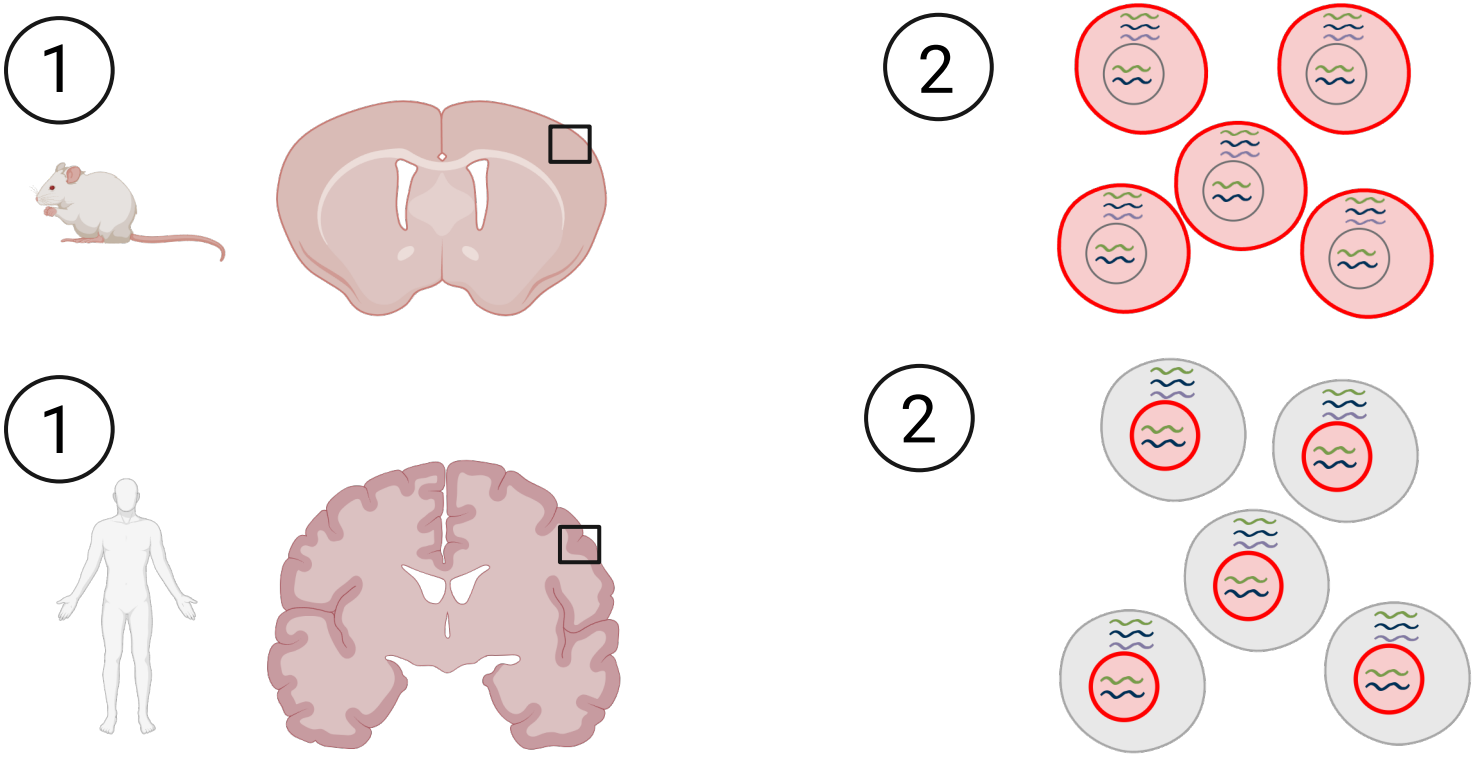
How does 10x single cell RNA-sequencing work?

This document was created to introduce Allen Brain Cell (ABC) Atlas users to the single cell RNA-sequencing methods (10x Genomics) used in the atlas. This document was created using BioRender.

