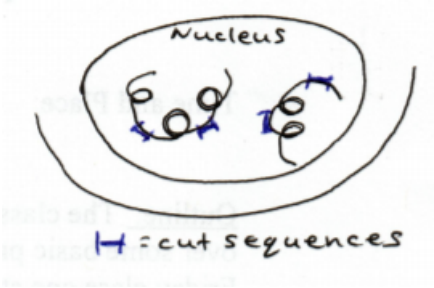

Figure 4.8: ATAC-Seq enzyme digestion.

The DNA cuts are then sequenced and aligned to a reference genome. The resulting output is a set of locations where the enzyme attacked for each cell, corresponding to open sections of the DNA.

4.2.4.2 Hi-C

Hi-C is a chromosome conformation capture technique which connects adjacent DNA strands within a certain radius (see Figure 9).

The linear DNA strand outputs are then sequenced and aligned to a reference genome. The adjoined DNA strands correspond to strand found in the same physical location. A matrix of the number of times you have a collision between two regions can then be constructed for downstream analysis.

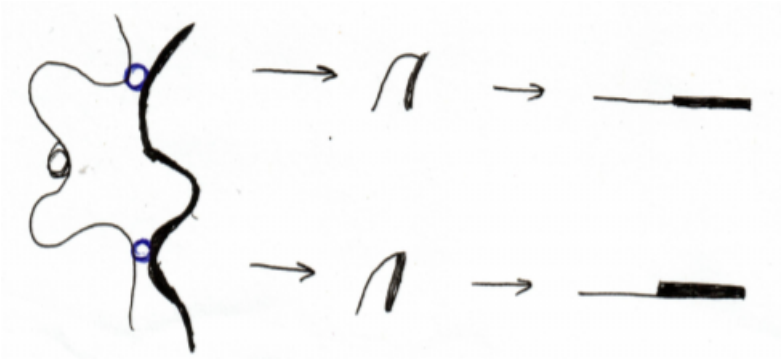

Figure 4.9: Workflow for Hi-C.