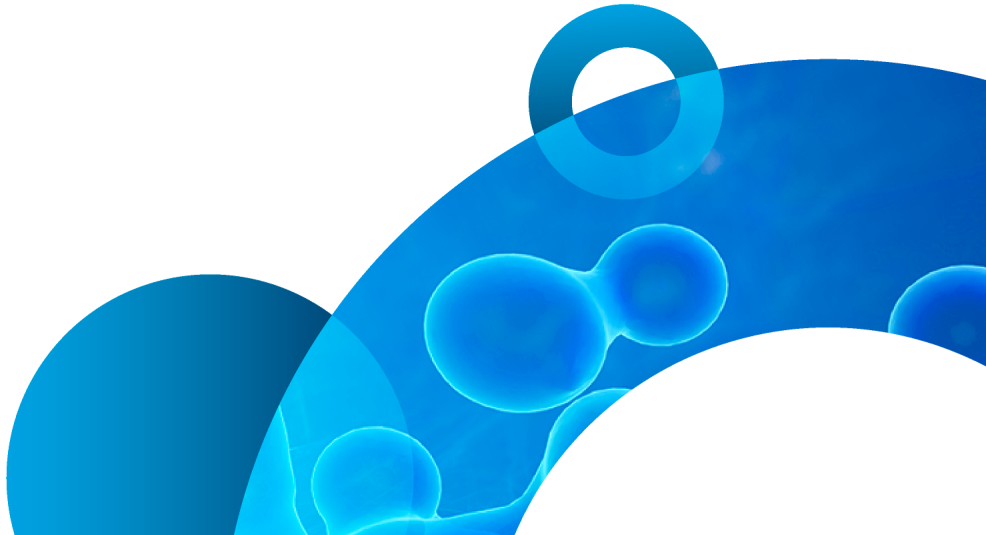
Chromium Single Cell Immune Profiling

Software Training



CG000096 RevH

Objectives

- Overview of how to analyze Chromium Immune Profiling data
- Review key steps in Cell Ranger V(D)J and Gene Expression pipelines
- Overview of key output files generated by the pipelines
- Review the run metrics and know what to look for to quickly gauge the quality of the run

10x GENOMICS

Outline

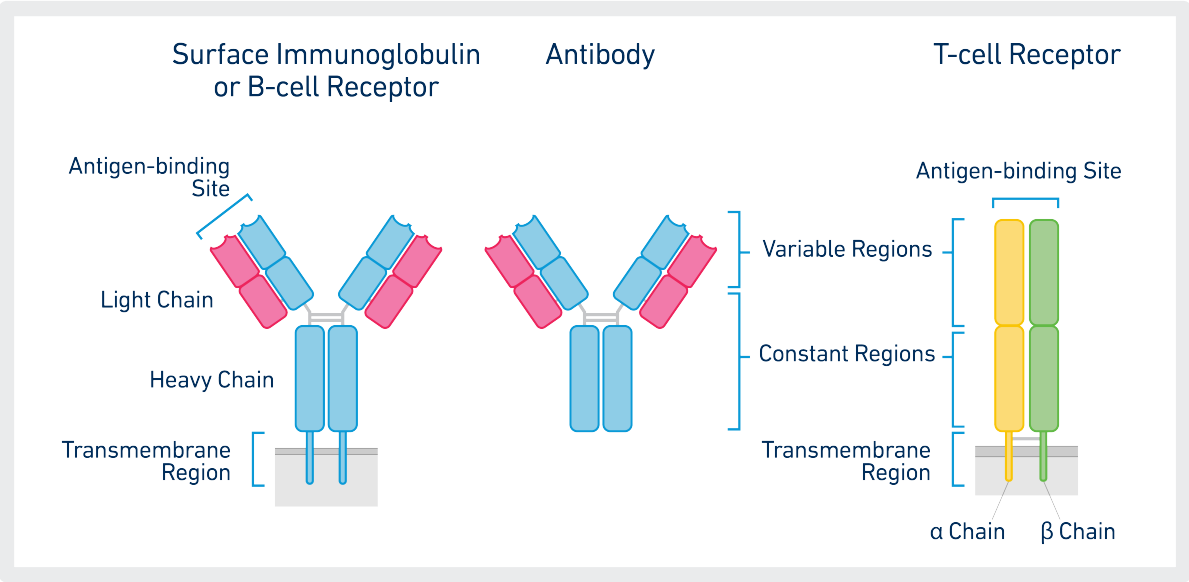
- Workflow review
- Cell Ranger overview
- Demultiplexing with “mkfastq”
- Cell Ranger “vdj” pipeline for Repertoire analysis
- Cell Ranger “multi” pipeline for combined VDJ and Gene Expression analysis
- Cell Ranger “aggr” for combining VDJ data
- Gene Expression
  - Cell Ranger “count” Pipeline for Gene Expression and/or Feature Barcoding analysis
  - Gene Expression and Feature Barcoding analysis outputs
  - Gene Expression Analysis and Feature Barcoding QC metrics

10x GENOMICS

Key components of the adaptive immune system

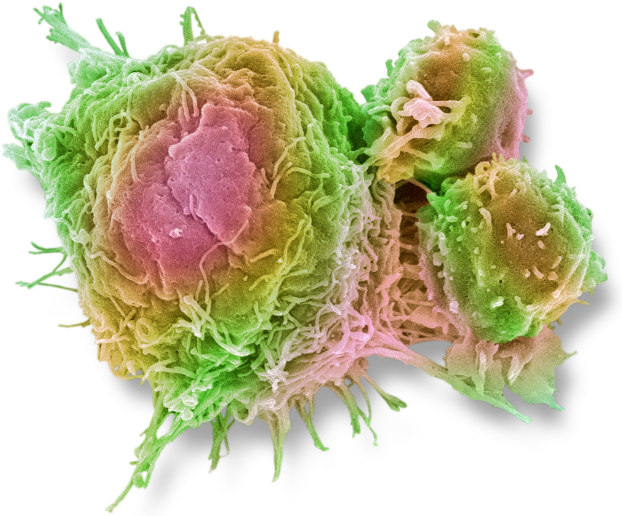
Diverse immune receptors provide broad surveillance

Determinants of antigen specificity



- In both T and B cells specificity is determined by two distally encoded, co-expressed genes
- Diversity is generated by V(D)J recombination and somatic hypermutation

T cells attacking a tumor cell



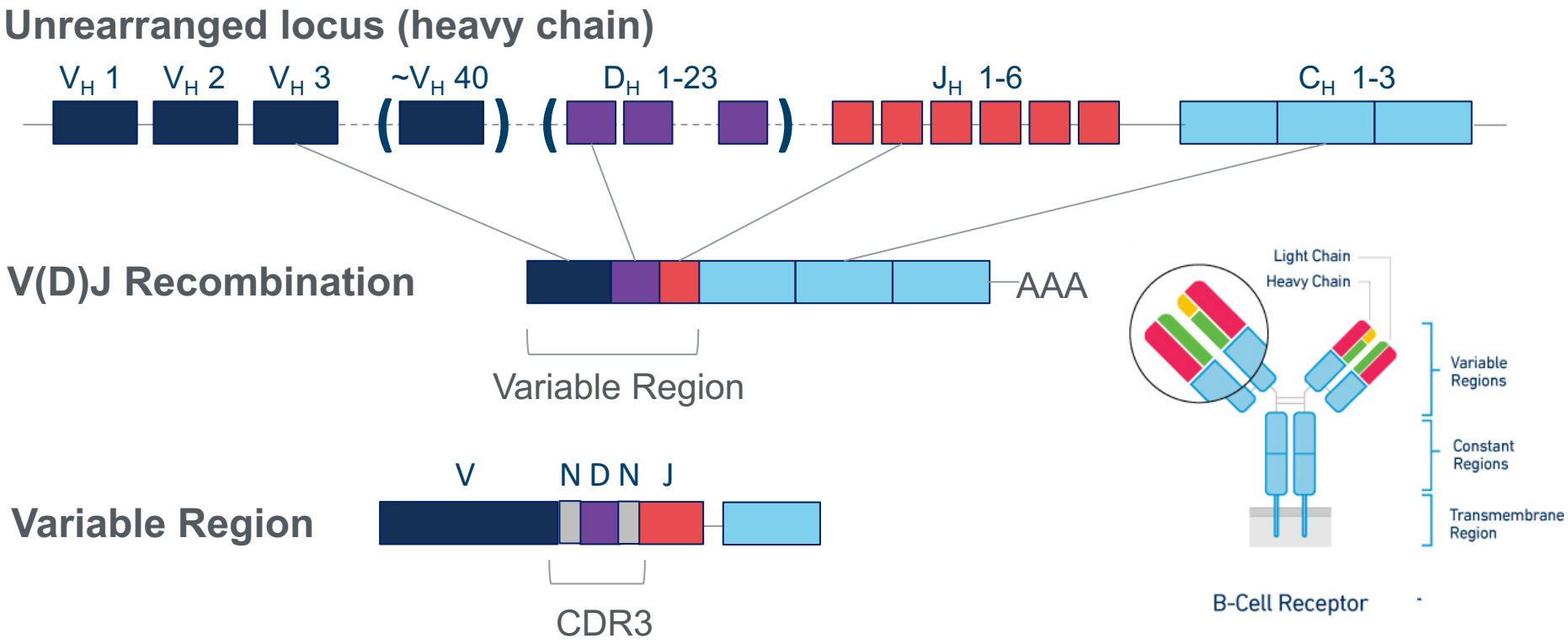
- Paired characterization of the TCR α and β transcripts in each T cell is critical to dissecting cellular interactions

10x GENOMICS

V(D)J recombination generates immune repertoire diversity

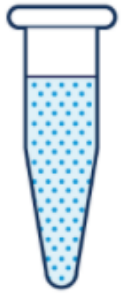
Enormous diversity of T and B cell antigen-specific receptors

- Diversity is generated by V(D)J recombination + N nucleotide addition or deletion + Somatic hypermutation
- Full-length sequencing of the paired heavy and light chains (B cells) or α and β chains (T cells) is critical to dissecting these interactions

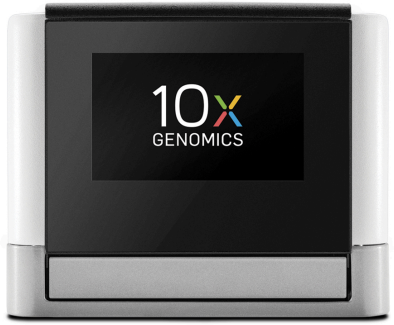


Chromium Single Cell Immune Profiling Solution

Input | Library creation | Sequence | Data analysis & visualization

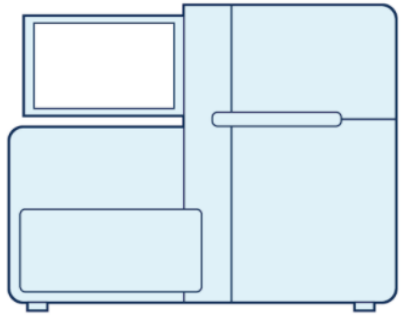


User-supplied cells

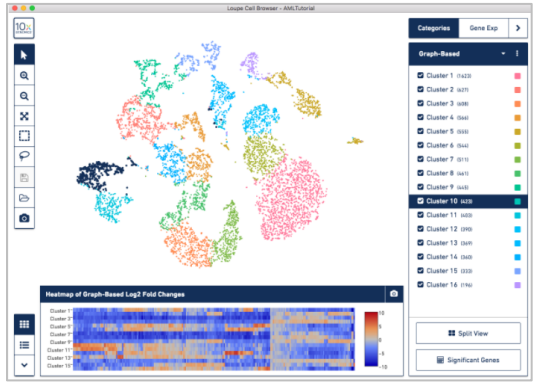


Chromium Controller and Single Cell Immune Profiling Reagents

- Cell partitioning and lysis
- Barcoding and library creation
- T cell or B cell Ig enrichment
- Human or mouse species
- Feature Barcode Library Kit



Illumina Sequencer compatibility



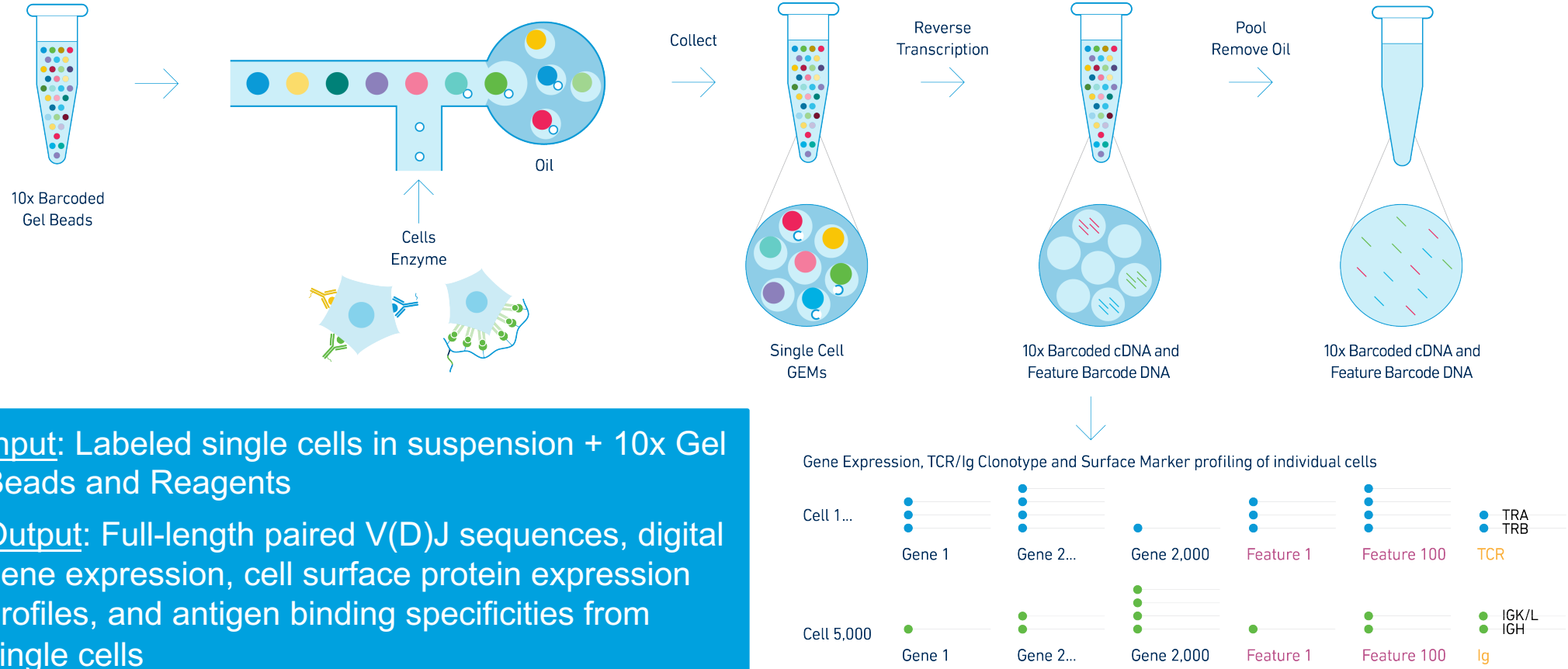
Cell Ranger Pipeline
Loupe Browser
Loupe V(D)J Browser

10X GENOMICS

Workflow review

Single Cell Immune Profiling solution with Feature Barcode technology

Workflow overview

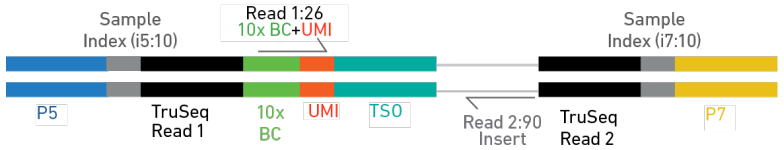


Input: Labeled single cells in suspension + 10x Gel Beads and Reagents

Output: Full-length paired V(D)J sequences, digital gene expression, cell surface protein expression profiles, and antigen binding specificities from single cells

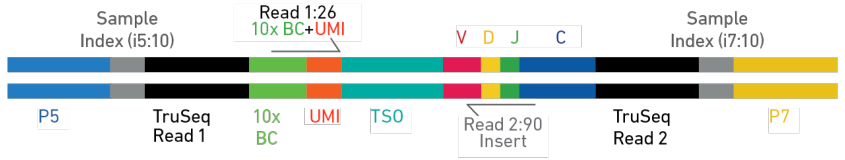
Sequencing

Chromium Single Cell 5’ Gene Expression Library



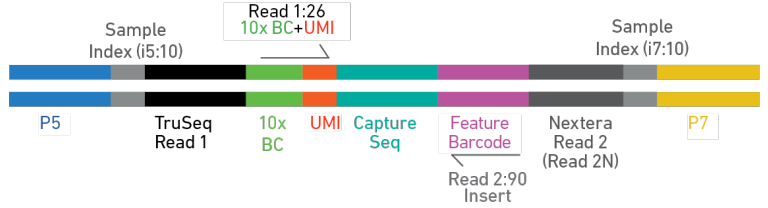
Recommended 20,000 read pairs per cell

Chromium Single Cell VDJ Amplified Library



Recommended 5,000 read pairs per cell

Chromium Single Cell 5’ Cell Surface Protein Library



Recommended 5,000 read pairs per cell

	Read 1	i7 Index	i5 Index	Read 2
Purpose	10x Barcode and UMI	Sample index	Sample index	Insert (Transcript or Feature Barcode sequence)
Length	26	10	10	90

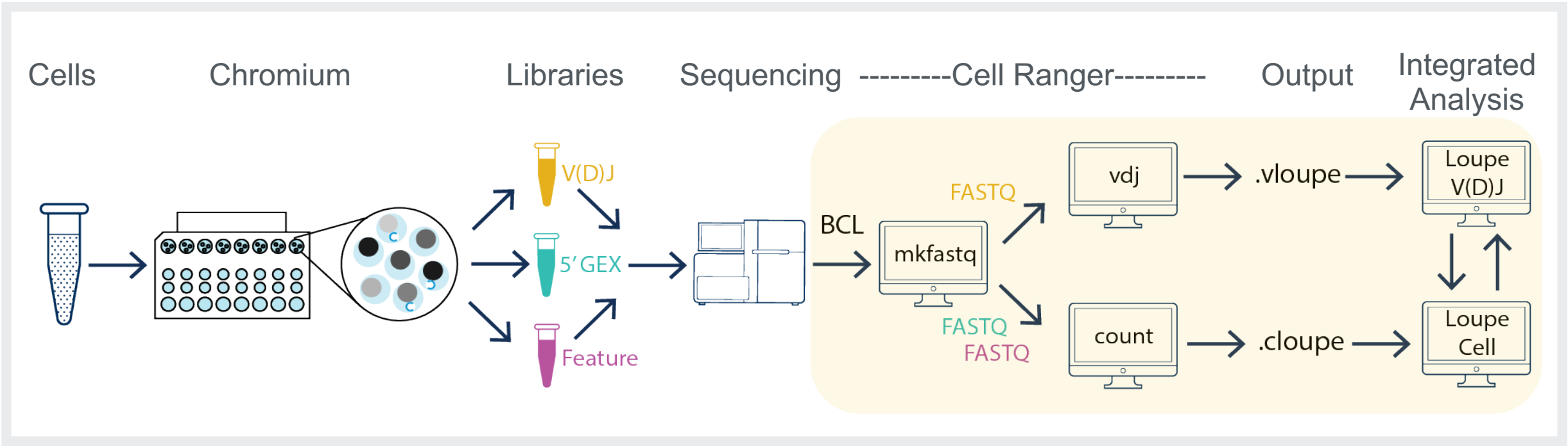
10x GENOMICS

Cell Ranger overview

Single Cell Immune Profiling

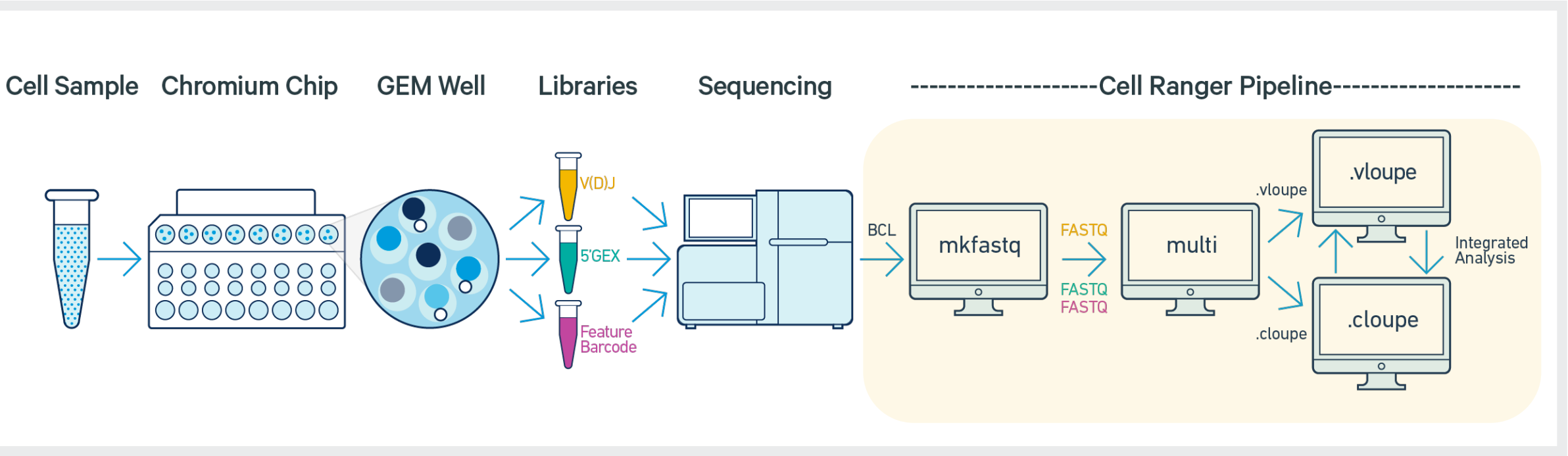
Analyzing immune profiling data

Overview



Analyzing immune profiling data

Overview: analyzing V(D)J with gene



Laboratory

Analysis

10x GENOMICS

Cell Ranger

Introduction

- A set of analysis pipelines that process Chromium Single Cell Immune Profiling data
- Capable of analyzing V(D)J, Gene Expression, and Feature Barcode data
- Contains various pipelines for:
  - Demultiplexing (*mkfastq*)
  - Single sample V(D)J assembly and annotation (*vdj*)
  - Single sample V(D)J with gene expression and/or feature barcoding analysis (*multi*)
  - Single sample gene expression and/or feature barcode analysis (*count*)
  - Combining data from multiple samples into an experiment-wide analysis (*aggr*)
  - Reanalyzing GEX and feature barcode data with custom parameters (*reanalyze*)
- Easy to download and run on Linux

10x GENOMICS

Cell Ranger

System requirements

Local Mode

- Run on single, standalone Linux system
- 64-bit CentOS/RedHat 6.0 or Ubuntu 12.04
- 8 cores, 64GB RAM (minimum)

Cluster Mode

- Run on SGE and LSF
- Each node must have 8+ cores and 8GB+ RAM/core
- Shared filesystem between nodes (e.g. NFS)

Runtime

- VD)J
  - 1200 cells, 10k read pairs/cell
  - ~1 hours wall-time with 16 cores, 64 GB RAM
- GEX only
  - ~8000 cells, 35k reads/cell
  - ~4 hours wall-time with 16 cores, 64 GB RAM
- Antibody only
  - ~8000 cells, 16k reads/cell
  - ~1 hours wall-time at 16 cores, 64 GB RAM

10x GENOMICS

Cell Ranger

Easy to download and install

- Download Cell Ranger and 10x Genomics reference data from:
  http://support.10xgenomics.com
- Simple installation:

```
$ tar -xzvf cellranger-x.y.x.tar.gz
$ tar -xzvf refdata-cellranger-x.y.z.tar.gz
$ export PATH=/opt/cellranger-x.y.z:$PATH
```

No Make. No Compile.