Cell/Nucleus Collection



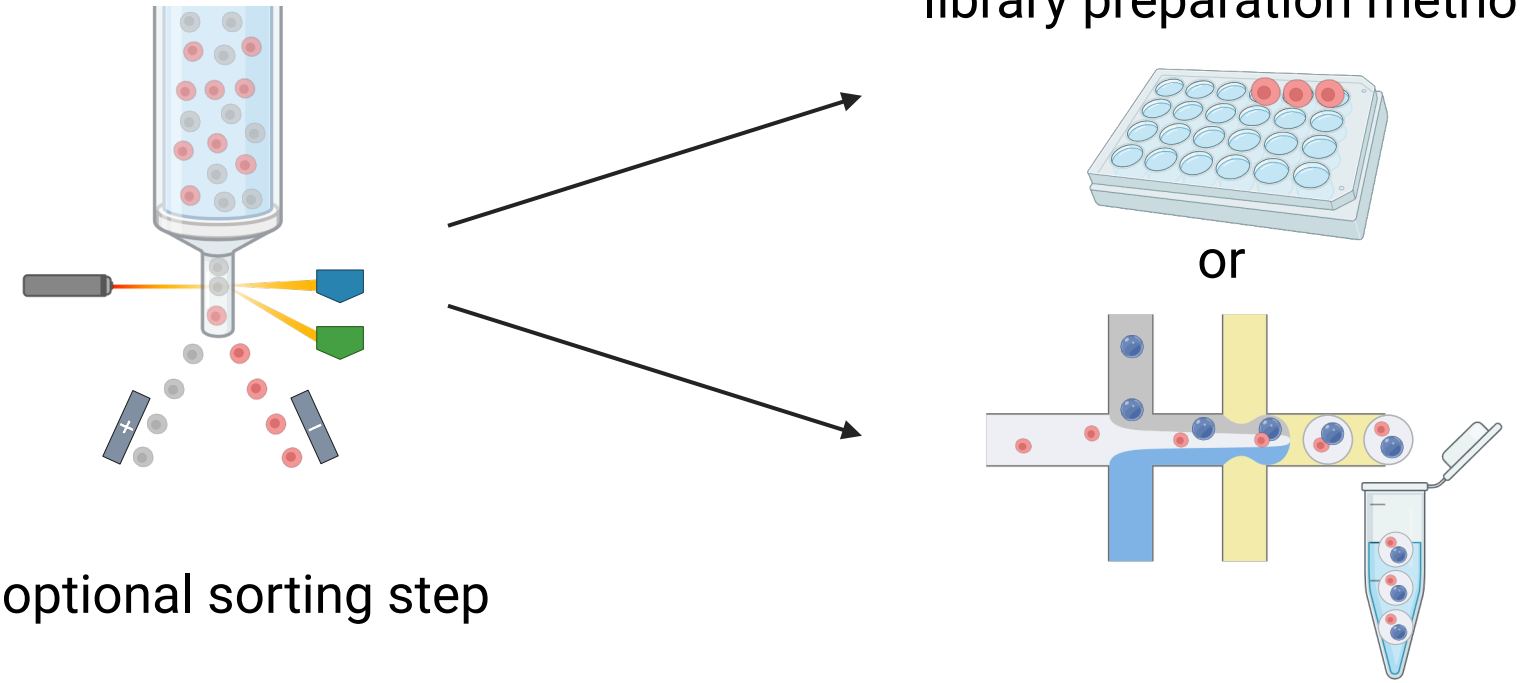
1) Obtain sample from tissue. Mouse is shown in the upper panel, while human in shown below.
2) Dissociate cells/nuclei from tissue. Cells are used for the majority of mouse studies at the Allen Institute, while nuclei are used for human and non-human primate studies (cells vs. nuclei highlighted in red).

In general, single cell RNAseq: can be done with fresh tissue, and captures more transcripts (cytosolic RNA) and more mature transcripts (fewer introns). Single nucleus RNAseq captures less transcripts (nuclear RNA) and less mature transcripts (more introns), can be done with fresh or flash frozen tissue, and captures RNA from cells that are harder to dissociate.

Cell/Nucleus Sorting

ALLEN INSTITUTE for BRAIN SCIENCE

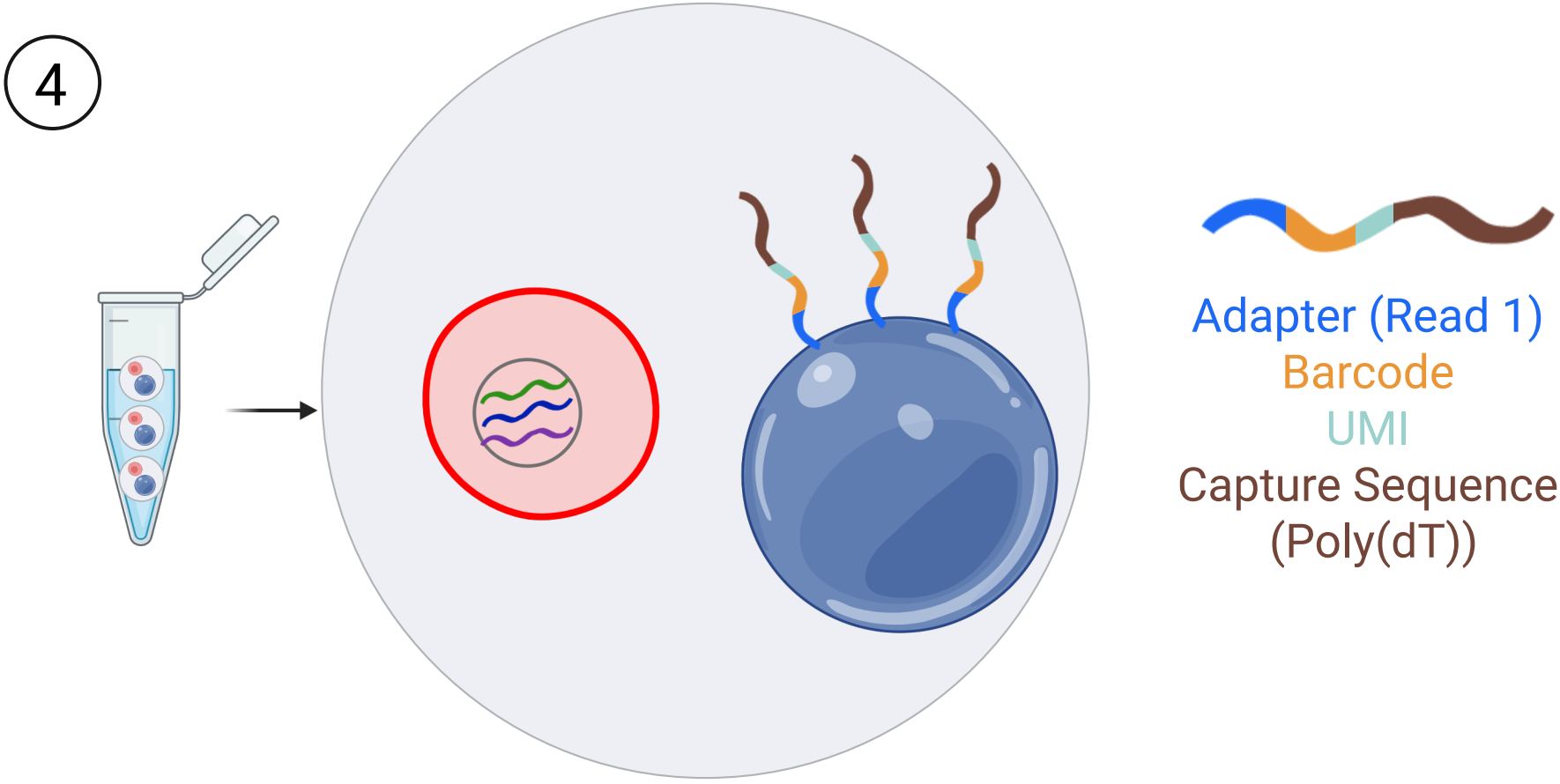
3



3) Once cells/nuclei are dissociated, they need to be sorted and/or isolated (sorting is optional depending on the experiment). Fluorescence Activated Cell Sorting (FACS) (on the left) is often used for sorting, which sorts for cells/nuclei with fluorescent labels (TdTomato, antibodies (NeuN, Oligo2), DAPI, etc.), and/or based on size. Some RNA-seq library preparations use well plates to isolate cells (upper right), while others use oil droplets (lower right). There are many variations of droplet-based and plate-based library prep methods, but for this document, 10x Genomics (droplet-based) methods will be presented.

Library Preparation (10x droplet-based)

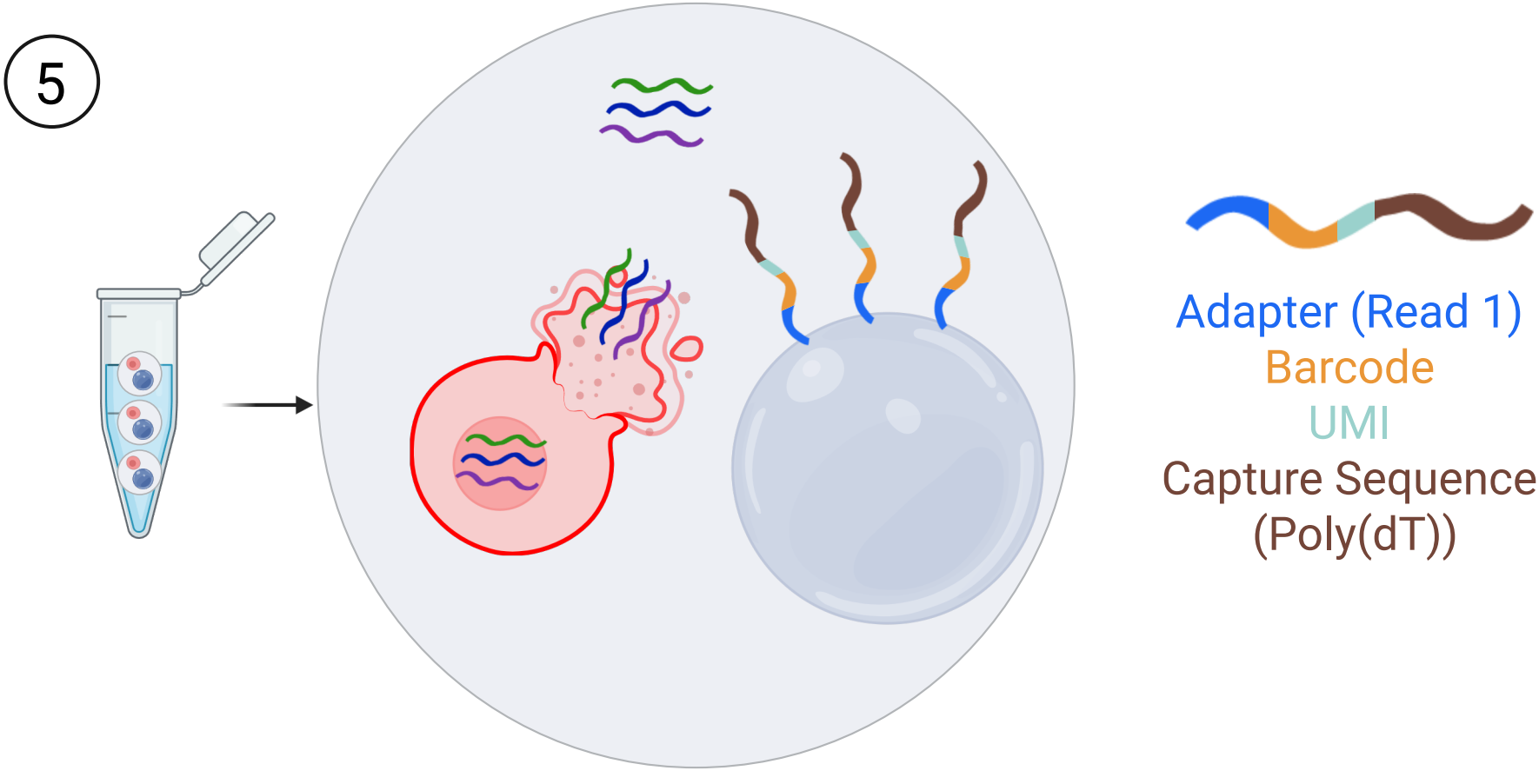
ALLEN INSTITUTE for BRAIN SCIENCE



4) For droplet-based (10x) sequencing, individual cells and gel beads are inserted into oil droplets. Gel beads contain oligonucleotides which have an adapter, a barcode, a unique molecular identity (UMI), and a capture sequence. Each bead has a unique barcode, while each oligonucleotide has a unique UMI. Our example is focused on mRNA and the capture sequence contains deoxy-thymidine to capture Poly(A) tails. Thousands of oil droplets are together in the same tube.