One Dependency. Illumina's bcl2fastq installed and added to PATH variable

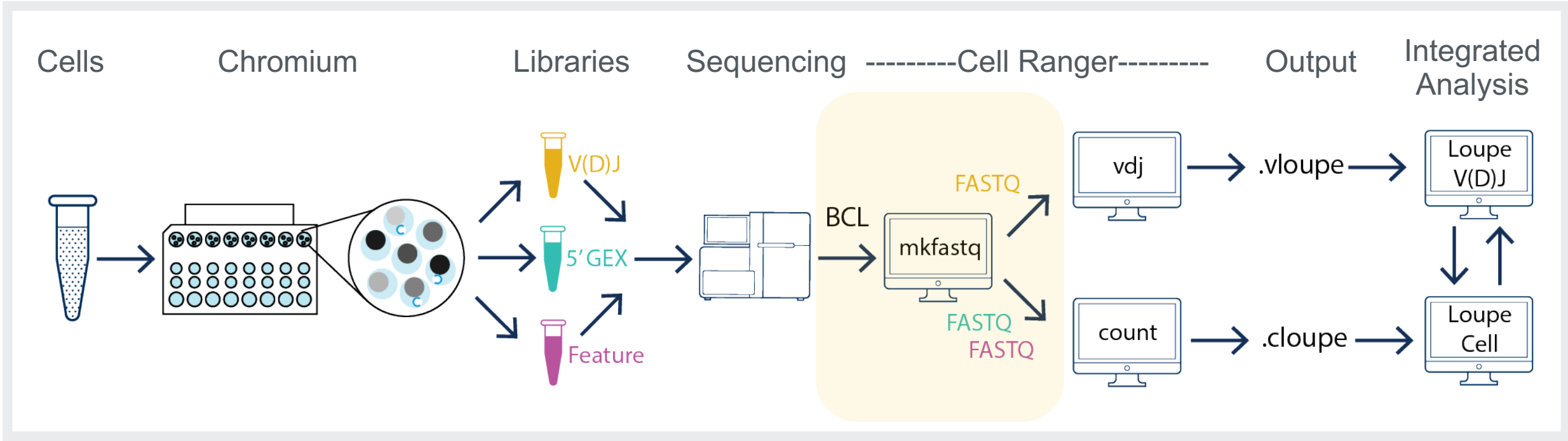
10x GENOMICS

Running Cell Ranger

Demultiplexing and V(D)J pipelines

Analyzing Immune Profiling data

Overview



Executing `cellranger mkfastq`

Demultiplexing

- Run the `cellranger mkfastq` pipeline to demultiplex an Illumina sequencing run folder into FASTQ files

```
$cellranger mkfastq --id=run1 \
                    --csv=my_samples.csv \
                    --run=/mnt/hiseq/run1_bcl
```

Output directory name

Sample sheet

Path to bcls

- Output

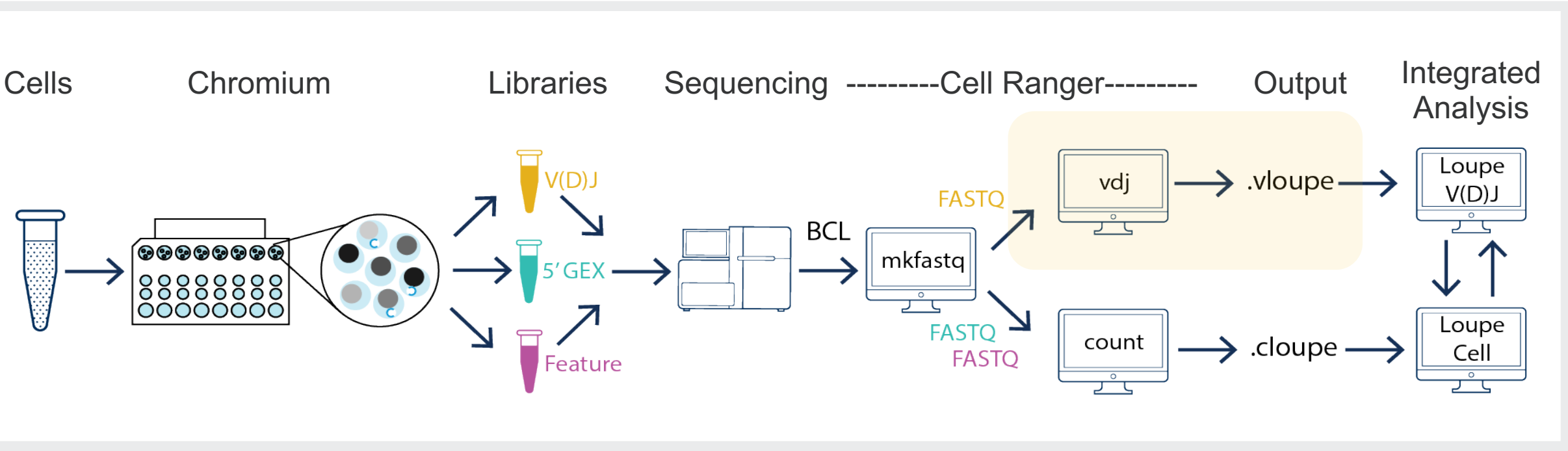
```
$ls -l /opt/bar/run1/outs/fastq_path
drwxr-xr-x 3 jdoe jdoe           3 Sep 24 19:19 HAWT7ADXX
drwxr-xr-x 3 jdoe jdoe           3 Sep 24 19:17 Reports
drwxr-xr-x 2 jdoe jdoe          10 Sep 24 19:17 Stats
-rw-r--r-- 1 jdoe jdoe    27971547 Sep 24 18:07 Undetermined_S0_L001_I1_001.fastq.gz
-rw-r--r-- 1 jdoe jdoe    44432222 Sep 24 18:07 Undetermined_S0_L001_R1_001.fastq.gz
-rw-r--r-- 1 jdoe jdoe   912086076 Sep 24 18:07 Undetermined_S0_L001_R2_001.fastq.gz

$tree opt/bar/run1/outs/fastq_path/HAWT7ADXX/sample1/
opt/bar/sample1/outs/fastq_path/HAWT7ADXX/sample1
|── sample1_S1_L001_I1_001.fastq.gz
|── sample1_S1_L001_R1_001.fastq.gz
|── sample1_S1_L002_R2_001.fastq.gz
```

10x GENOMICS

Analyzing Immune Profiling data

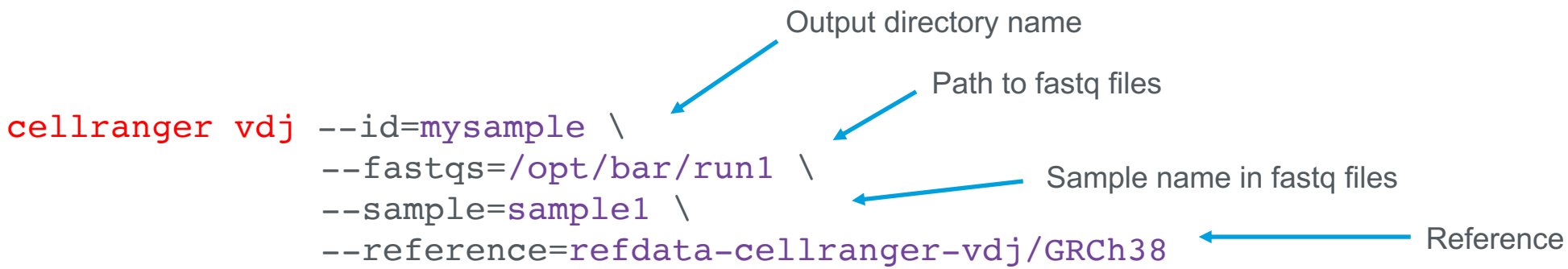
Overview



Executing “cellranger vdj”

- This pipeline performs contig assembly, and annotation for a sample

- Typical command line



Cell Ranger “vdj” pipeline: Reference

- 10x Genomics pre-built references
  - Human, mouse with 10x Genomics-curated V(D)J annotations
  - Download from here: https://support.10xgenomics.com/single-cell-vdj/software/downloads/latest

- Custom references
  - `cellranger mkvdjref` utility generates a 10x V(D)J reference package from any FASTA and GTF gene file
  - Can also create IMGT based custom reference
  - More information available here: https://support.10xgenomics.com/single-cell-vdj/software/pipelines/latest/advanced/references

Cell Ranger “vdj” pipeline algorithms

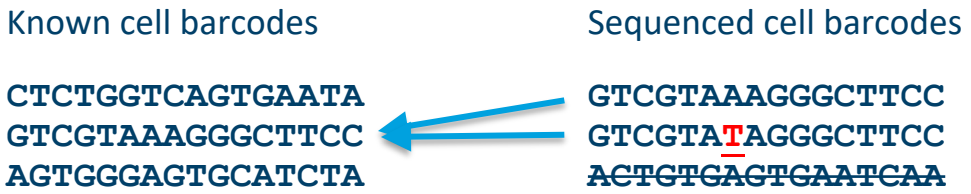
Pipeline steps

Correct barcodes, UMIs

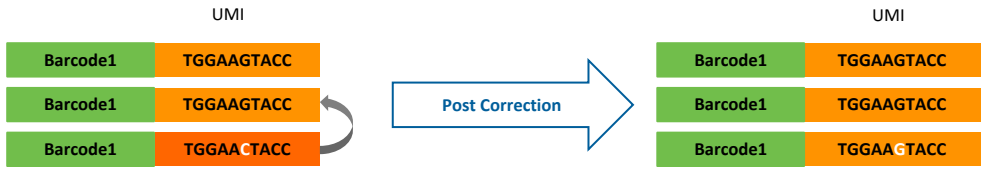
- Correct Barcodes, UMIs
- Trim Reads
- De novo Assembly
- V(D)J Annotation
- Filter Contigs
- Call Cells
- Generate Clonotypes

- Goal: Filter out reads with noisy barcodes and UMI that may arise from PCR errors, sequencing errors etc.

- Cell barcode filtering
  - Must be on list of ~737k known barcode sequences
  - May be 1 mismatch away from the list if the mismatch occurs at a low-quality position (the barcode is then corrected)



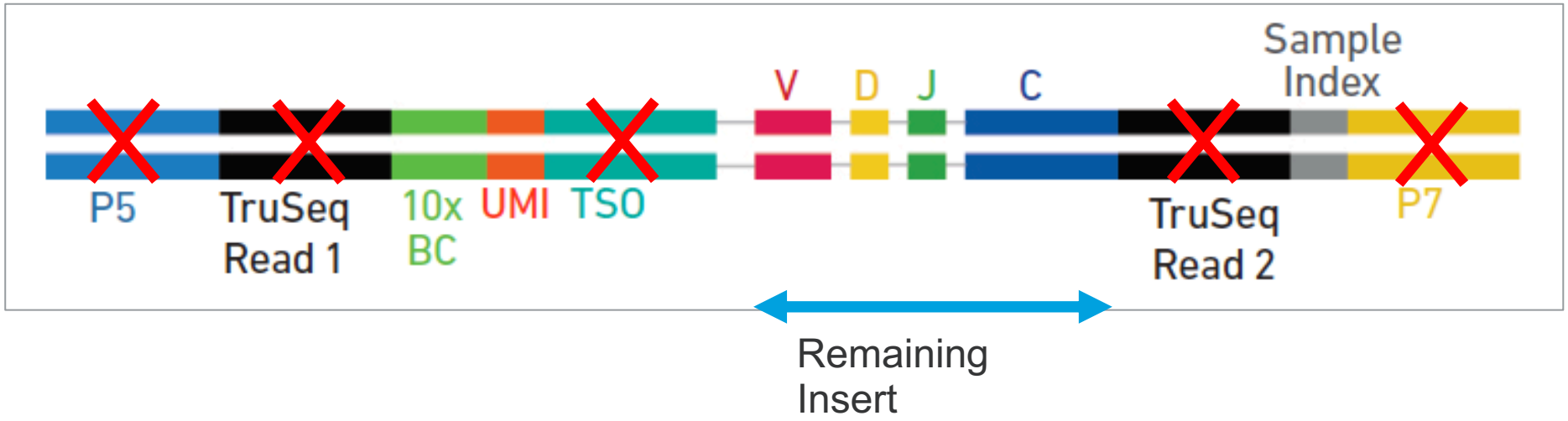
- UMI correction
  - UMIs that are 1 mismatch away from a higher-count UMI are corrected to that UMI if they share a cell barcode



Trim reads

- Correct Barcodes, UMIs
- Trim Reads
- De novo Assembly
- V(D)J Annotation
- Filter Contigs
- Call Cells
- Generate Clonotypes

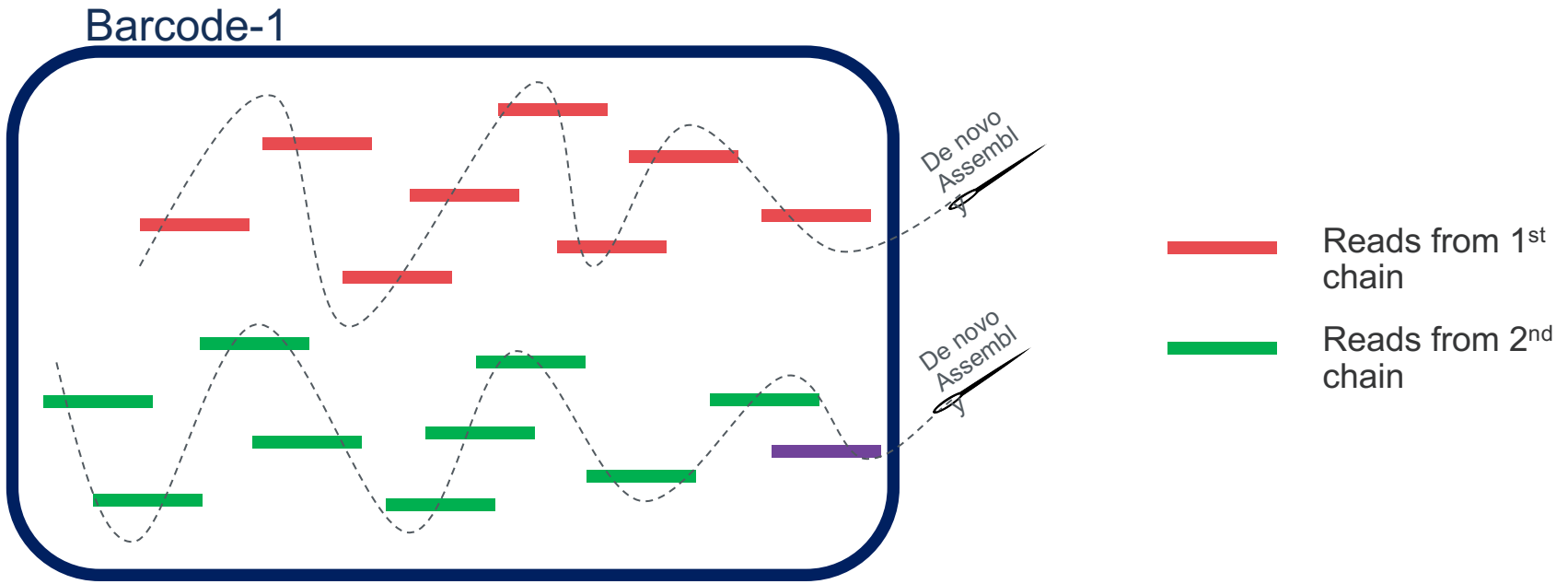
- Trim known adapter and primer sequences from the 5’ and 3’ ends of reads using the `cutadapt` tool
- This tool uses Smith-Waterman alignment and allows for a small number of differences from the expected primer sequences



De novo assembly

- Correct Barcodes, UMIs
- Trim Reads
- **De novo Assembly**
- V(D)J Annotation
- Filter Contigs
- Call Cells
- Generate Clonotypes

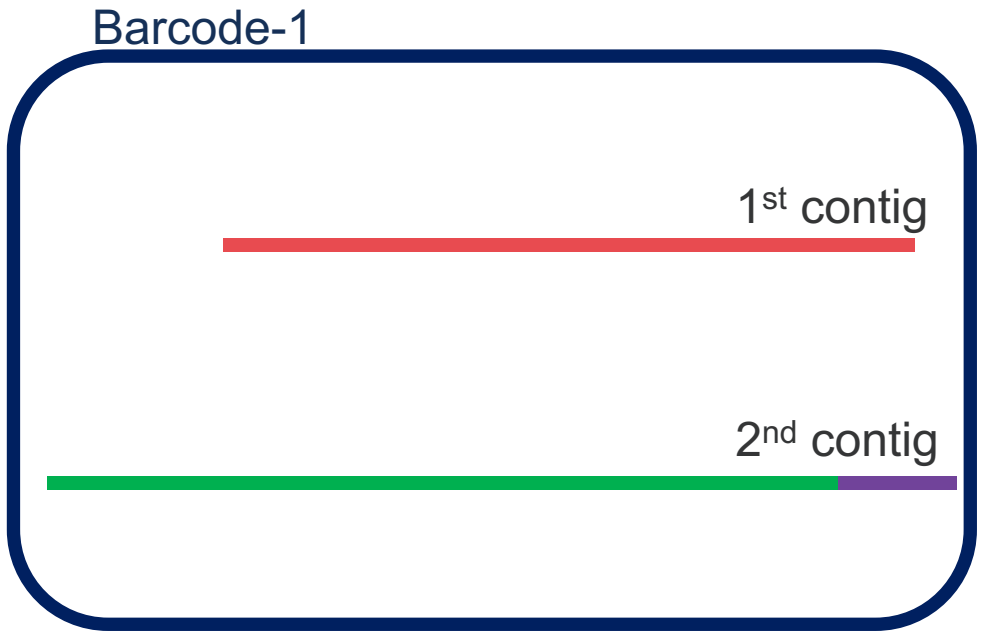
- Goal: Generate full length transcripts (contig) for each chain in all observed GEMs/barcodes
- De novo assembly is used to stitch the reads into contigs for each barcode



De novo assembly

- Correct Barcodes, UMIs
- Trim Reads
- De novo Assembly
- V(D)J Annotation
- Filter Contigs
- Call Cells
- Generate Clonotypes