Library Preparation (10x droplet-based)

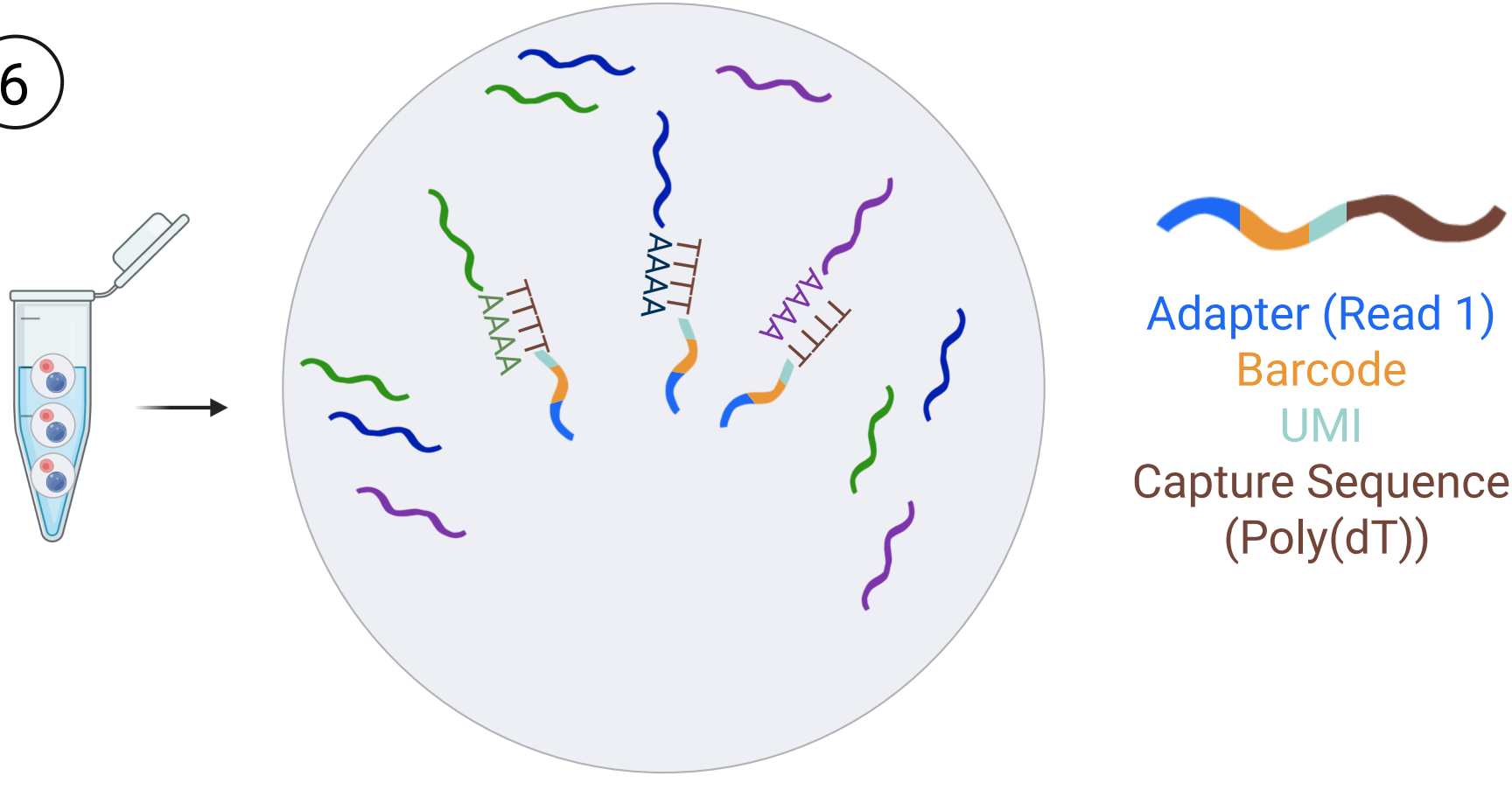
ALLEN INSTITUTE for BRAIN SCIENCE



5) In each oil droplet, lysis buffer lyses the cells and allows for RNA to be released into the oil droplet. During this time, the gel bead dissolves and releases the oligonucleotides.

Library Preparation (10x droplet-based)