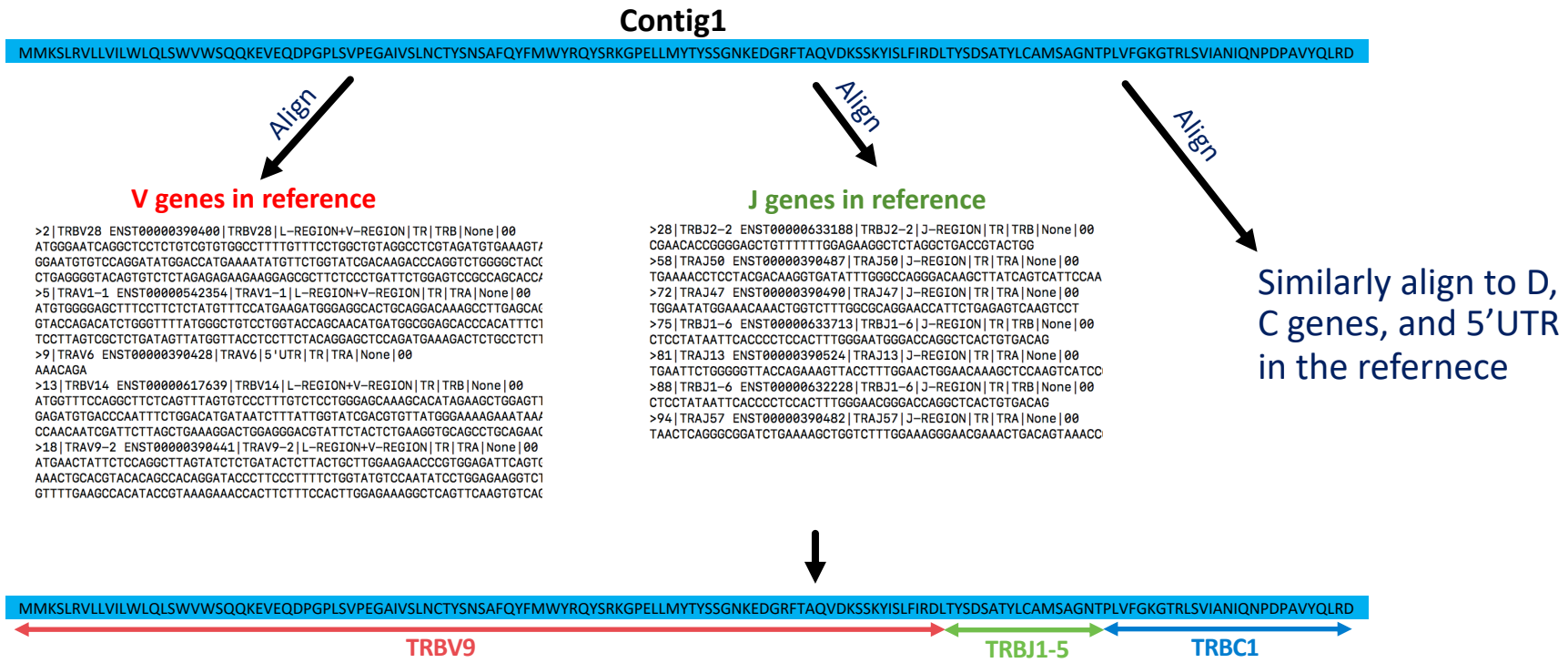
- Goal: Generate full length transcripts (contig) for each chain in all observed GEMs/barcodes
- De novo assembly is used to stitch the reads into contigs for each barcode



V(D)J annotation

- Correct Barcodes, UMIs
- Trim Reads
- De novo Assembly
- V(D)J Annotation
- Filter Contigs
- Call Cells
- Generate Clonotypes

- Goal: Annotate the contig with VDJ segments that form the transcript



Similarly align to D, C genes, and 5'UTR in the refernece

- To annotate:
  - Each assembled contig is aligned against all germline genes in the reference
  - The genes with best alignments to each segment are used

Cell Ranger “vdj” pipeline: V(D)J annotation

- Correct Barcodes, UMIs
- Trim Reads
- De novo Assembly
- V(D)J Annotation
- Filter Contigs
- Call Cells
- Generate Clonotypes

- CDR3 region: one of the Complementarity Determining Regions
  - It is the main hypervariable region responsible for recognizing the antigen
  - Located at the intersection of V,J region for TRA, Light chains (Ig), and V,D,J for TRB, Heavy chain (Ig)

- To identify CDR3 region
  - Pipeline searches for conserved V and J segment motifs that flank a CDR3 as below
  - CDR3 must start with C, be 5-27 amino acids in length and not have stop codon
  - In cases where more than one putative CDR3s are found, the one with best score is selected

Cell Ranger “vdj” pipeline: V(D)J annotation

Correct Barcodes, UMIs

Trim Reads

De novo Assembly

V(D)J Annotation

Filter Contigs

Call Cells

Generate Clonotypes

- Full length:
  - The contig matches the initial part of annotated V gene
  - The contig spans until the 3’ end of J gene
- Productive:
  - Full length
  - Contains a start-codon
  - Contains no stop codons in the V-J spanning region
  - Start codon of V gene and last codon in J gene are in frame
  - Contains a CDR3 region
  - Length of V-J spanning region in contig is within a reasonable distance of the length of the annotated V+J genes

10x GENOMICS

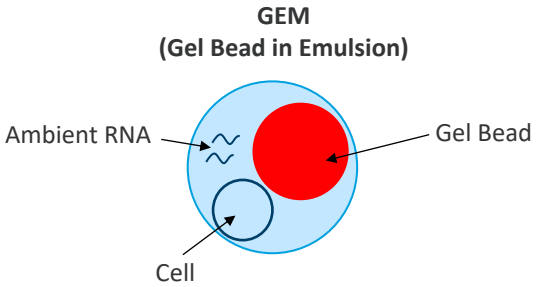
Cell Ranger “vdj” pipeline: Filter contigs

- It is expected that each cell-barcode typically contains
  - For T cells: one productive TRA and one productive TRB chain
  - For B cells: one productive Heavy and one productive Light chain (Kappa or Lambda)
- Extra productive contigs are less likely to be legitimate
  - For example these might arise from ambient mRNA or doublets
  - These suspicious contigs are annotated as low confidence
- Productive contigs are marked as low confidence if
  - They are not biologically likely or they have low UMI support

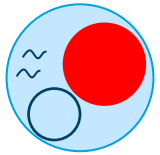
10x GENOMICS

Cell Ranger “count” pipeline: Call cells

- Correct Barcodes, UMIs
- Trim Reads
- De novo Assembly
- V(D)J Annotation
- Filter Contigs
- **Call Cells**
- Generate Clonotypes

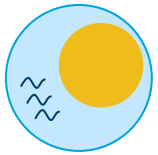


- One Chromium Single Cell channel generates approximately 100,000 GEMs
- What is in the ~100,000 GEMs?
  - A fraction of GEMs contains 1 cell each
    - A smaller fraction of GEMs contain 1 T or B cell (unless sorted)



  - Typically very small number (depending on cell load) contain more than one cells

  - A large fraction of GEMs are considered “empty,” without any cells, but do contain ambient RNA. These are called background barcodes.



- Goal: Distinguish barcodes containing targeted cells from background

10x GENOMICS

Cell Ranger “vdj” Pipeline: Call targeted cells

- Correct Barcodes, UMIs
- Trim Reads
- De novo Assembly
- V(D)J Annotation
- Filter Contigs
- **Call Cells**
- Generate Clonotypes

- Goal: Identify GEMs that contain T or B cells
- Each barcode must satisfy below conditions to be called as a cell
  - Filter barcodes based on presence of productive contigs
    - Only T and B cells produce fully rearranged transcripts. So this filters against other cell types that might express V/D/J genes
  - Filter barcodes based on sufficient UMI support
    - This helps filter against calling cells based on ambient mRNA
  - Filter barcodes based on sufficient read support for UMIs
    - This helps against transcripts seen due to cross-library contamination or sample index-hopping

10x GENOMICS

Cell Ranger “vdj” pipeline: Generate clonotypes

Correct Barcodes, UMIs

Trim Reads

De novo Assembly

V(D)J Annotation

Filter Contigs

Call Cells

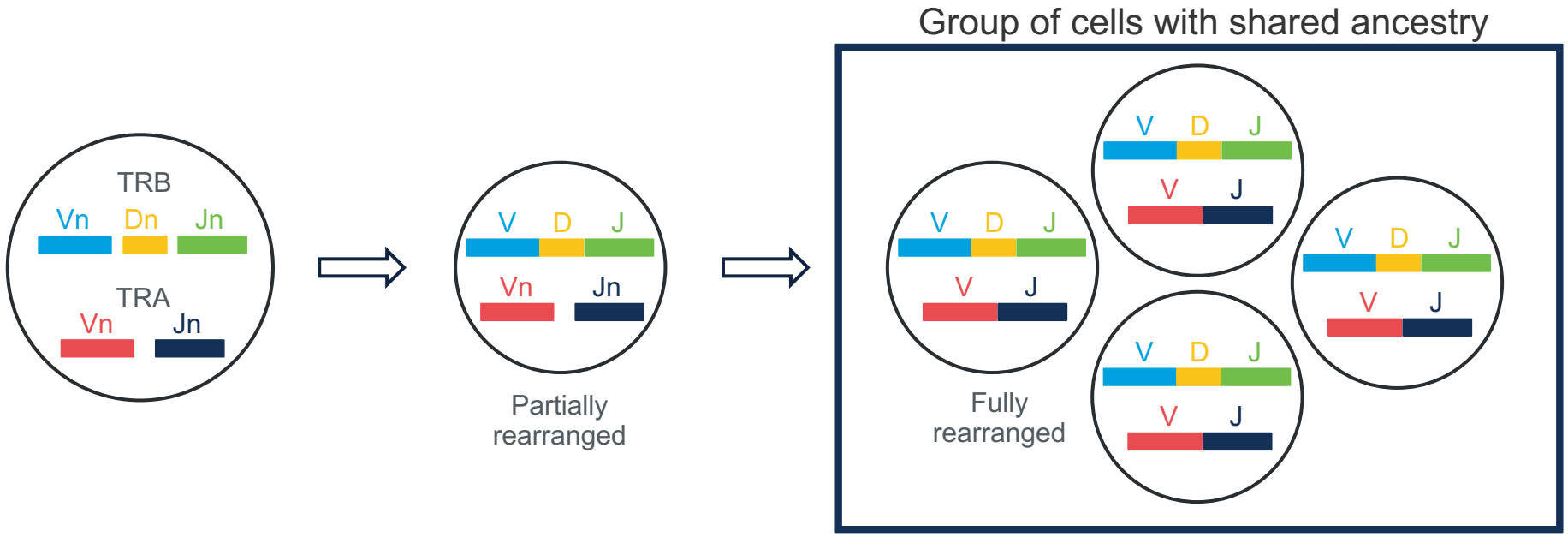
Generate Clonotypes

- Goal: Group all cell associated barcodes into clonotypes
- A clonotype is a group of cells that are have presumably derived from same progenitor cell

10x GENOMICS

What is a Clonotype ?

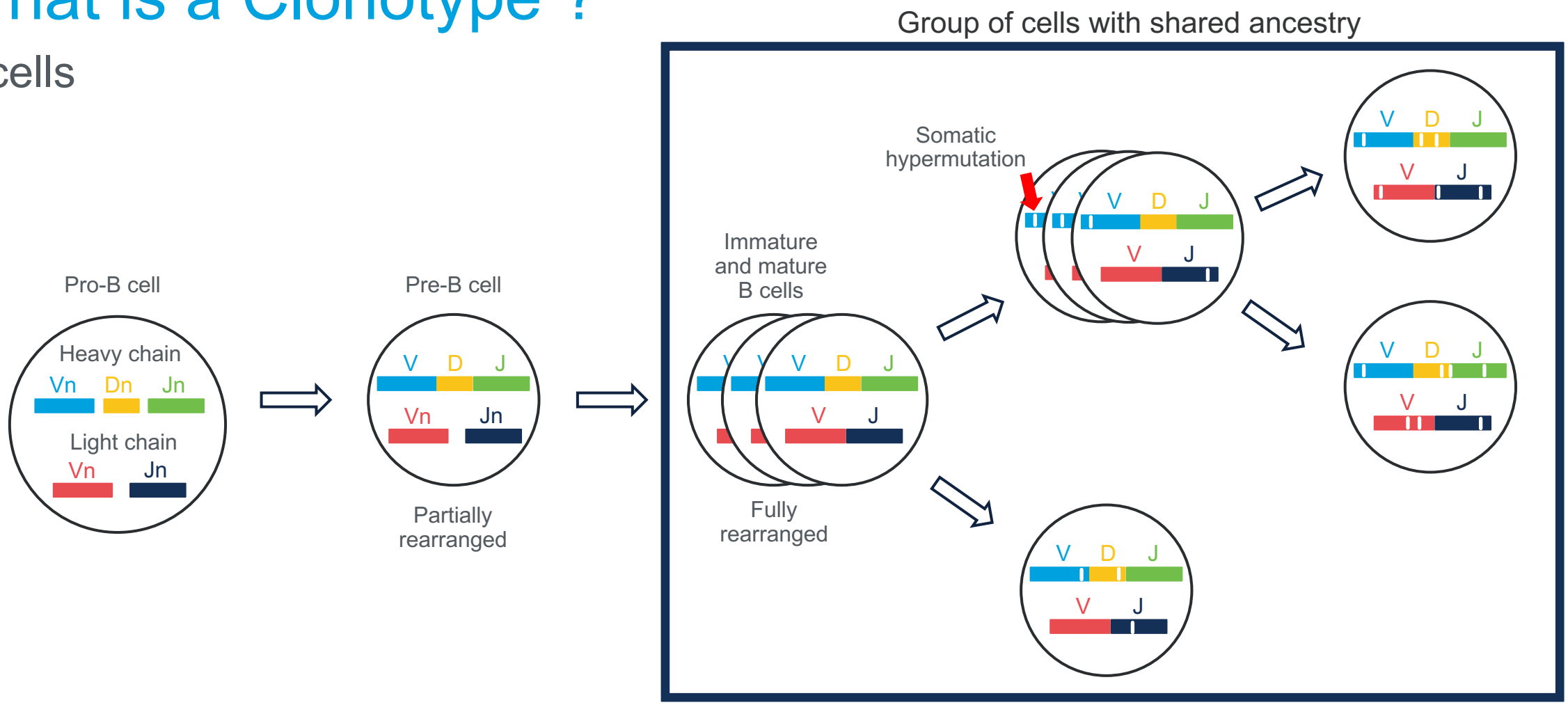
T cells



- The VDJ transcripts in cells in a clonotype are copies of each other
- Cell Ranger v4.0 merges cells based on exact match in CDR3 sequence
- This works okay for T cells

What is a Clonotype ?

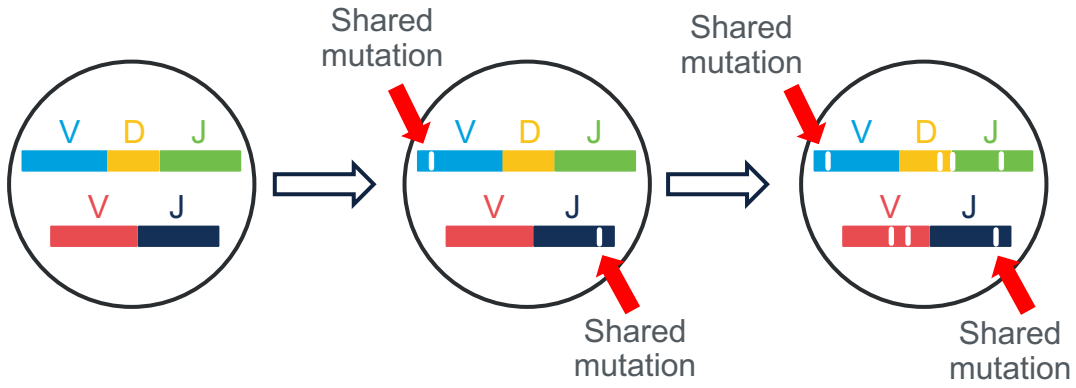
B cells



Grouping Cells into Clonotypes

- Correct Barcodes, UMIs
- Trim Reads
- De novo Assembly
- V(D)J Annotation
- Filter Contigs
- Call Cells
- **Generate Clonotypes**

- Look for evidence of shared ancestry
- Similar CDR3
  - $score_d$ = *dissimilarity between CDR3*
  - Low $score_d$ indicates common ancestry
- Shared somatic mutations



10x GENOMICS

Cell Ranger V(D)J analysis outputs

Outputs

- Correct Barcodes, UMIs
- Trim Reads
- **De novo Assembly**
- **V(D)J Annotation**
- **Filter Contigs**
- Call Cells
- **Generate Clonotypes**

Assembled V(D)J sequences for each cell-barcode in FASTA format

V(D)J annotations (such as V/D/J/C genes, CDR3 sequence etc. for each assembled sequence in CSV format

barcode	is_cell	contig_id	high_confidence	length	chain	v_gene	d_gene	j_gene	c_gene	full_length	productive	cdr3	cdr3_nt	reads
AAACCTGCAATGACCT-1	TRUE	AAACCTG	TRUE	573	TRB	TRBV6-3	TRBD1	TRBJ1-1	TRBC1	TRUE	TRUE	CASSAPRGRL	TGTGCCAGC	6524
AAACCTGCAATGACCT-1	TRUE	AAACCTG	TRUE	551	TRB	TRBV6-3	TRBD1	TRBJ1-1	TRBC1	FALSE	None	CASSAPRGRL	TGTGCCAGC	505
AAACCTGCAATGACCT-1	TRUE	AAACCTG	TRUE	576	TRA	TRAV19	None	TRAJ13	TRAC	TRUE	TRUE	CALSGGGYQ	TGTGCTCTGA	1441
AAACCTGCAATGACCT-1	TRUE	AAACCTG	TRUE	263	None	None	None	None	None	FALSE	None	None	None	442
AAACCTGGTTAAAGAC-1	TRUE	AAACCTG	TRUE	648	TRB	TRBV23-1	TRBD1	TRBJ2-5	TRBC2	FALSE	None	CASSRRCAGA	TGCGCCAGC	
AAACCTGGTTAAAGAC-1	TRUE	AAACCTG	TRUE	504	TRB	TRBV4-1	TRBD2	TRBJ2-5	TRBC2	TRUE	TRUE	CASSQAGYG	TGCGCCAGC	7733
AAACCTGGTTAAAGAC-1	TRUE	AAACCTG	TRUE	567	TRA	TRAV27	None	TRAJ44	TRAC	TRUE	TRUE	CAADAGTAS	TGTGCAGCC	3434
AAACCTGTCGAGAACG-1	TRUE	AAACCTG	TRUE	607	TRB	TRBV7-2	None	TRBJ2-1	TRBC2	TRUE	TRUE	CASSTNGWN	TGTGCCAGC	8581

Clonotypes and their frequencies in CSV format

clonotype_id	frequency	proportion	cdr3s_aa	cdr3s_nt
clonotype1	4	0.00449438	TRA:CAVMDSNYQLIW;TRB:CSAKTGQGEADTQYF	TRA:TGTGCTGTGATGGATA
clonotype9	2	0.00224719	TRB:CASSQESLAGGLTDTQYF	TRB:TGCGCCAGCAGCCAAG
clonotype8	2	0.00224719	TRA:CAVRNQGGKLIF	TRA:TGCGCTGTGAGGAATC
clonotype5	2	0.00224719	TRA:CAPSGYSTLTF;TRA:CAVHAGNNRKLIW;TRB:CASSAAGTGGAGANVLTF	TRA:TGTGCCGTCCATGCTG
clonotype4	2	0.00224719	TRB:CSAKTGQGEADTQYF	TRB:TGCAGTGCTAAAACAG
clonotype7	2	0.00224719	TRA:CAYSLSGNTGKLIF;TRB:CSARQMNTEAFF	TRA:TGTGCTTATAGCCTCTC
clonotype6	2	0.00224719	TRA:CALISGGYQKVTF;TRB:CASSERDSQETQYF	TRA:TGTGCTCTGATTTCTGC
clonotype3	2	0.00224719	TRA:CIVREGGDDKIIF	TRA:TGCATCGTCAGAGAGG

vloupe file for interactive analysis with Loupe V(D)J Browser

Cell Ranger V(D)J analysis: QC metrics

web_summary.html

QC Metrics: PBMCs of a healthy donor - TCR

3,928
Estimated Number of Cells

20,184
Mean Read Pairs per Cell

3,364
Number of Cells With Productive V-J Spanning Pair

Enrichment

Reads Mapped to Any V(D)J Gene	80.7%
Reads Mapped to TRA	24.4%
Reads Mapped to TRB	56.0%

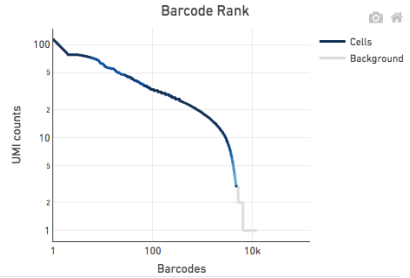
V(D)J Expression

Median TRA UMIs per Cell	4.00
Median TRB UMIs per Cell	10.00

V(D)J Annotation

Cells With Productive V-J Spanning Pair	85.6%
Cells With Productive V-J Spanning (TRA, TRB) Pair	85.6%
Paired Clonotype Diversity	2,597.91
Cells With TRA Contig	92.6%
Cells With TRB Contig	99.2%
Cells With CDR3-annotated TRA Contig	89.9%
Cells With CDR3-annotated TRB Contig	98.4%
Cells With V-J Spanning TRA Contig	91.8%
Cells With V-J Spanning TRB Contig	98.5%
Cells With Productive TRA Contig	87.9%
Cells With Productive TRB Contig	97.8%