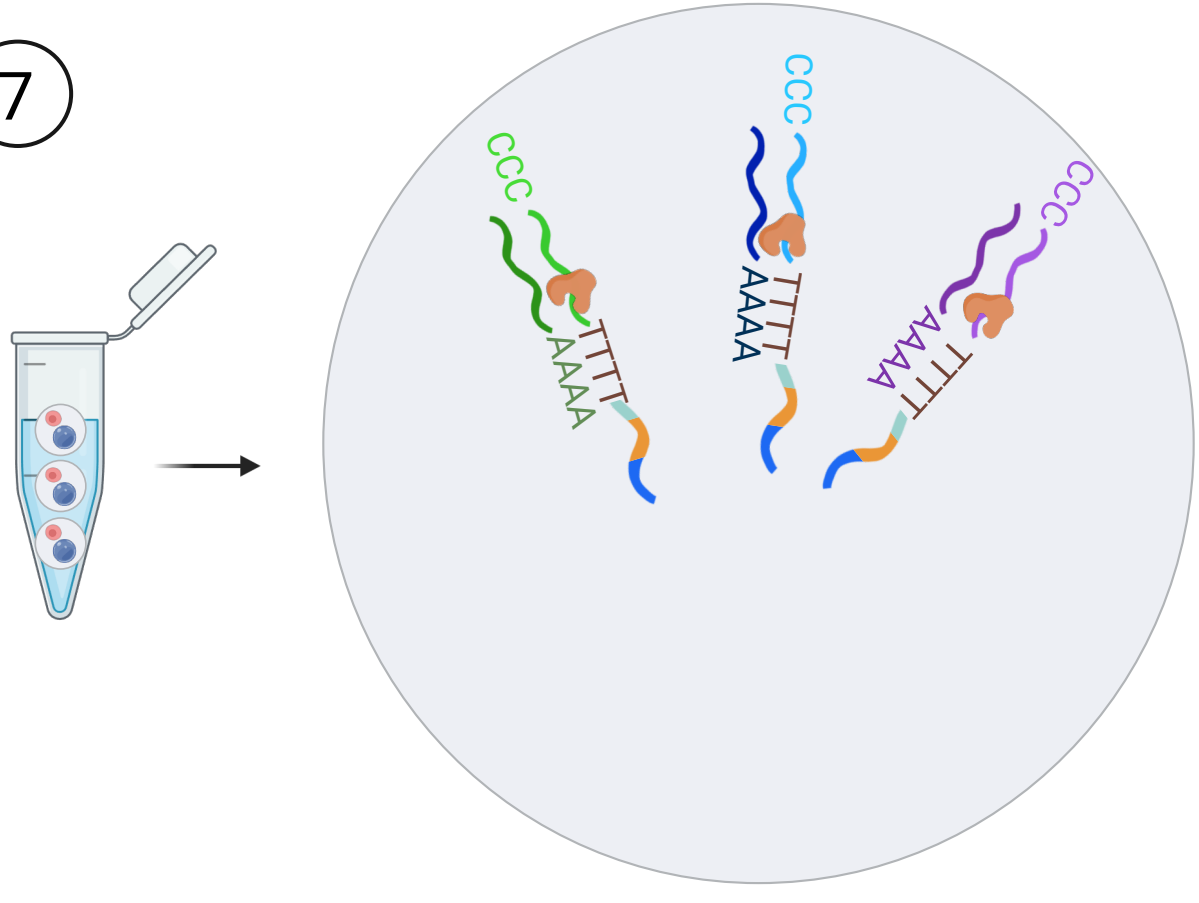
ALLEN INSTITUTE for BRAIN SCIENCE



6) Once the oligonucleotides and RNAs are released into the oil droplet, the Poly(dT) capture sequence on the oligonucleotides hybridizes to the poly(A) tails on the 3' ends of the mRNA. This ensures that only mRNA and not other RNAs (rRNA, tRNA, etc.) are captured for sequencing.

Library Preparation (10x droplet-based)

ALLEN INSTITUTE for BRAIN SCIENCE



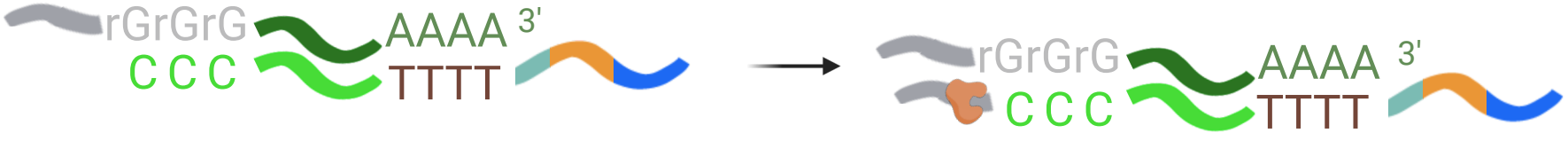
Adapter (Read 1)
Barcode
UMI
Capture Sequence (Poly(dT))
cDNA

7) After hybridizing to the Poly(A) tails, reverse transcriptase (RT) is used to generate cDNA complimentary to the mRNA. At the end of the sequence, the RT adds three cytosine nucleotides (these are used in the next step).

Library Preparation (10x droplet-based)

ALLEN INSTITUTE for BRAIN SCIENCE

8



Adapter (Read 1) Barcode UMI
Capture Sequence cDNA TSO mRNA

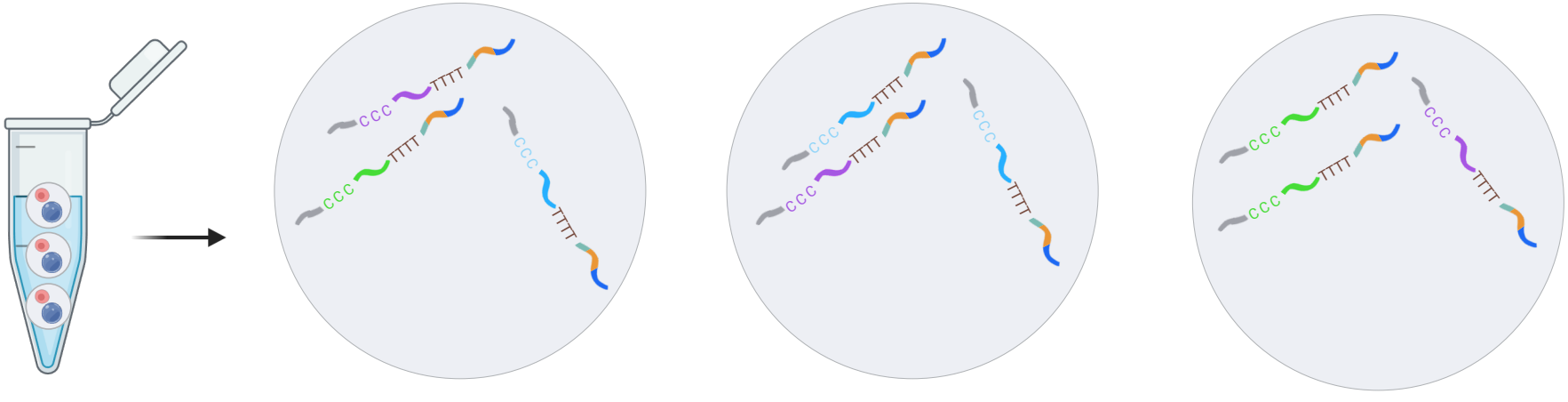
8) Next, a template switch oligo (TSO) containing three riboguanosine nucleotides (rG) hybridizes to the three cytosines. This allows RT to switch templates and reserve transcribe the sequence following the rGs.