Cells



Estimated Number of Cells	3,928
Mean Read Pairs per Cell	20,184
Mean Used Read Pairs per Cell	14,432
Fraction Reads in Cells	86.6%

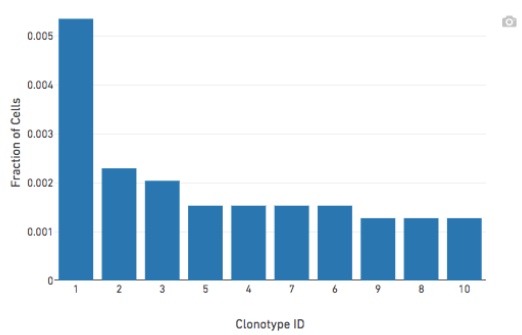
Sequencing

Number of Read Pairs	79,280,913
Valid Barcodes	93.0%
Q30 Bases in Barcode	96.0%
Q30 Bases in RNA Read 1	93.9%
Q30 Bases in RNA Read 2	91.9%
Q30 Bases in Sample Index	92.6%
Q30 Bases in UMI	95.8%

Sample

Sample ID	Demo
Sample Description	PBMCs of a Healthy Donor
Chemistry	Single Cell V(D)J
V(D)J Reference	vdj_GRCh38_alts_ensembl-3.1.0-3.1.0
Pipeline Version	3.1.0

Top 10 Clonotype Frequencies



Top 10 Clonotype CDR3 Sequences

Clonotype ID	CDR3s	Frequency	Proportion
1	TRA:CIVRGPRNTGNQFYF TRA:CAYRSVYRSFMYSGGGADGLTF TRB:CASSLEVGGGEETQYF	21	0.53%
2	TRA:CAYRSVYRSFMYSGGGADGLTF TRB:CASSLEVGGGEETQYF	9	0.23%
3	TRA:CAAYSGSRLTF TRB:CASSEAPETSGTKGYEQFF	8	0.20%
5	TRA:CAVDQAGTALIF TRB:CASSQLLGQGPYEQYF	6	0.15%
4	TRA:CAVRAYGNKLVF TRB:CASMGATASYEQYF	6	0.15%
7	TRA:CAMRPSFSGGYNKLIF TRB:CASSQDSSSNSPLHF	6	0.15%
6	TRA:CAAWNDYKLSF TRB:CASSQSTFEASGELFF	6	0.15%
9	TRA:CAVVRRKSGANNLFF TRB:CASSFTTGANGYTF	5	0.13%
8	TRB:CASSQLLGQGPYEQYF	5	0.13%
10	TRA:CAVITYGNNRLAF TRB:CASSSGRMIQETQYF	5	0.13%

QC Metrics: PBMCs of a healthy donor - Ig

High expressing plasma cell subpopulation; typical for B cells

775
Estimated Number of Cells

17,455
Mean Read Pairs per Cell

730
Number of Cells With Productive V-J Spanning Pair

Enrichment

Reads Mapped to Any V(D)J Gene	82.9%
Reads Mapped to IGH	21.8%
Reads Mapped to IGK	43.1%
Reads Mapped to IGL	17.8%

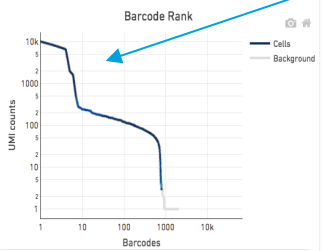
V(D)J Expression

Median IGH UMIs per Cell	19.00
Median IGK UMIs per Cell	46.00
Median IGL UMIs per Cell	41.00

V(D)J Annotation

Cells With Productive V-J Spanning Pair	94.2%
Cells With Productive V-J Spanning (IGK, IGH) Pair	54.8%
Cells With Productive V-J Spanning (IGL, IGH) Pair	41.4%
Paired Clonotype Diversity	706.09
Cells With IGH Contig	96.0%
Cells With IGK Contig	63.2%
Cells With IGL Contig	49.9%
Cells With CDR3-annotated IGH Contig	94.6%
Cells With CDR3-annotated IGK Contig	59.4%
Cells With CDR3-annotated IGL Contig	44.4%
Cells With V-J Spanning IGH Contig	95.7%
Cells With V-J Spanning IGK Contig	61.4%
Cells With V-J Spanning IGL Contig	45.7%
Cells With Productive IGH Contig	94.3%
Cells With Productive IGK Contig	58.6%
Cells With Productive IGL Contig	44.1%

Cells



Estimated Number of Cells	775
Mean Read Pairs per Cell	17,455
Mean Used Read Pairs per Cell	7,790
Fraction Reads in Cells	89.9%

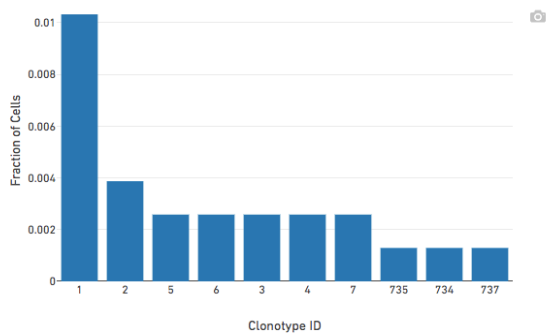
Sequencing

Number of Read Pairs	13,527,419
Valid Barcodes	95.1%
Q30 Bases in Barcode	95.5%
Q30 Bases in RNA Read 1	94.1%
Q30 Bases in RNA Read 2	90.3%
Q30 Bases in Sample Index	93.1%
Q30 Bases in UMI	95.5%

Sample

Sample ID	Demo
Sample Description	PBMCs of a Healthy Donor - Ig
Chemistry	Single Cell V(D)J
V(D)J Reference	vdj_GRCh38_alts_ensembl-3.1.0-3.1.0
Pipeline Version	3.1.0

Top 10 Clonotype Frequencies

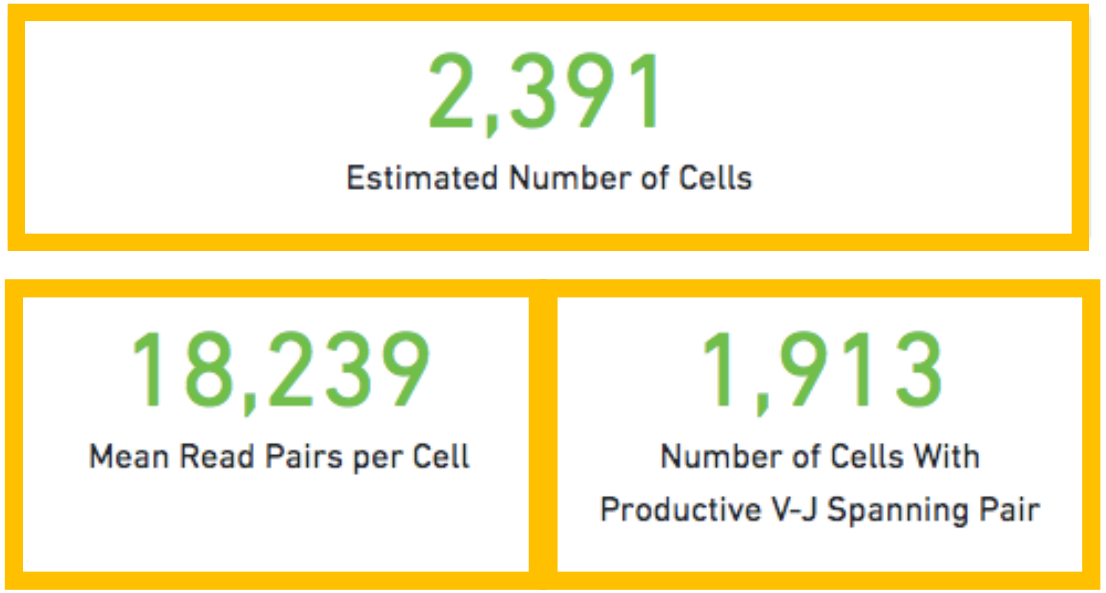


Top 10 Clonotype CDR3 Sequences

Clonotype ID	CDR3s	Frequency	Proportion
1	IGK:CQQYDNWPPYTF	8	1.03%
2	IGK:CQQYDNWPPYTF IGL:CCSYAGSYTPWVF	3	0.39%
5	IGH:CAKSFFTGTGQFHYW IGK:CQQYSTYSQTF	2	0.26%
6	IGH:CARPLTLADSARFDYW IGK:CQQYDNWPPYTF	2	0.26%
3	IGL:CQSYDSSLSGVF	2	0.26%
4	IGH:CARRPWDSGWYKYW IGK:CQQYDNWPPYTF	2	0.26%
7	IGH:CARPGTTGTTGLKNW IGK:CQQYNNWPLTF	2	0.26%
735	IGH:CAYSSGYSSSWSLGYMDVW IGK:CLQHDNFPPLTF IGL:CAAWDDSLNGHVVF	1	0.13%
734	IGH:CATVIVVHRGYYFDYW IGL:CNSRDSSGNHLVVF	1	0.13%
737	IGH:CARDPDPHYYGSGSYSNYFDYW IGL:CSSYTSSSTPLVF	1	0.13%

Cell Ranger V(D)J: QC metrics

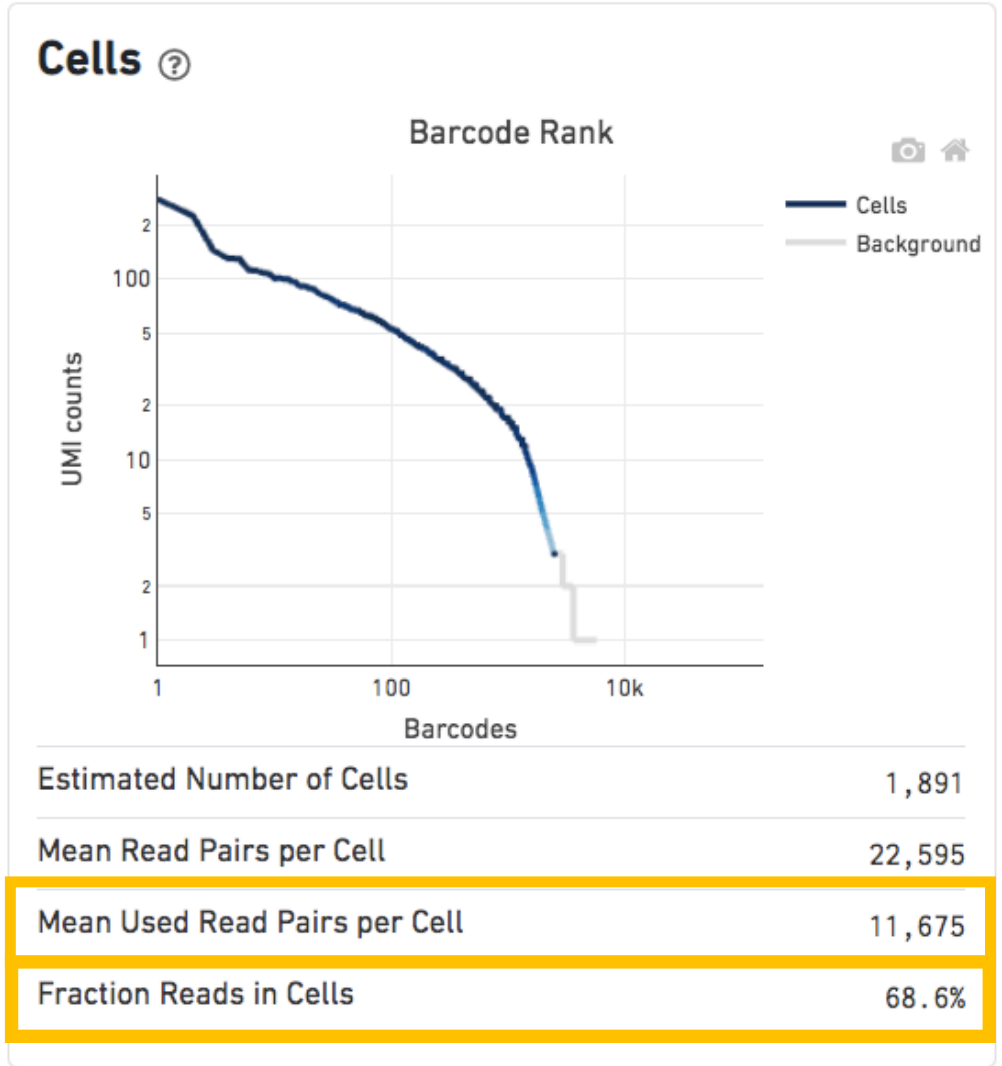
- Estimated Number of Cells
  - Number of cells expressing TCR/Ig chains
- Mean Read Pairs per Cell
  - Number of Read Pairs divided by the estimated Number of Cells
- Number of Cells with Productive V-J Spanning Pair
  - Number of cells with at least one TRA/TRB pair or Ig heavy/light chain pairs where both contigs span full length V(D)J (with in-frame CDR3, and has no stop codons in the V-J region)



Cell Ranger V(D)J: QC metrics

Cell calling

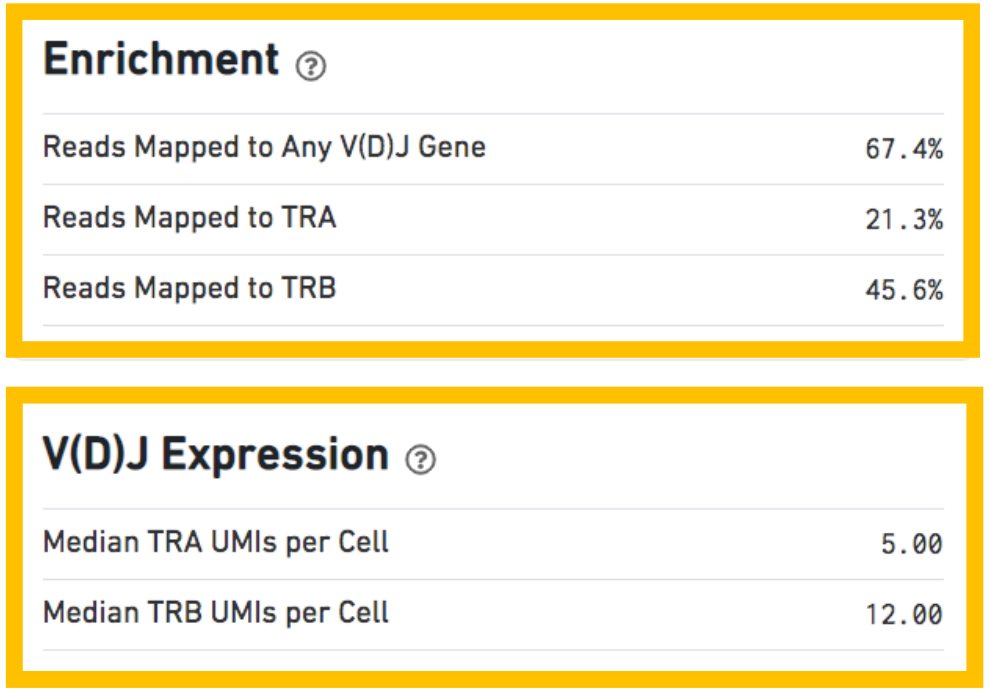
- Mean Used Read Pairs per Cell
  - Mean number of read pairs used in assembly per cell-associated barcode
- Fraction of Reads in Cells
  - Number of reads with cell-associated barcodes divided by the number of reads with valid barcodes



Cell Ranger V(D)J: QC metrics

Enrichment and TCR/Ig expression

- Enrichment
  - On-target rate
  - Typical values are over 80%
  - Low values for this metric indicate potential PCR undercycling
- V(D)J Expression
  - Metric indicates TCR / Ig expression levels
  - Values depend on sample type
  - TCR expression levels are typically lower than Ig



Enrichment

Reads Mapped to Any V(D)J Gene	67.4%
Reads Mapped to TRA	21.3%
Reads Mapped to TRB	45.6%

V(D)J Expression

Median TRA UMIs per Cell	5.00
Median TRB UMIs per Cell	12.00

10x GENOMICS

Cell Ranger V(D)J: QC metrics

Annotation

- Paired Clonotype Diversity
  - Measures clonal diversity. A value of 1 indicates a minimally diverse sample - only one distinct clonotype
- Metrics for various annotation steps:
  - These metrics tell us where we lost contigs on the road to calling Productive
  - Can indicate problems with Cell Ranger's annotation algorithm if the CDR3 rate is low (user can try their own annotation tool)
  - Low V-J spanning indicates poor assembly potentially due to insufficient depth

V(D)J Annotation

Cells With Productive V-J Spanning Pair	86.6%
Cells With Productive V-J Spanning (TRA, TRB) Pair	86.6%
Paired Clonotype Diversity	671.77
Cells With TRA Contig	93.0%
Cells With TRB Contig	98.9%
Cells With CDR3-annotated TRA Contig	91.4%
Cells With CDR3-annotated TRB Contig	98.1%
Cells With V-J Spanning TRA Contig	92.4%
Cells With V-J Spanning TRB Contig	98.3%
Cells With Productive TRA Contig	89.4%
Cells With Productive TRB Contig	97.2%

10x GENOMICS

Cell Ranger V(D)J: QC metrics

Sequencing

- Number of Read Pairs
  - Input raw read pairs count
- Valid Barcodes
  - Percent barcodes that are within 1 base mismatch of the whitelist
- Q30 Bases in Reads
  - Fraction of cell barcode bases with Q-score >= 30, excluding very low quality/no-call (Q <= 2) bases from the denominator

Sequencing

Number of Read Pairs	42,727,096
Valid Barcodes	88.1%
Q30 Bases in Barcode	96.1%
Q30 Bases in RNA Read 1	93.0%
Q30 Bases in RNA Read 2	92.0%
Q30 Bases in Sample Index	93.1%
Q30 Bases in UMI	95.4%

10x GENOMICS

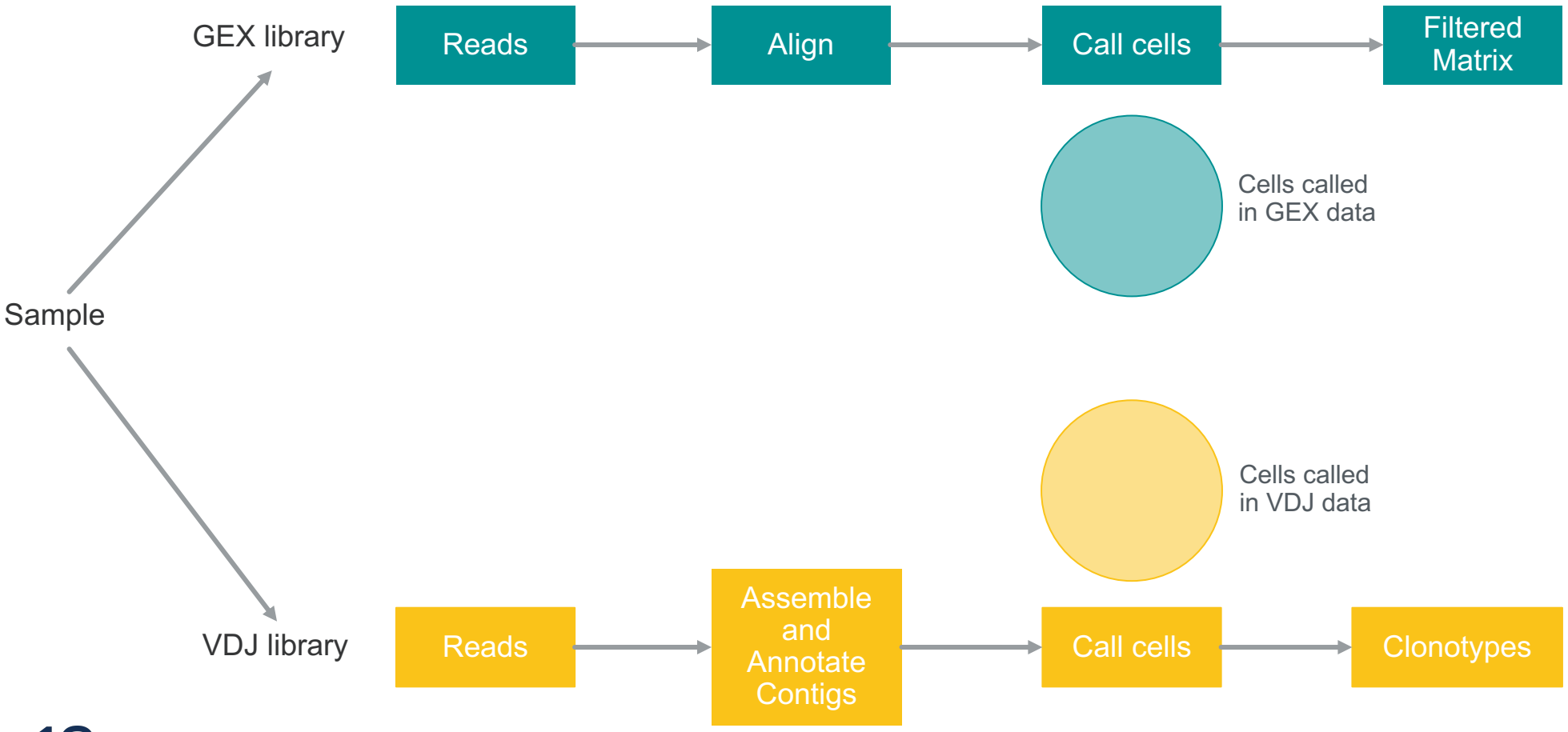
multi pipeline

New pipeline to analyze VDJ and GEX data together for streamlined cell calls.

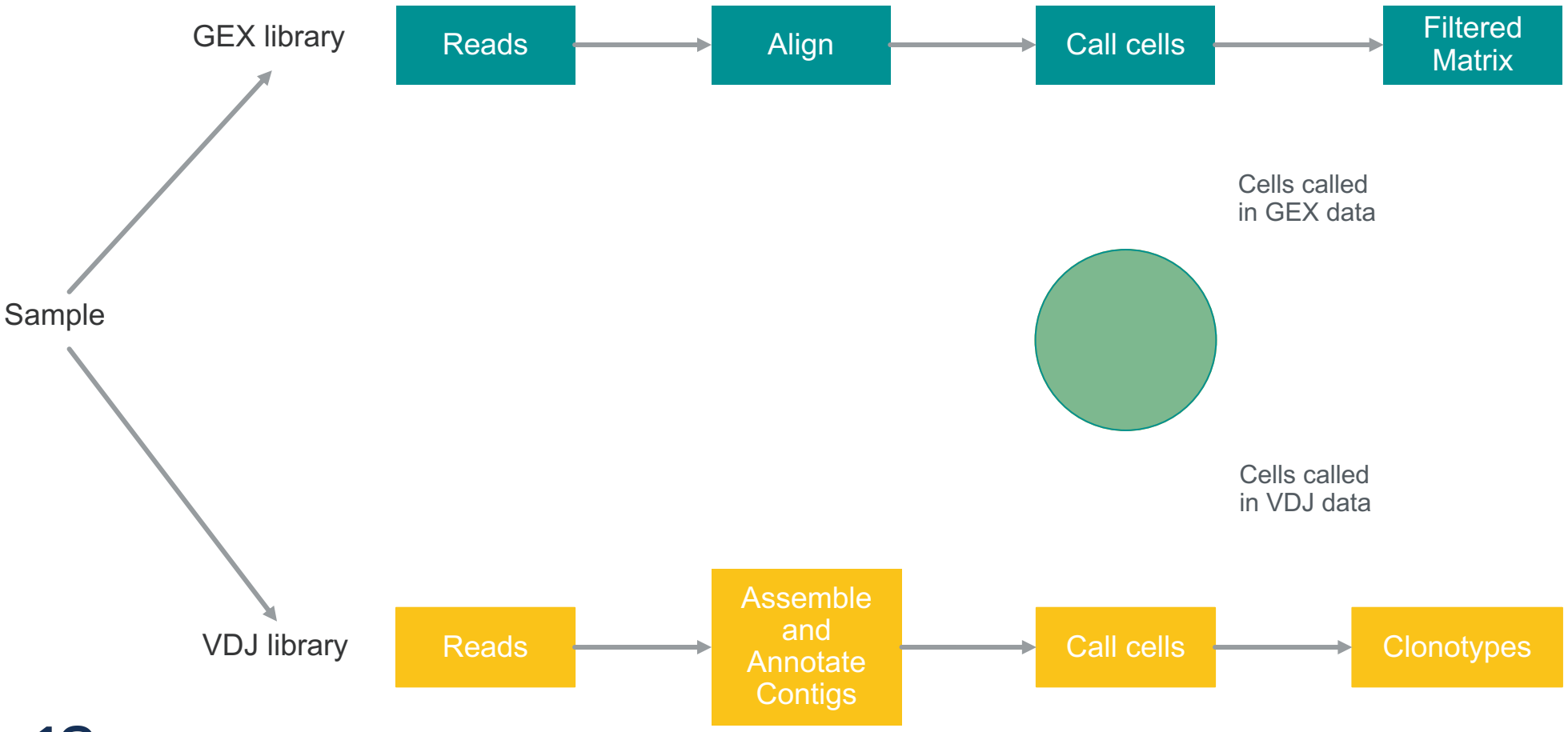
multi pipeline

- Motivation
  - **Simplify analysis** of VDJ data with Gene Expression and/or Feature Barcoding data.
  - Enable more **consistent cell calling** between various library types.

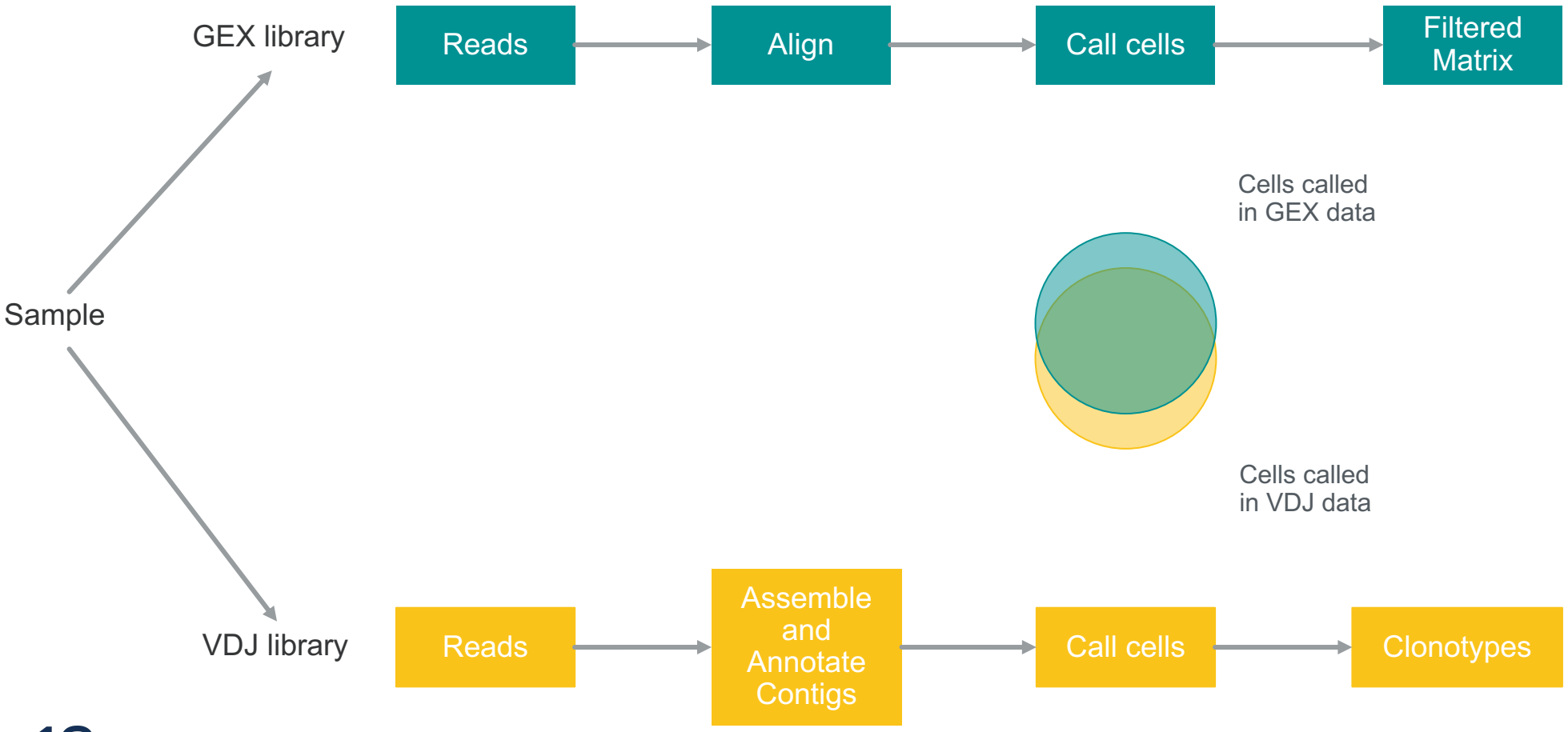
Cell calling differences



Cell calling differences



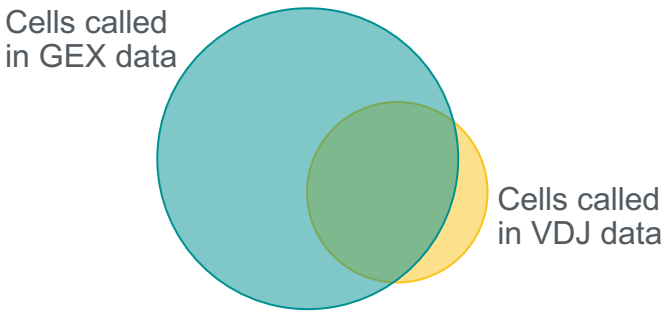
Cell calling differences



Cell calling differences

Under calling cells in VDJ

- More often seen in TCR data.
- Often due to low expression of TCR genes.
- Improved sensitivity of 5’ v2 chemistry will help this phenotype



Over calling cells in VDJ

- More often seen in BCR data.
- Likely due to ambient mRNA from high expressing plasma cells.
- New filters in VDJ pipeline also reduce false positives.

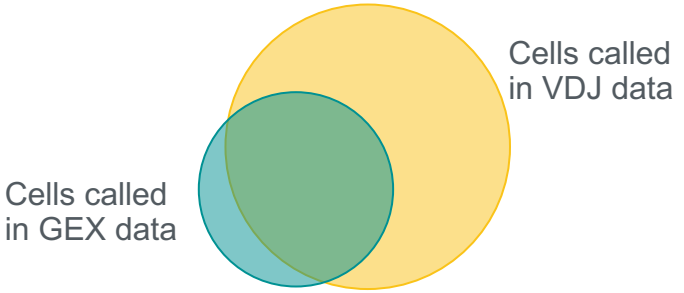
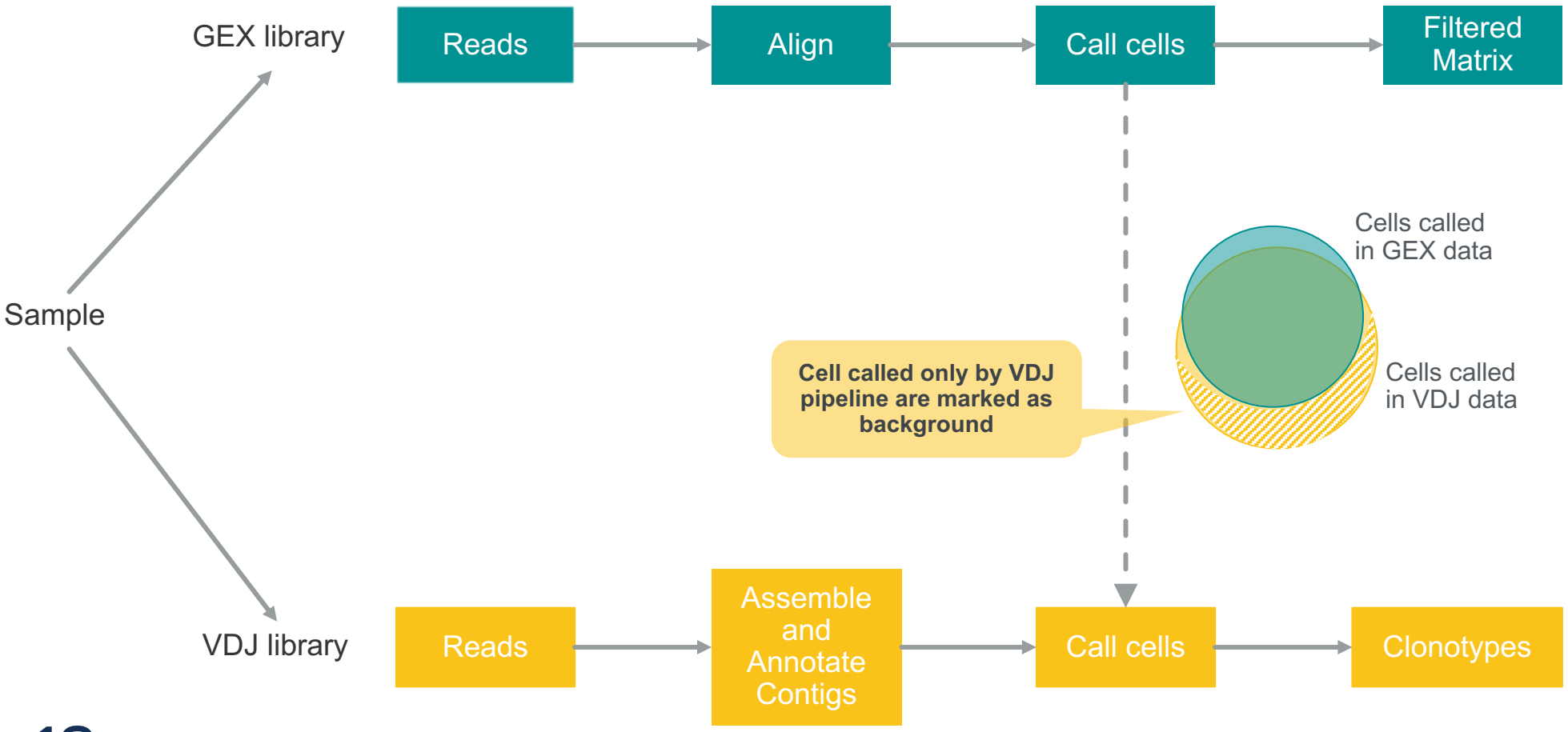


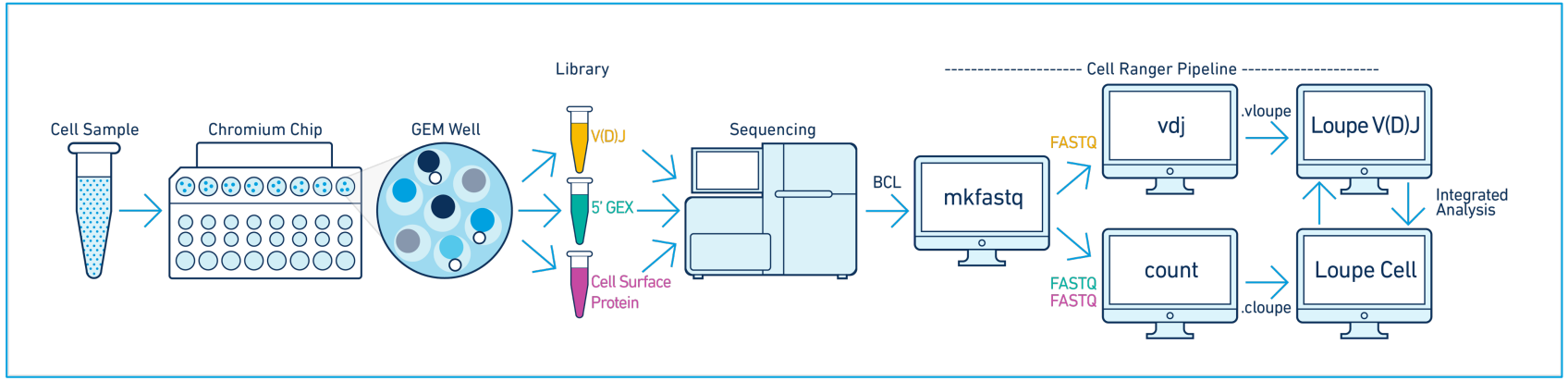
Illustration of potential cell calling differences between library types

Cell calling in *multi*

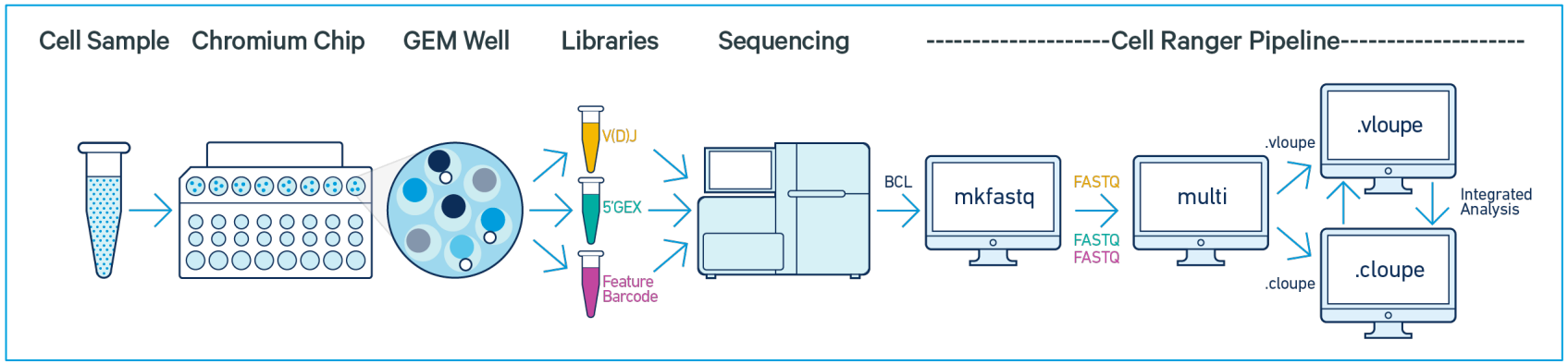


Overview of *multi* pipeline workflow

Old



New



How to run *multi*

- Example command

```
cellranger multi --id=OutputDirectory --csv=./experiment.csv
```

Input CSV

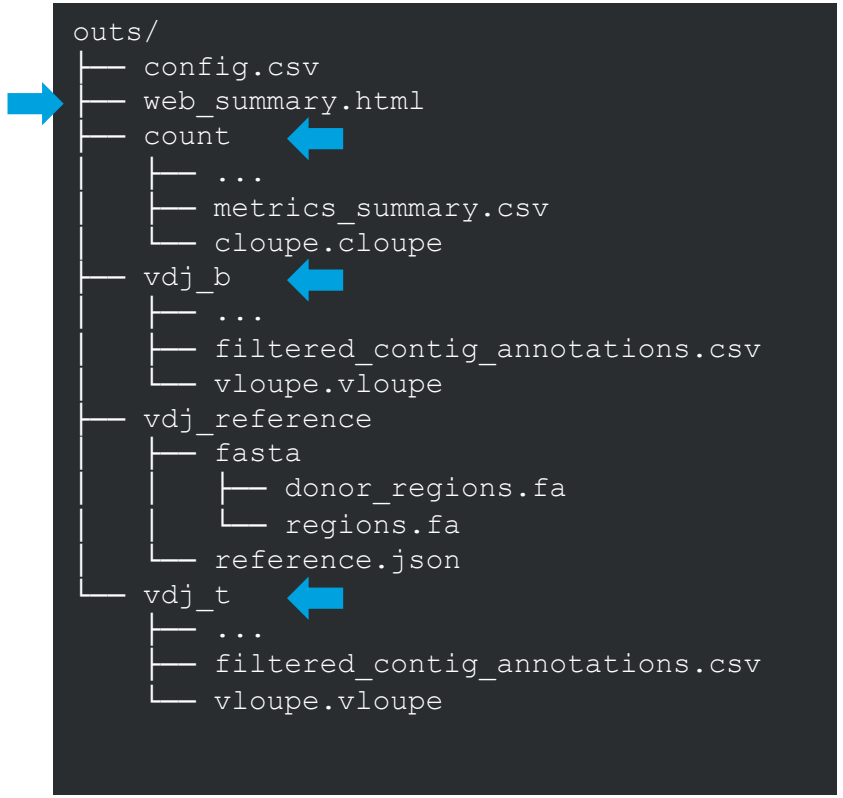
```
[gene-expression]                    # Section describing the GEX options/parameters.
ref,/mnt/data/ref/gex/refdata-gex-GRCh38-2020-A


[vdj]                                # Section describing the VDJ parameters.
ref,/mnt/data/ref/vdj/refdata-cellranger-vdj-GRCh38-alts-ensembl-4.0.0


[libraries]                          # Section describing the input sequencing libraries.
fastq_id,fastqs,lanes,feature_types,subsample_rate
MySample_GEX,/mnt/data/fastqs,"any",gene expression,
MySample_Ig,/mnt/data/fastqs,"any",vdj,
```

10x GENOMICS

Output files

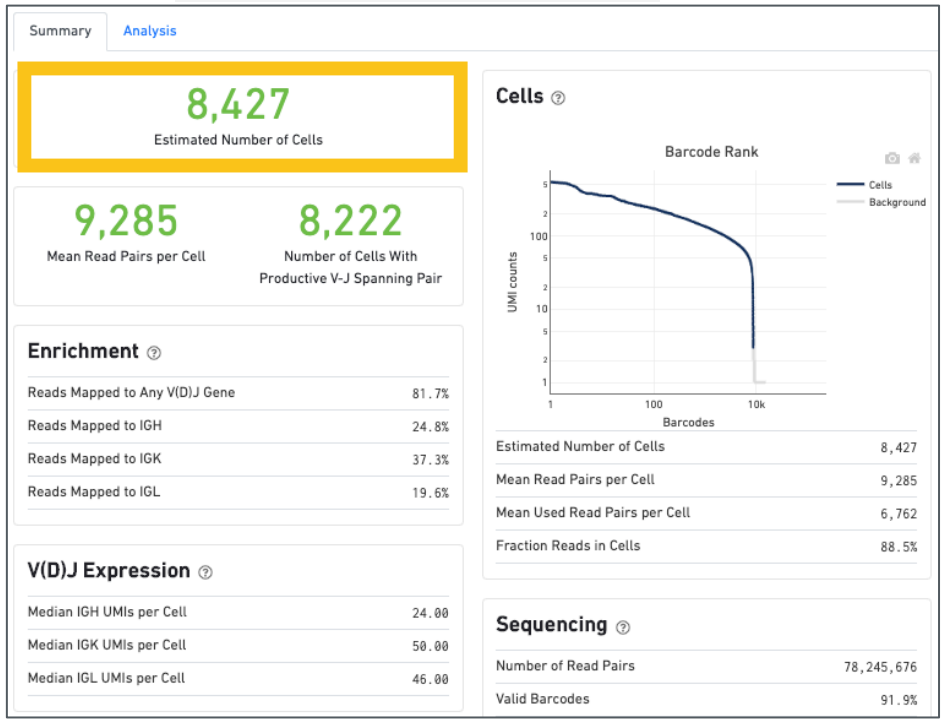


- One combined web_summary.html
- Subdirectories for each library.
  - “count” for gene expression and, feature barcoding outputs.
  - vdj_t, vdj_b for TCR, BCR library outputs.
- Content of subdirectories identical to ”count” and “vdj” pipelines.
- Separate per library vloupe and cloupe files.