Library Preparation (10x droplet-based)

ALLEN INSTITUTE for BRAIN SCIENCE

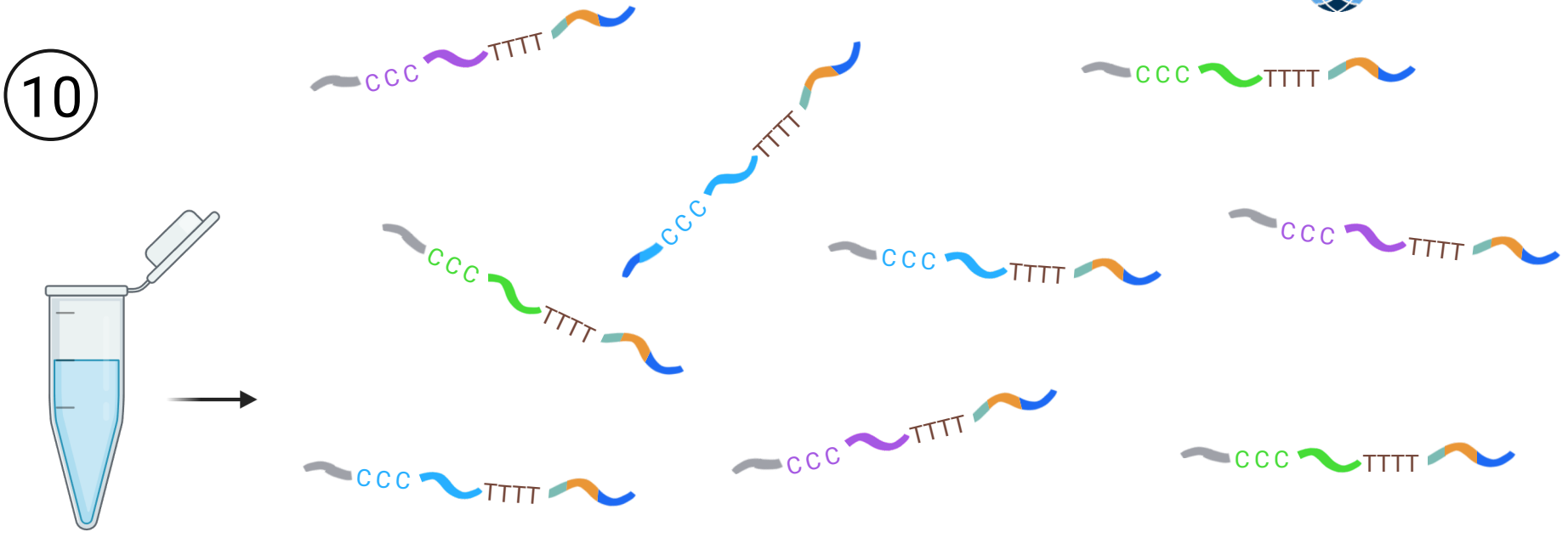
9

Adapter (Read 1) Barcode UMI
Capture Sequence cDNA TSO mRNA



9) This process is happening inside each oil droplet that originally contained both a cell and a gel bead. Not every oil droplet contains both one bead and one cell; many have no cells, while some have two. Those with two cells are identified as "doublets" and are removed during later analysis.

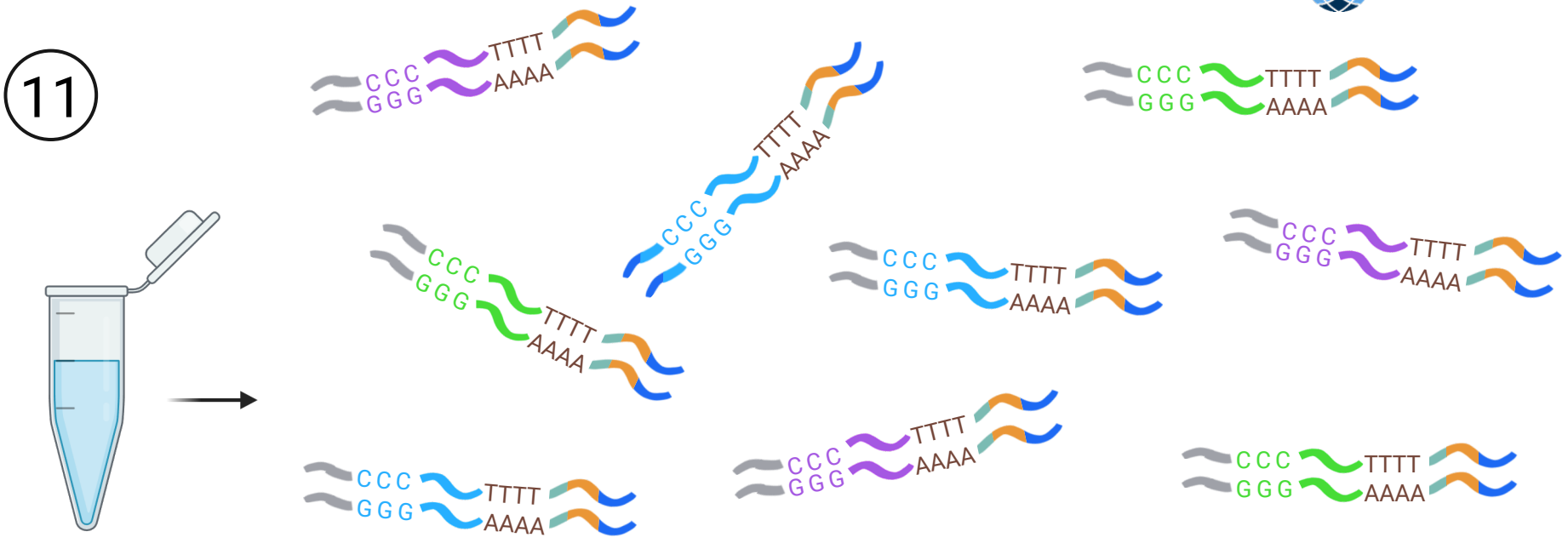
Library Preparation (10x droplet-based)



10) After generating single strands of cDNA, the next step is to remove the oil droplets. After removing oil, the cDNA strands from all cells are pooled together in the tube. The barcode is used to identify the cellular origin of each cDNA, while the UMI is used to identify which mRNA it was reverse transcribed from. The UMI is also important for quantification later on after PCR.