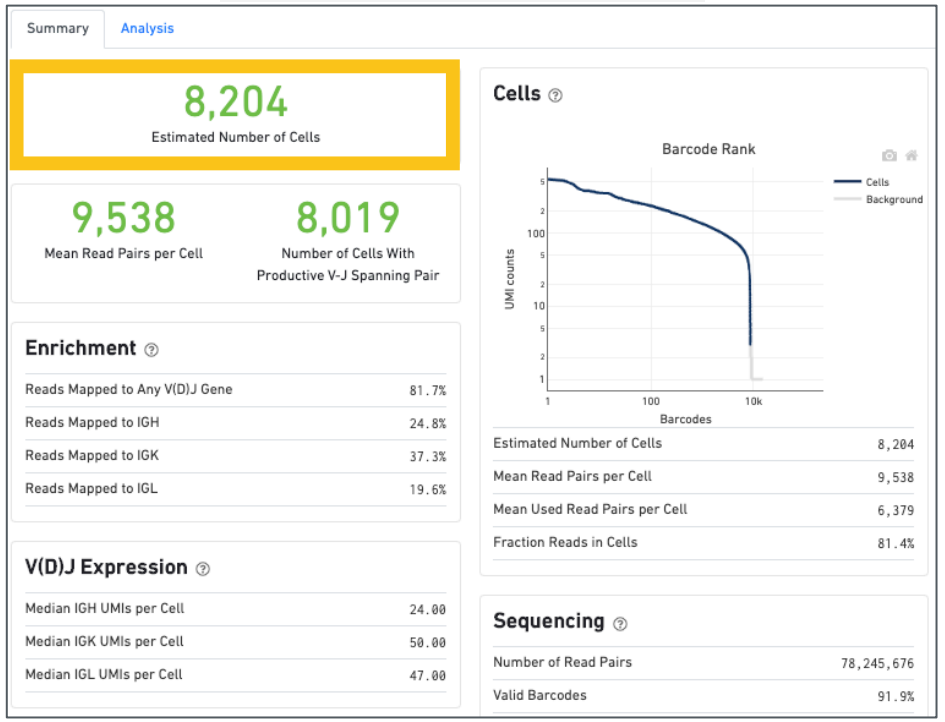
10x GENOMICS

vdj/count vs multi

vdj pipeline in v5.0



multi pipeline in v5.0



No change expected in GEX data between *count* versus *multi*

Take home for *multi*

- If you are working with VDJ data together with Gene Expression, *multi* pipeline is recommended.
- Helps in reducing false cell calls in VDJ data.

VDJ	5' GEX	5' FB	Use *multi* ?
Yes	Yes	Yes	Recommended
Yes	Yes	No	Recommended
Yes	No	Yes	Optional. No effect in cell calling.
Yes	No	No	Optional
No	No	Yes	Optional
No	Yes	No	Optional
No	Yes	Yes	Optional

10x GENOMICS

VDJ aggr

Combine VDJ data from multiple libraries

VDJ *aggr*

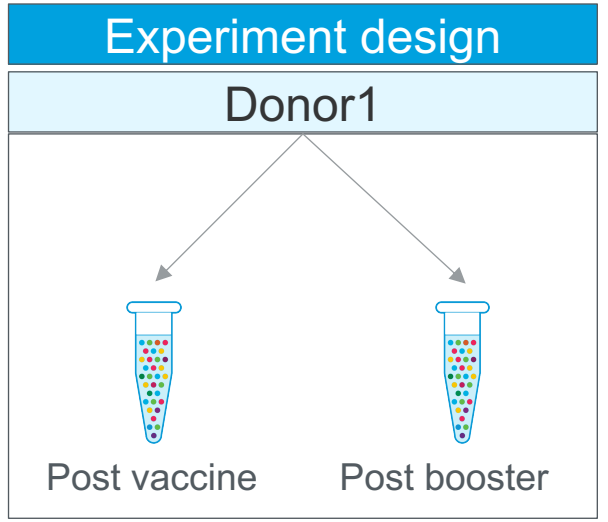
- "*cellranger aggr*" command supports combining VDJ data from Cell Ranger v5
  - Combine clonotypes
  - Generate a combined vloupe file

- Why ?
  - Analyze clonotypes between experimental conditions
  - Boost the power of clonotype grouping by combining more cells
  - Interactive analysis of combined data in vloupe

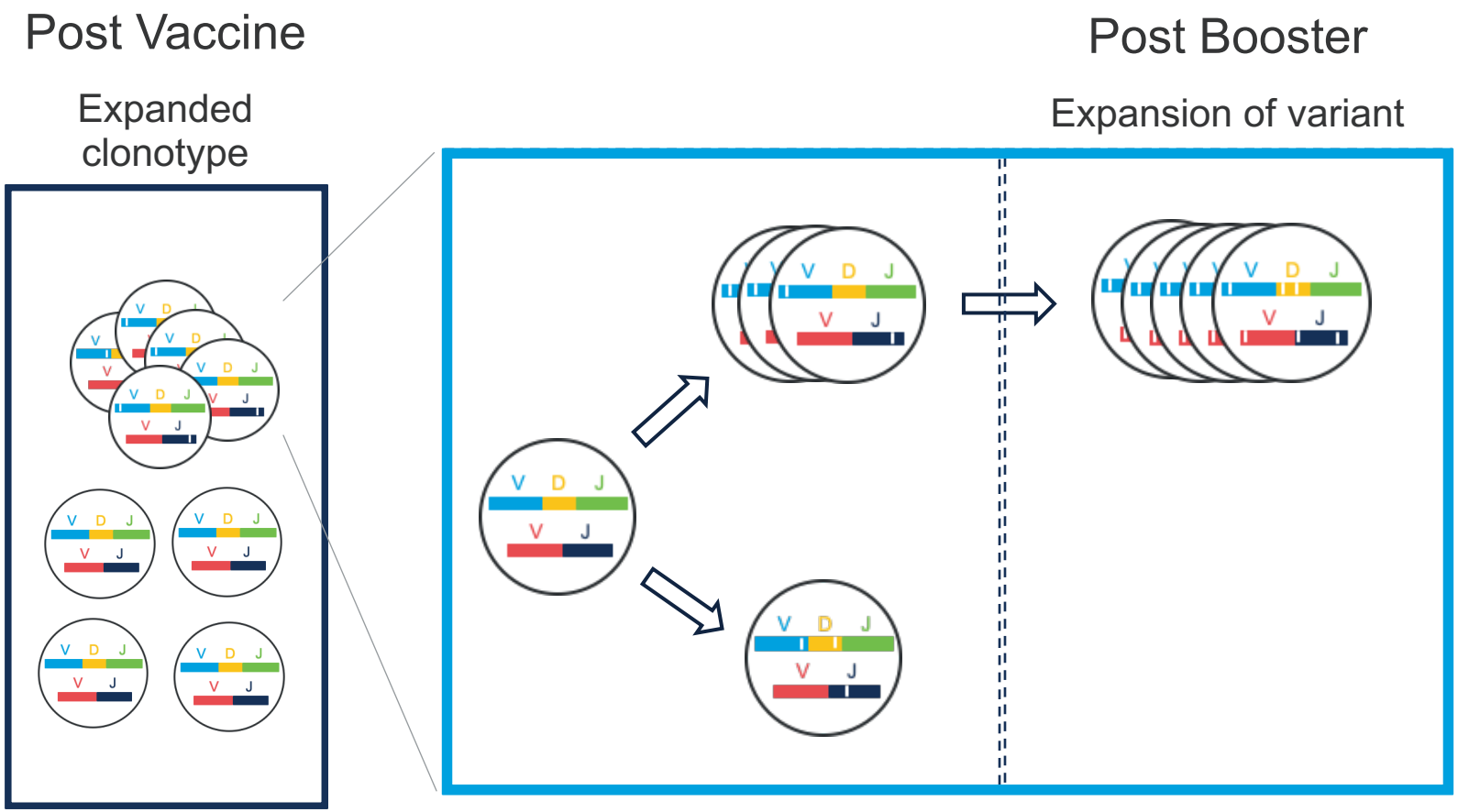
10x GENOMICS

Example use case

- Potential questions within donor
  - Which are clonotypes found at each timepoint
  - Which clonotypes are unique to each time point and which ones are persistent
  - Are there expanded clonotype seen post vaccination and booster shot



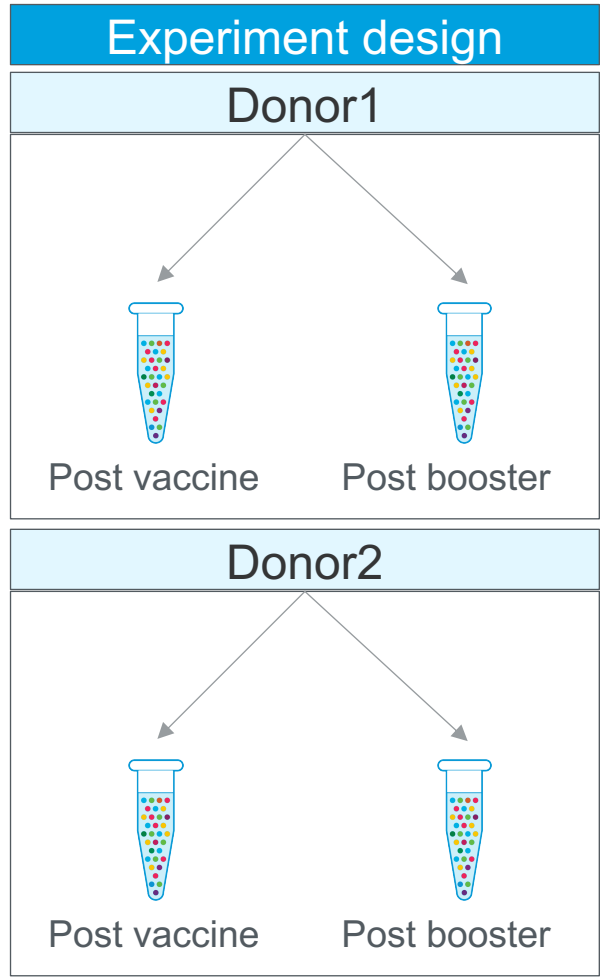
Example use case



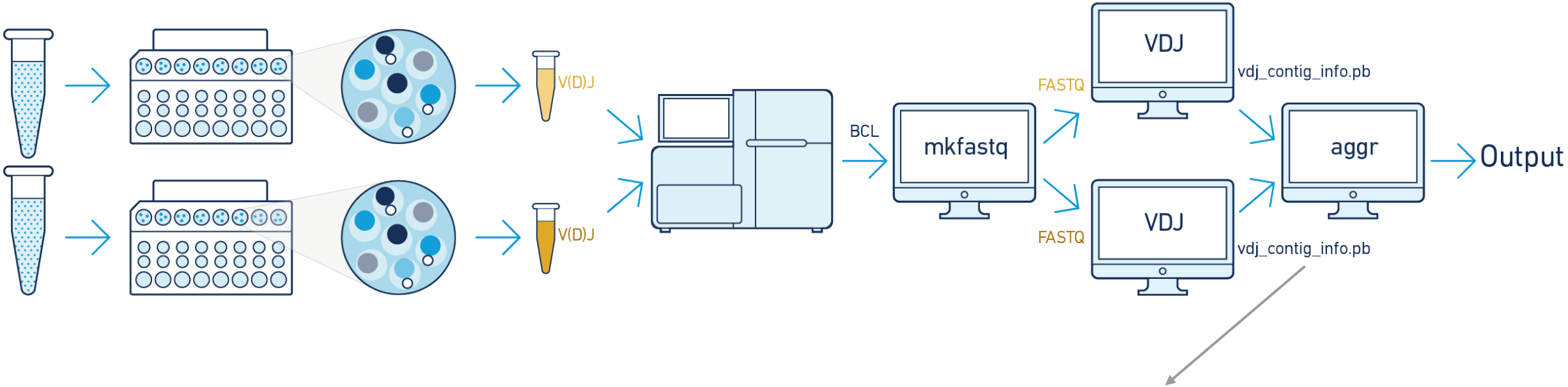
- *aggr* reruns clonotype grouping for all cells from same donor
- Enables combining clonotypes from same lineage

Example use case

- Potential questions across donors
  - Are there shared CDR3
  - Trends in the V/D/J gene usage



Overview of *aggr* workflow



- New file generated by VDJ pipeline from v5.0
- Binary file to be used specifically for *aggr*
- Combining GEX+VDJ runs from *multi* is enabled

How to run *aggr*

- Example command

```
cellranger aggr --id=OutputDirectory --csv=./experiment.csv
```

Making the input CSV

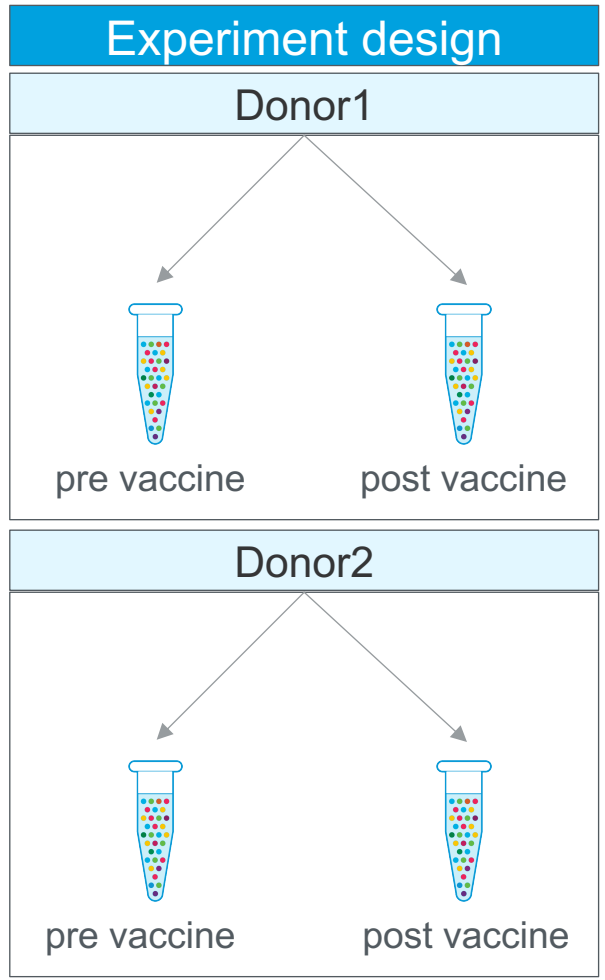
----------Required---------- ----------Required---------- -----Optional-----

library_id	vdj_contig_info	donor	origin	Treatment
donor1_pre	/home/donor1_pre/outs/vdj_contig_info.pb	donor1	donor1_pre	pre_vaccine
donor1_post	/home/donor1_post/outs/vdj_contig_info.pb	donor1	donor1_post	post_vaccine
donor2_pre	/home/donor2_pre/outs/vdj_contig_info.pb	donor2	donor2_pre	pre_vaccine
donor2_post	/home/donor2_post/outs/vdj_contig_info.pb	donor2	donor2_post	post_vaccine

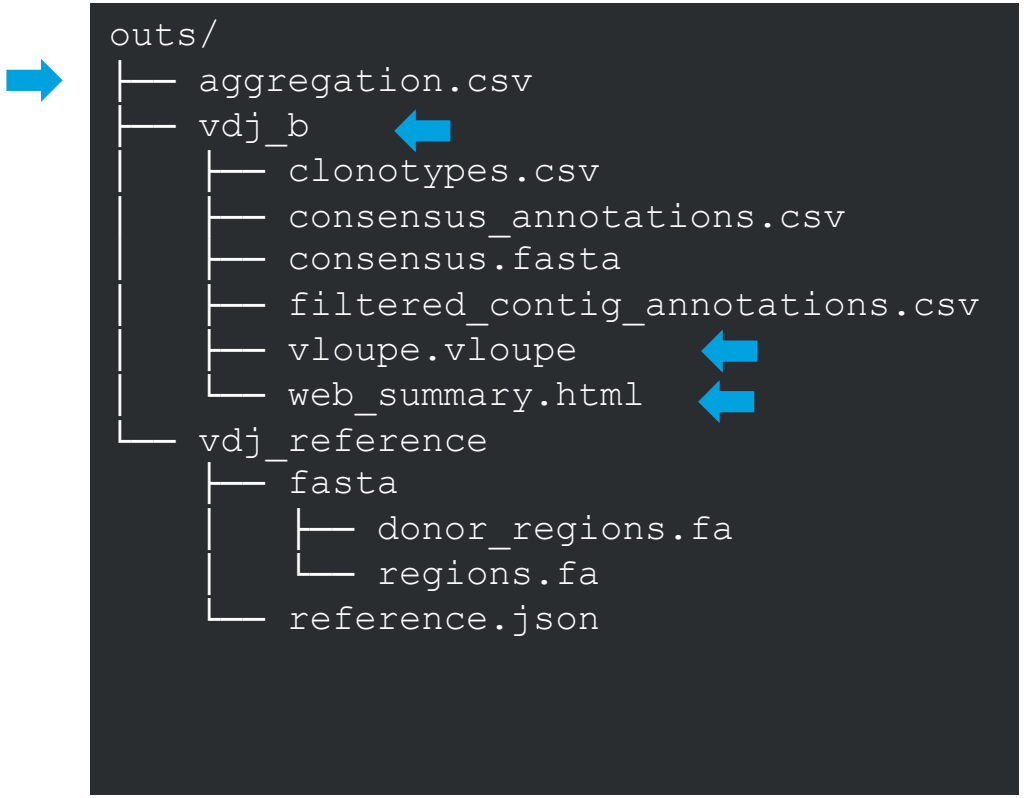
- Identifier for the donor
- Combine clonotypes

Identifier for cells from a donor for an experimental condition

Custom grouping



Output files



- Input csv file
- Subdirectories for each library type.
  - “count” for gene expression and, feature barcoding outputs.
  - vdj_t, vdj_b for TCR, BCR library outputs.
- web_summary.html with combined information
- vloupe and cloupe for VDJ and GEX libraries

VDJ aggr web_summary.html

44,687
Total Number of Cells

43,222
Total Number of Clonotypes

43,279
Number of Cells With Productive V-J Spanning Pair

Sample

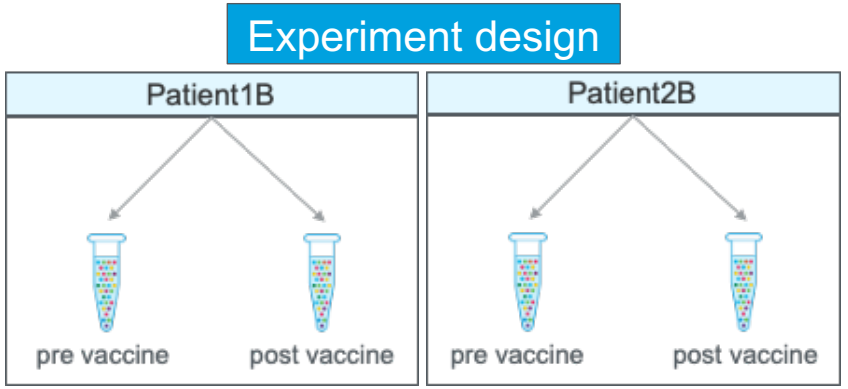
Sample ID	aggr_donor1B_donor2B_5p0
Sample Description	
V(D)J Reference	vdj_GRCh38_alts_ensembl-4.0.0
Pipeline Version	cellranger-5.0.0
Number of Libraries	4
Number of Donors	2
Number of Origins	4

Annotation

Cells With Productive V-J Spanning Pair	96.85%
Cells With Productive V-J Spanning (IGK, IGH) Pair	57.09%
Cells With Productive V-J Spanning (IGL, IGH) Pair	44.33%

Cells

Donor	Origin	Cells	Clonotypes	Diversity
patient1B	post_p1	8,438	8,319	8,138.99
patient1B	pre_p1	9,125	8,982	8,802.79
patient2B	post_p2	13,163	12,954	12,685.01
patient2B	pre_p2	13,961	13,080	12,336.83



- Total Number of Cells: Number of cells across all libraries
- Breakdown of cell counts for each donor, origin combination.
- Cell counts may be different from those obtained from original VDJ pipeline.

Top clonotypes

Per library clonotypes

Top 10 Clonotype CDR3 Sequences

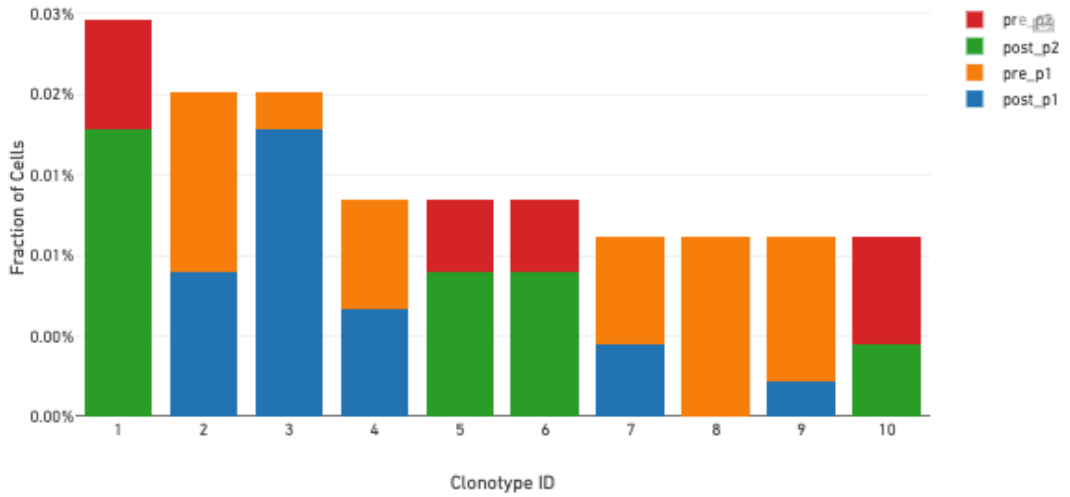
Clonotype ID	CDR3s	Frequency	Proportion
1	IGH:CSRSLNYW IGK:CQQYNSYPRTF	5	0.06%
2	IGH:CATSVAAPGNYW IGL:CSAWDSSLSAWVF	5	0.06%
3	IGH:CARVREGDYFGPLEYW IGK:CQQSYTAPGTF	4	0.04%

Top 10 Clonotype CDR3 Sequences

Clonotype ID	CDR3s	Frequency	Proportion
1	IGH:CARGFDGRDGFDIW IGL:CAAWDDSLSGQVF	8	0.10%
2	IGH:CSRSLNDW IGK:CQQYNSYPRTF	4	0.05%
3	IGH:CVRGDPRDYW IGL:CGTWDSSLNAWVF	4	0.05%

Combined clonotypes

Top 10 Clonotype Frequencies



Top 10 Clonotype CDR3 Sequences

Clonotype ID	CDR3s	Frequency	Proportion
1	IGH:CARDGSDDDFWNAYFKWFDPW IGL:CGTWDSSLSAVVF	11	0.02%
2	IGH:CSRSLNYW IGK:CQQYNSYPRTF	9	0.02%

Take home for VDJ *aggr*

- Combining VDJ data enabled in “cellranger aggr”
- Generate clonotypes for combined cells from same donor
- Generate combined vloupe for all combination types

Additional resources

Cell Ranger

Troubleshooting

- Cell Ranger Pipeline automatically produces a diagnostic file
- Send us the diagnostic file for
  - Test
  - Training run
  - Troubleshooting failed runs
  - Successful runs
- Send it to us directly from command line:

```
$ cellranger upload your@email.edu SampleA/SampleA.mri.tgz
```

10x GENOMICS

Cell Ranger

Support

Online Documentation:

http://support.10xgenomics.com

Q&A Knowledgebase:

https://kb.10xgenomics.com/hc/en-us

Please send questions, comments, and feedback to:

support@10xgenomics.com

Appendix A

Changes in Cell Ranger 3.1

Summary of Cell Ranger 3.1 updates for V(D)J

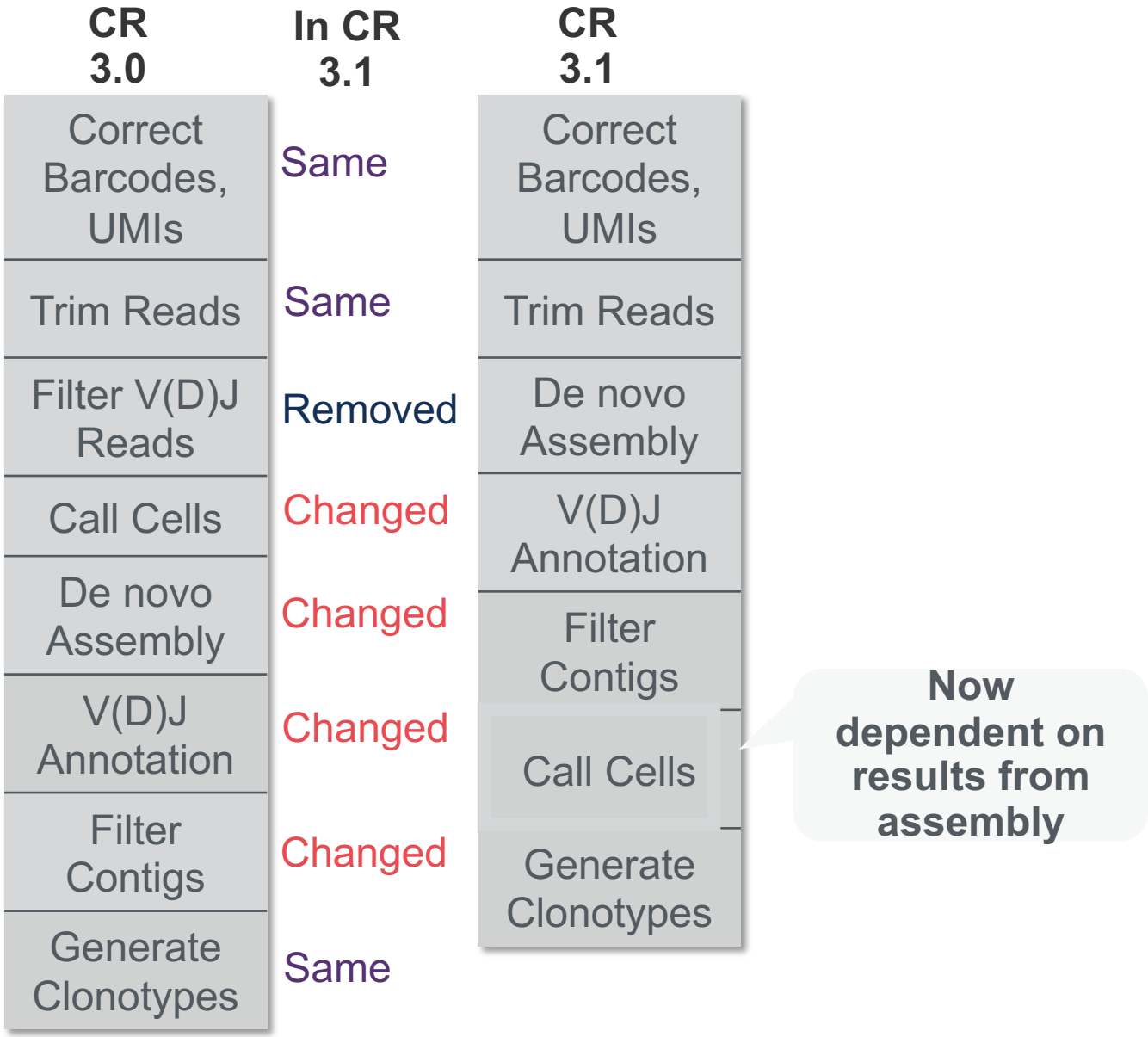
- Complete rewrite of key steps
  - Assembly
  - Annotation
  - Cell calling
- Human/mouse references improved
- Outcome of these changes
  - Sensitivity increased, particularly at lower coverage
  - Sequencing depth requirement reduced 2.5-fold
  - 26x91 sequencing supported, enabling VDJ+GEX in one run
  - Many output errors corrected, including for CDR3
  - Pipeline up to two-fold faster
- How to run pipeline
  - Remains same for human and mouse
  - Non-human/mouse: customers need to input primers
- Outputs
  - File formats unchanged
  - vloupe file and visualization unchanged

Overall recommendation is to switch to 3.1 including for preexisting datasets

10x GENOMICS

Changes in the pipeline

- Complete rewrite of key steps of the pipeline



New reference

- Manual improvement was done to Human and Mouse V(D)J annotations from Ensembl
- Examples of improvement:
  - Trimmed V genes such that now all begin with ATG (start codon)
  - Pseudogenes are excluded from the reference except in cases where we think pseudogene annotation was incorrect based on observed data
  - Trimmed 3 bases from the right of human TRAJ37 because otherwise we find a consistent three-base indel in observed data
  - Added an alternate form of mouse TRAV16N, differing from the reference form by having a 3-base insertion, because we observe this in data
- Details here:
  - https://support.10xgenomics.com/single-cell-vdj/software/pipelines/latest/advanced/built-in-refs

10x GENOMICS

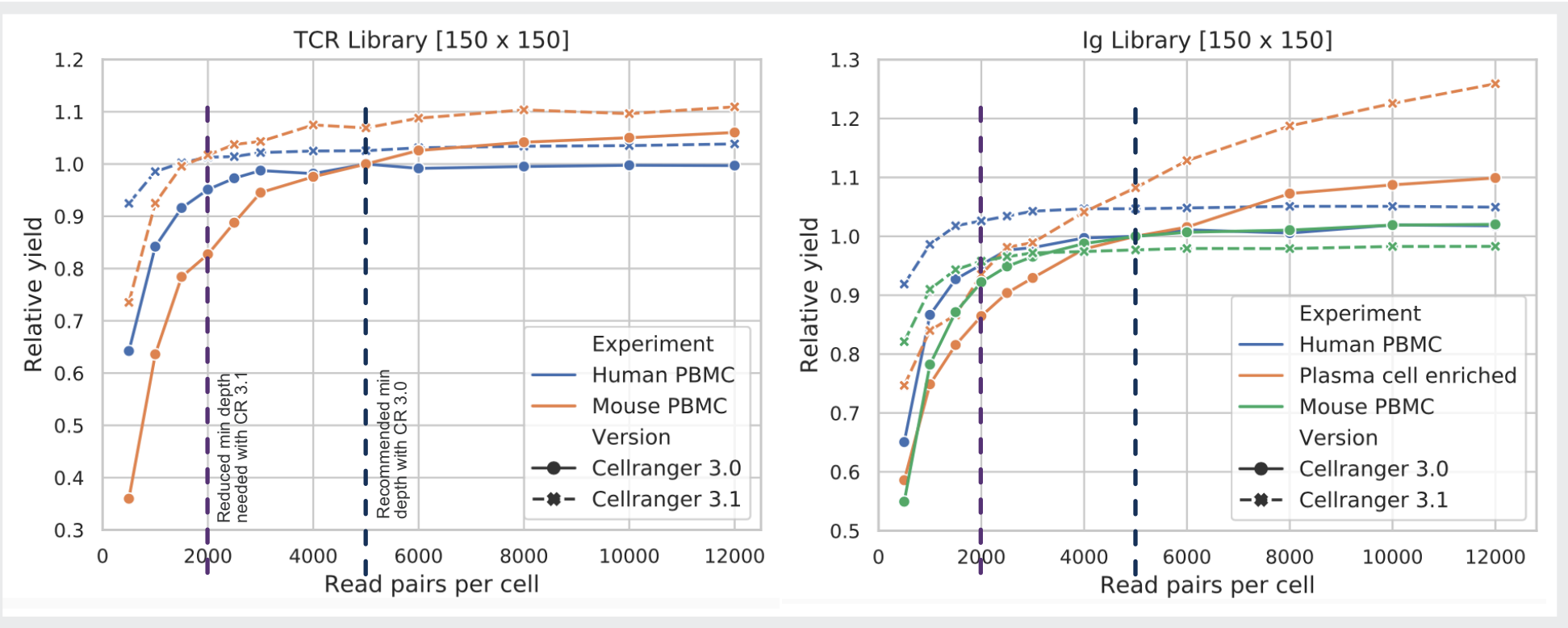
Outcome of these changes in Cell Ranger 3.1

- Sensitivity increased, particularly at lower coverage
- Sequencing depth requirement reduced 2.5-fold
- 26x91 sequencing supported, enabling pooling of VDJ, GEX, Cell Surface Protein libraries in one sequencing run
- Many output errors corrected, including CDR3 identification
- Pipeline up to two-fold faster

10x GENOMICS

Better sensitivity

- On an average we observed about 10% increase in sensitivity for internal datasets
- Cell Ranger 3.1 has much higher sensitivity at lower sequencing depth
- Conversely same yield with 2.5 fold less sequencing depth

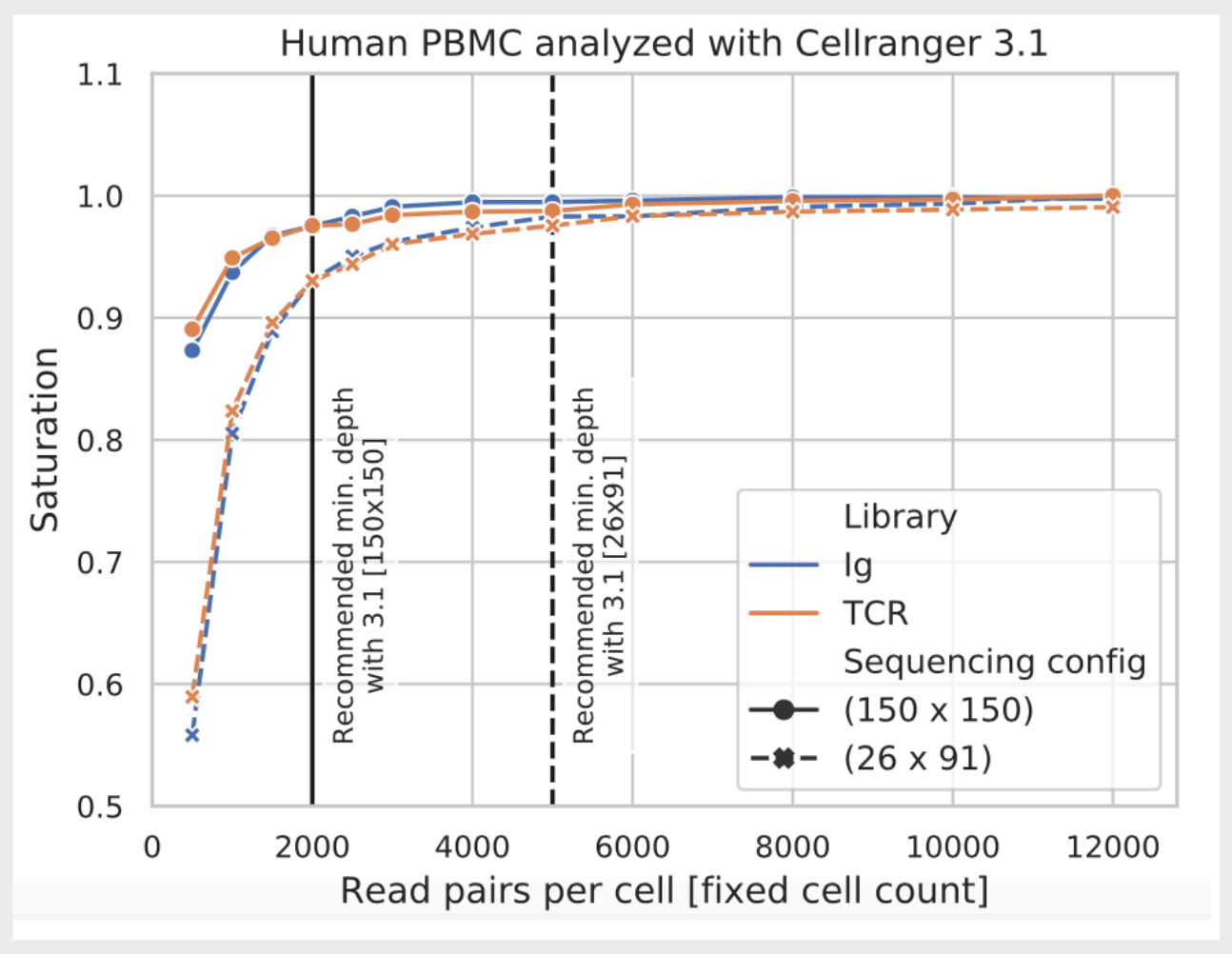


Relative Yield

- Numerator: Number of productive cells
- Denominator: Number of productive cells reported by 3.0 at 5k depth for each sample

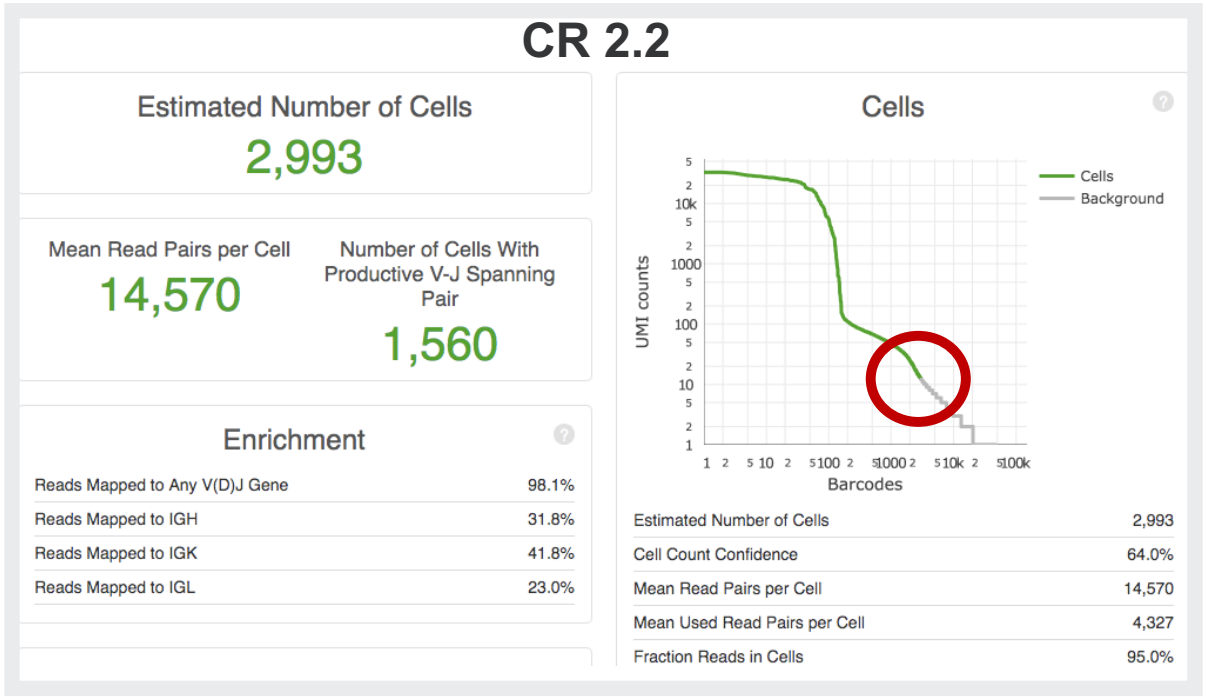
26x91 sequencing configuration is officially supported