Library Preparation (10x droplet-based)

ALLEN INSTITUTE for BRAIN SCIENCE



Barcode = cell identity

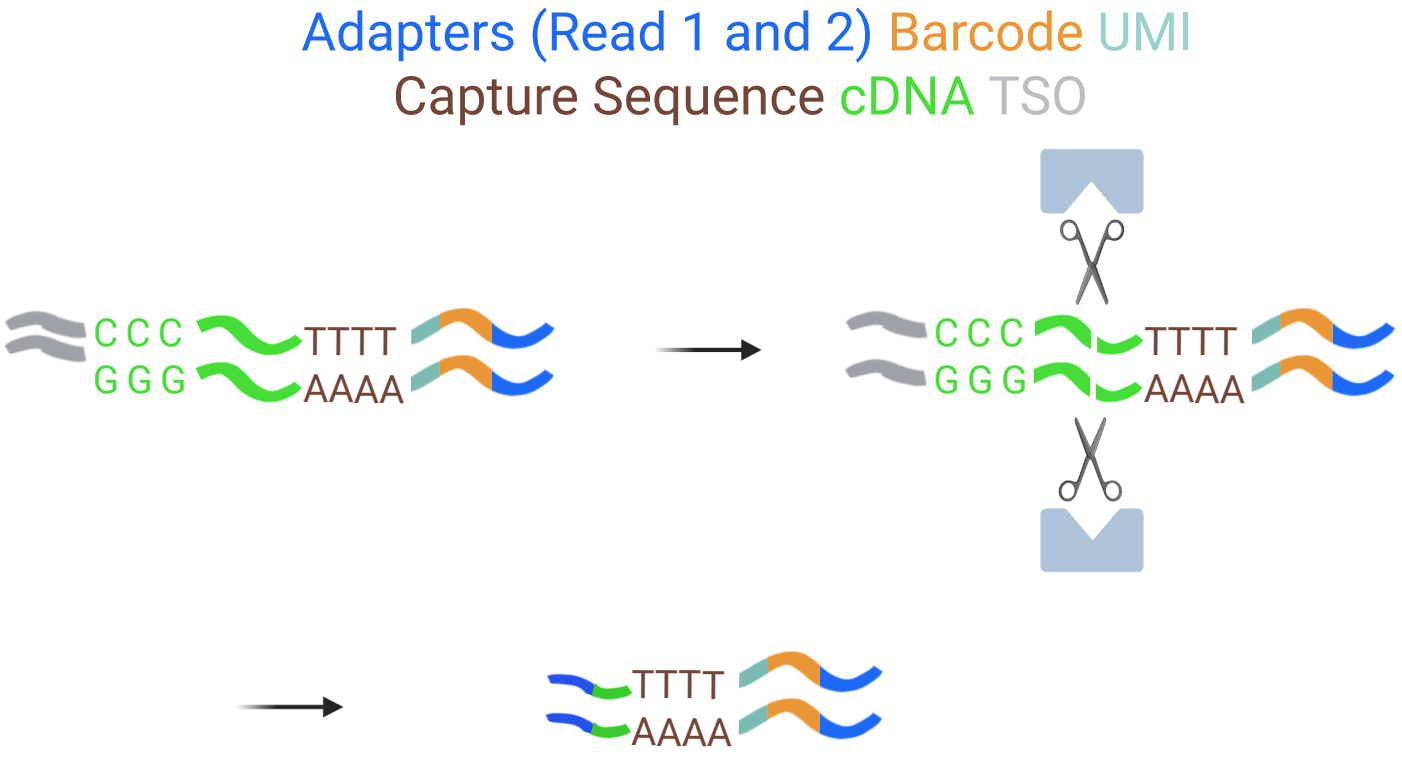
UMI = transcript identity

11) After all single strand cDNAs are pool together, PCR is performed to generate the second strand of cDNA.

Library Preparation (10x droplet-based)

ALLEN INSTITUTE for BRAIN SCIENCE

12



12) After amplifying the full-length cDNA, the next step is enzymatic fragmentation. The protocol highlighted here focuses on the 3' end, while some other protocols use 5'. The cDNA is fragmented so that the final ~200-800 bp from the 3' end remain. Next, an adapter sequence is ligated onto the fragmented end. This adapter will be used for sequencing later and is called "Read 2."

Library Preparation (10x droplet-based)

ALLEN INSTITUTE for BRAIN SCIENCE

13

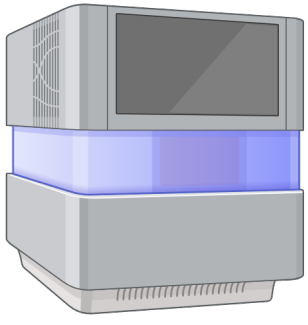
Adapters (Read 1 and 2) Barcode UMI
Capture Sequence cDNA Sample Index
Sequencing Flow Cell Adapters

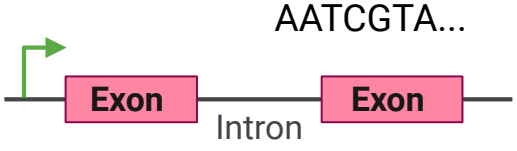
13) The next step involves adding sequencing flow cell adapters and the sample index to create the final sequence. The sequencing flow cell adapters are used for anchoring the construct to the flow cell during the bridge amplification step of sequencing. The sample index is used to track samples when pooling and sequencing multiple experiments together. The Read 1 and Read 2 adapter sequences signal where sequencing starts. Read 1 is used for reading the UMI and barcode, while Read 2 is used to read the cDNA. **After this step, the library is ready to be sequenced.**

Sequencing

ALLEN INSTITUTE for BRAIN SCIENCE

14





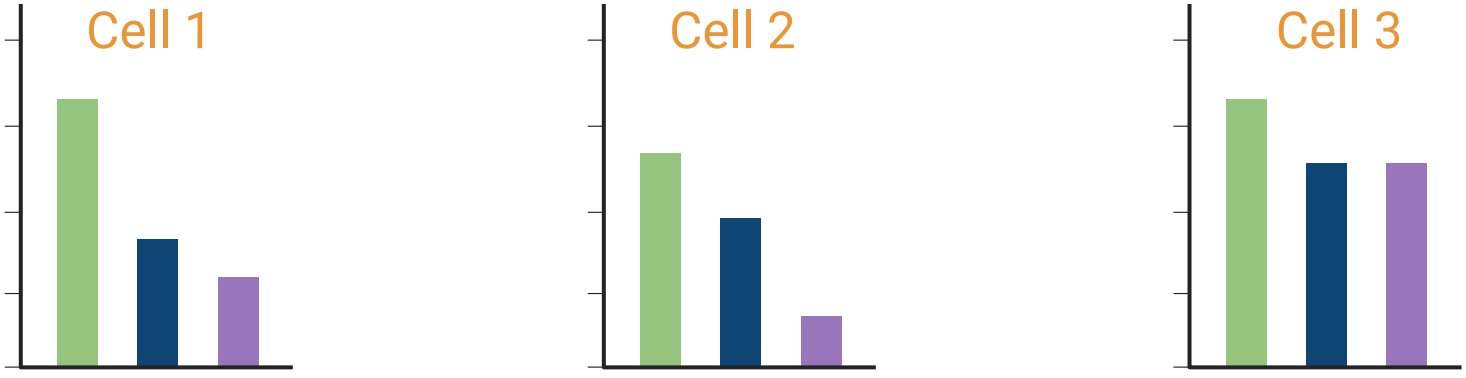
Adapters (Read 1 and 2) Barcode
UMI Capture Sequence cDNA
Sample Index Sequencing Primers

14) Once a cDNA library is generated, the next step is to sequence it. After sequencing, the transcripts are aligned to the reference transcriptome or genome.

Quantification and Quality Control

ALLEN INSTITUTE for BRAIN SCIENCE

15



	gene X	gene Y	gene Z
Cell 1	6	3	2
Cell 2	4	3	1
Cell 3	6	4	4

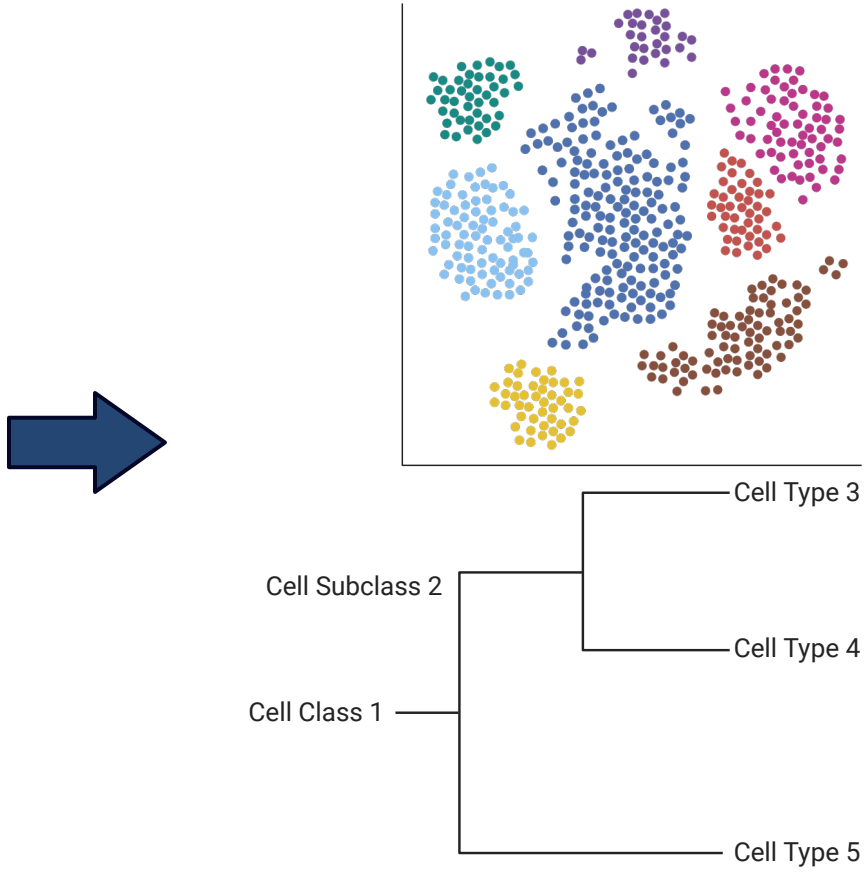
15) Thanks to the barcodes and the UMIs, we can quantify the number of individual transcripts per cell. During this step, doublets (where two cells were in one oil droplet) are identified and removed, in addition to partial cells.

Clustering

ALLEN INSTITUTE for BRAIN SCIENCE

16

	gene X	gene Y	gene Z
Cell 1	6	3	2
Cell 2	4	3	1
Cell 3	6	4	4



16) After quality control, the single cell RNA-seq data can be clustered into cell types based on their transcript counts. From this clustering, we can also identify which cell types are more or less similar to each other and create a cell type taxonomy.