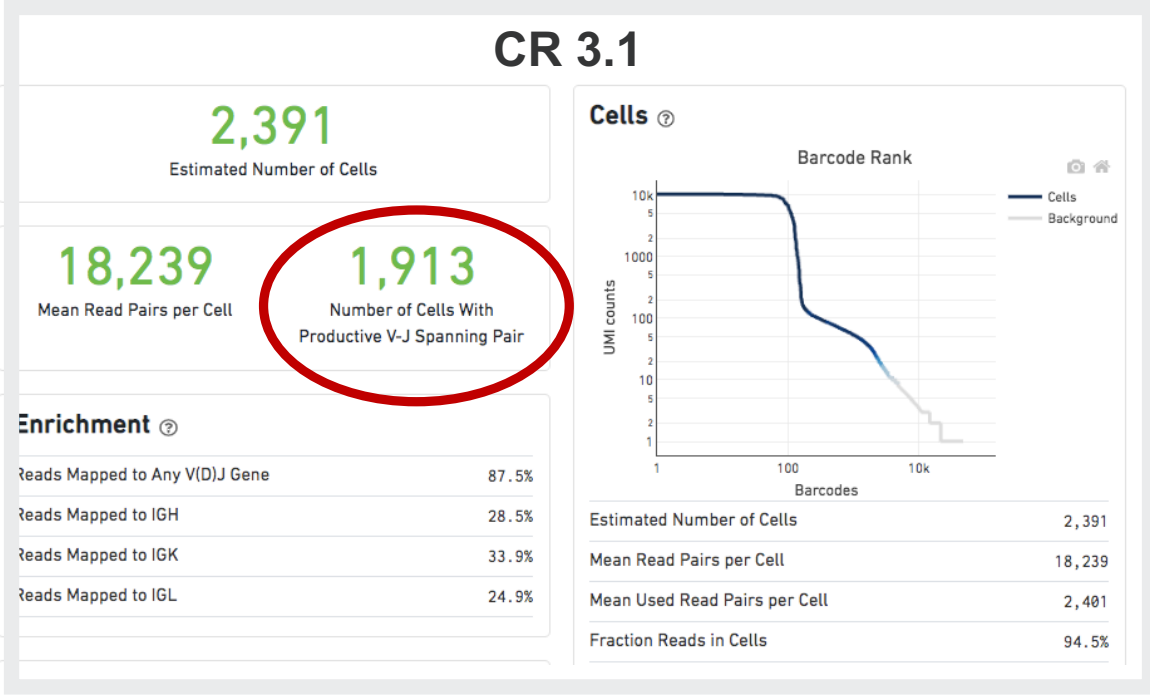
- Sequencing VDJ libraries at 150x150 might be inconvenient or expensive for some customers
- We now support 26x91 configuration
- Customers can now pool gene expression, cell surface protein and VDJ libraries more economically
- Our new recommendation is:
  - 26x91 with minimum depth of 5000 reads per cell
  - Can still do 150x150 with min. depth of 2000 reads per cell



Case Study: NSCLC tumor - Ig dataset

- NSCLC Ig dataset run using Cell Ranger 3.1 and compared to results from Cell Ranger 2.2 (available in public datasets)
- 600 more cells called by Cell Ranger 2.2 mostly with low UMI counts
  - Majority of the cells (~400) called by Cell Ranger 2.2 and not by Cell Ranger 3.1 are not called as cells by GEX also
- Higher number of productive pairs in Cell Ranger 3.1





Case Study: NSCLC tumor - Ig dataset

- Some clonal expansion was indicated in Cell Ranger 2.2 but Cell Ranger 3.1 does not show any clonal expansion

CR 2.2

Top 10 Clonotype CDR3 Sequences

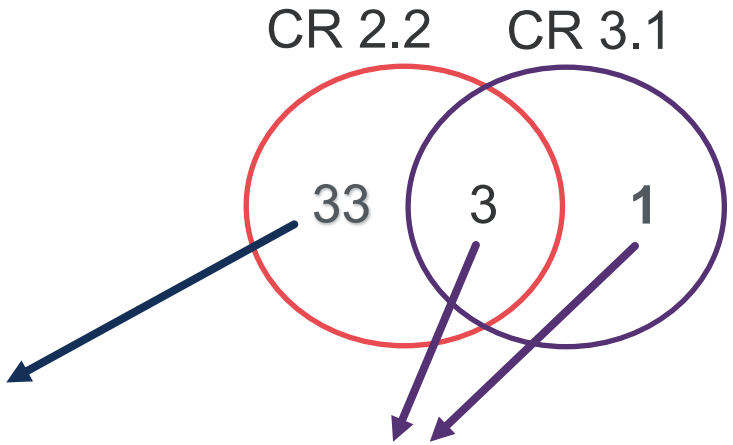
Clonotype ID	CDR3s	Frequency	Proportion
clonotype1	IGK:CMQALQTPLTF	98	3.3%
clonotype2	IGK:CMQALQTPWTF	92	3.1%
clonotype3	IGL:CSYIGGTTSRVF	56	1.9%
clonotype4	IGH:CASADPYSSTWHAWQGDFW IGK:CMQALQTPLTF	36	1.2%
clonotype5	IGH:CAKDSPLGGWFDPW IGL:CSYIGGTTSRVF	33	1.1%
clonotype6	IGH:CATTSPHVVVVPVADPPPFGHW IGL:CYSTDSSGNHRVF	11	0.4%
clonotype7	IGH:CAKDSPLGGWFDPW	10	0.3%
clonotype8	IGL:CYSTDSSGNHRVF	9	0.3%
clonotype9	IGL:CQVWDRSSVHYVF	5	0.2%
clonotype11	IGH:CARAGSGWPSNFYRYYYMDVW IGL:CATWDDSLSGPNWVF	5	0.2%

CR 3.1

Top 10 Clonotype CDR3 Sequences

Clonotype ID	CDR3s	Frequency	Proportion
1	IGK:CQQYYSTPTF	5	0.21%
2	IGH:CTTGRQGRTKDWFDPW IGL:CQTYDTSLSGSVF	5	0.21%
5	IGH:CASADPYSSTWHAWQGDFW IGK:CMQALQTPLTF	4	0.17%
4	IGH:CARGYCSGTSCYDLFDYW IGK:CQQYYSTPTF	4	0.17%
7	IGH:CARDWYCGANCYYFDHW IGL:CQVWDSNSDHVF	4	0.17%
6	IGH:CARQLSFNYYDTGFW IGL:CTSYTGNSILWVF	4	0.17%
3	IGH:CARKGGDFFGFDIW IGK:CQQYGNSLWTF	4	0.17%
9	IGH:CATPGPWATTFAYW IGK:CQQSYTTPWTF	3	0.13%
8	IGH:CARGGVDFLRNYHLAFDKW IGL:CCWSSFRGICTAAF	3	0.13%
10	IGH:CARAVSRSLGLPYFDSW IGL:CQVWDSSSDHCVF	3	0.13%

Case Study: NSCLC tumor - Ig dataset

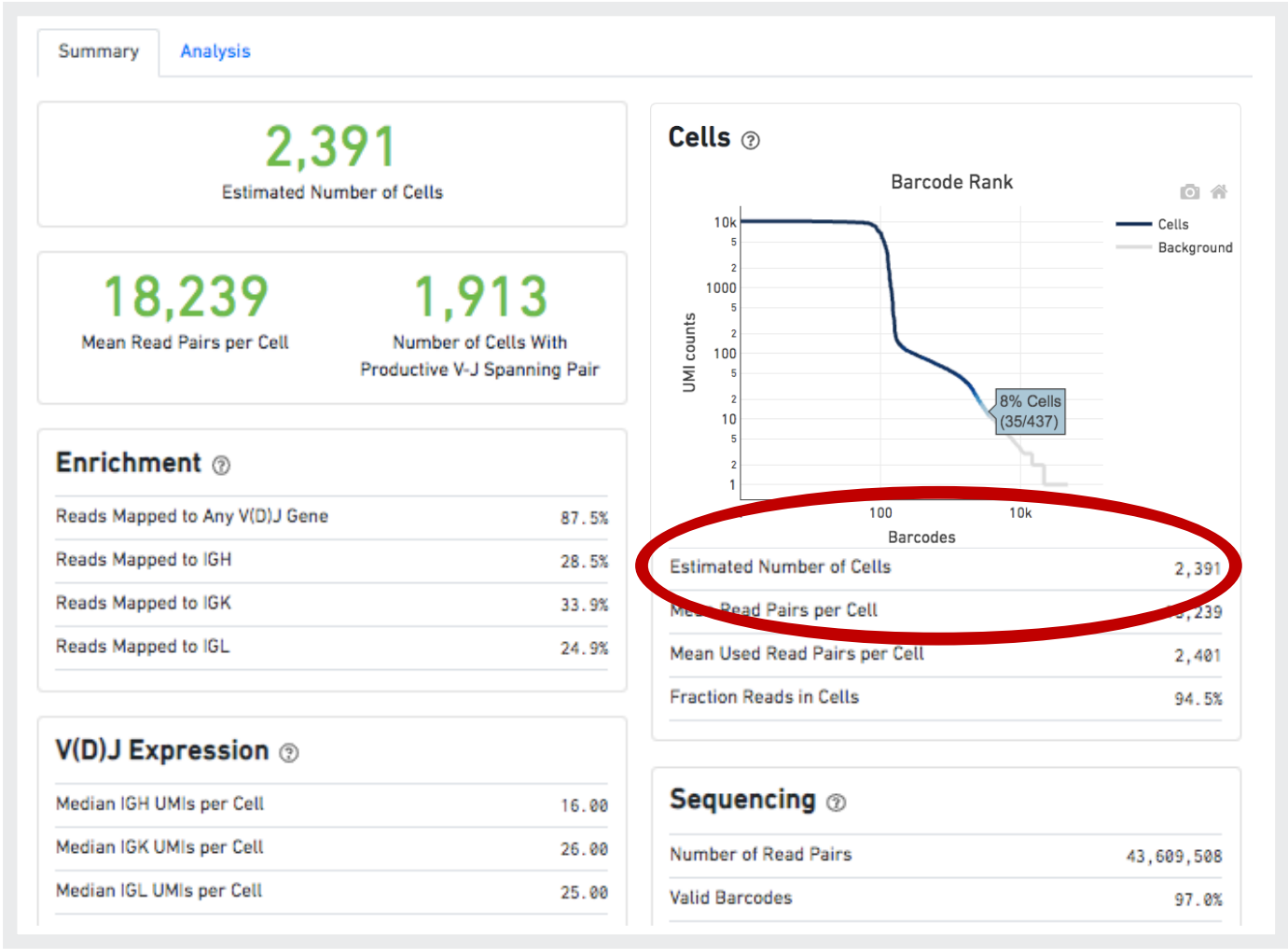


- Clonotype comparison:
  - v2.2 : 36 barcodes in clonotype4
  - v3.1 : 4 barcodes in clonotype5
- Validation from GEX
  - None of 33 were called as cells by GEX
  - All common and 1 cell in v3.1 clonotype were called as cells by GEX
- UMIs counts in these cells
  - v2.2 only cells have low UMI support (UMIs 1-10)
  - v3.1 cells have very high BCR UMI count. With one barcode having 11k and 9k UMIs for IGK and IGH respectively
- The 33 barcodes found only in clonotype4 in v2.2 might have been background GEMs miss-called as cells in v2.2
- v3.1 did the right thing by not calling them as cells

10x GENOMICS

Changes to metrics

- Barcode-Rank plot
  - In new cell calling algorithm, cells are not selected based solely on UMI and read support.
  - One criteria is them having at least 1 productive and confident contig
  - It is possible that some barcodes for a given UMI are called as cells and some not.
  - Local density of cells on the plot is illustrated by the shade of the line
- The metric “Cell Count Confidence” has been dropped. It was observed to be inaccurate in some instances



Key takeaways

- Complete rewrite of the key steps in VDJ pipeline:
  - Assembly
  - Annotation
  - Cell calling
- New improved references for human/mouse
- Outcome of above changes:
  - Sensitivity increased, particularly at lower coverage
  - Sequencing depth requirement reduced 2.5-fold
  - 26x91 sequencing supported, enabling pooling of VDJ, GEX, Cell Surface Protein libraries in one sequencing run
  - Many output errors corrected, including CDR3 identification
  - Pipeline up to two-fold faster

10x GENOMICS

Appendix B

VDJ Updates in Cell Ranger 4.0

V(D)J updates

5’ v2 chemistry

- **Question:** Can I combine 5’v1/v1.1 with 5’v2 data ?
- **Answer:** Yes, you can combine 5’ v1 chemistry data with 5’ v2 chemistry data.
- There are subtleties on exactly how to combine GEX versus CSP versus VDJ data. More details in KB: https://kb.10xgenomics.com/hc/en-us/articles/360045949291

V(D)J updates

- Support for dual indexing (See Gene Expression slides for details)
- Bug fix for annotating TRBD genes. Exact matches to D genes will be annotated.
- New output file: AIRR-compatible tsv file allows users to submit their VDJ data to Adaptive Immune Receptor Repertoire (AIRR) compliant databases.
- Workaround for combining samples via MRO file manipulation is not available
  - Use v3.1 to combine VDJ data
  - Use “aggr” to combine GEX data

10x GENOMICS

V(D)J updates

Enclone

- New flexible command line tool for immune profiling (http://bit.ly/enclone)
- Standalone Beta software. Enclone is not a part of Cell Ranger 4.0
- Limited support provided by developers (enclone@10xgenomics.com)
- What does enclone do?
  - Improves clonotype grouping algorithm
  - Enables exploratory display of clonotypes and phylogenies.
  - Allows optional integration of 5’ Gene Expression or Feature Barcode Data

10x GENOMICS

Appendix C

Updates in Cell Ranger 5.0

Summary of VDJ updates in Cell Ranger 5.0

- “cellranger vdj” pipeline:
  - New clonotype grouping algorithm.
  - New contig and clonotype filters introduced.
- “cellranger multi” will enable running GEX and/or FB with VDJ data.
- ”cellranger aggr” will support combining VDJ data.
- Output evidence of iNKT and MAIT cell types from VDJ data.
- New VDJ references (improved edge cases).

10x GENOMICS

Updates in Loupe VDJ Browser 4

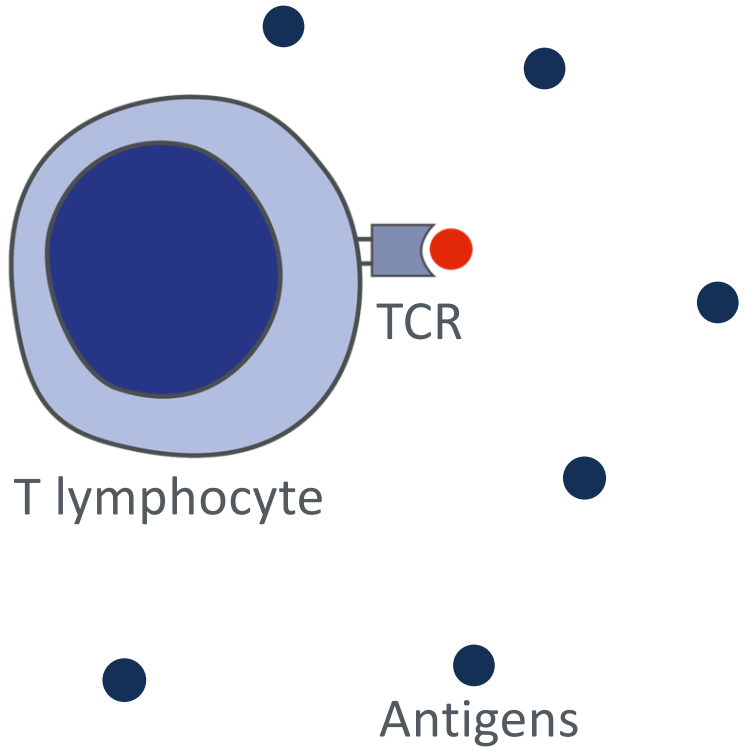
- Not backward compatible. vloupe files by Cell Ranger versions prior to 5.0 will not open in LVB 4
- Interface mostly same as before.
- What is new/updated
  - New filtering based on based on groupings in aggr, iNKT/MAIT cell types, isotype filters
  - For combined plots need to run aggr. Cannot load more than 1 vloupe files (as in earlier versions).
  - Ability to select and slice and dice the database on your grouping of choice
- What is deprecated
  - IGV like view of reads alignment to a chain
  - Clonotype comparison table

10x GENOMICS

Quick Review

Review of concepts

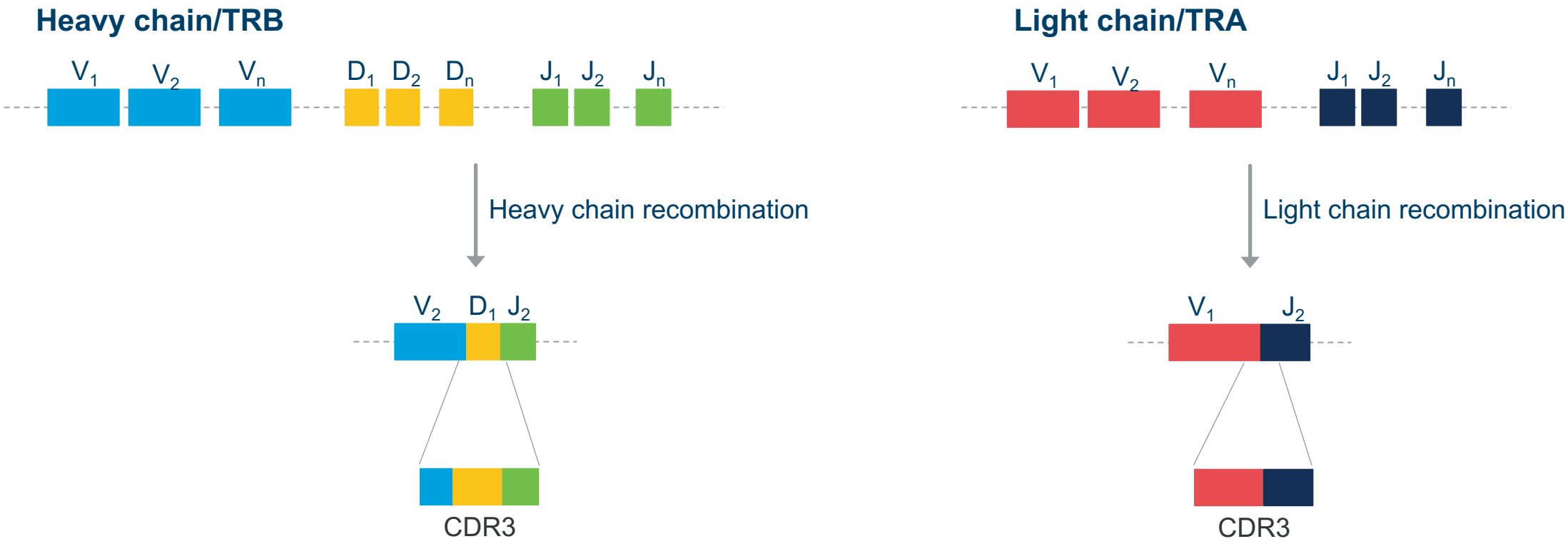
Antigen recognition by immune system



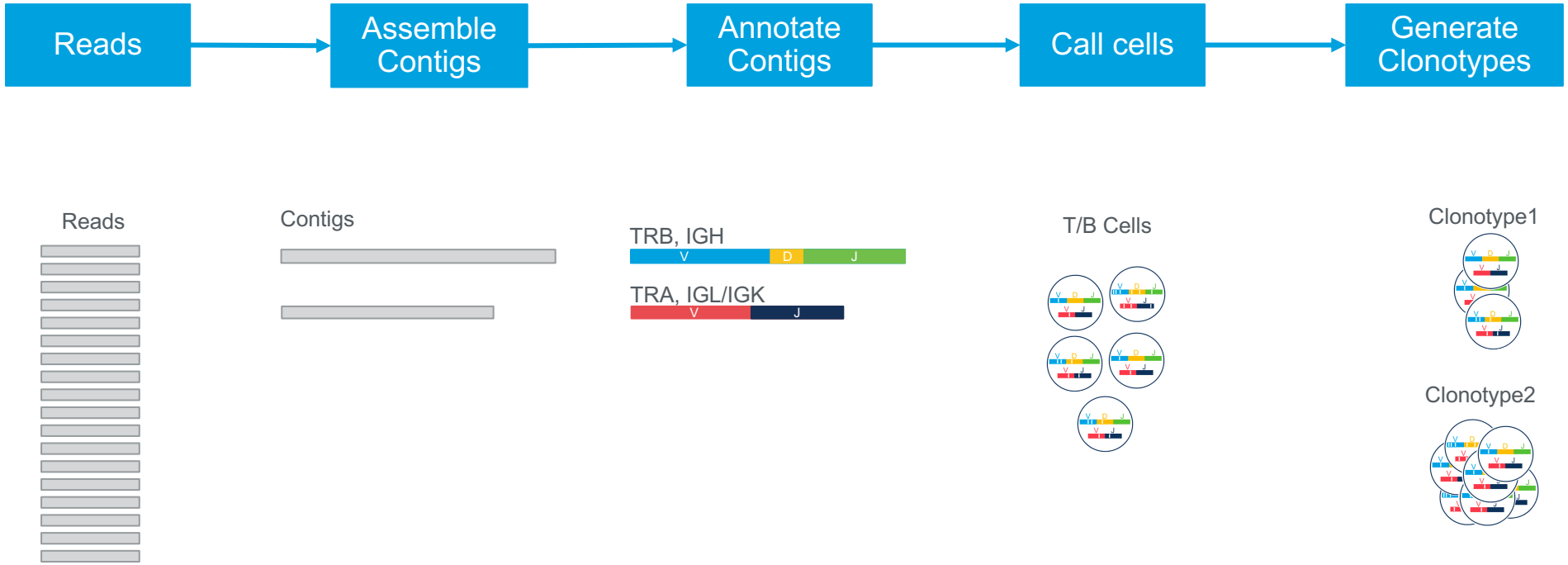
- Antigen receptors on **T and B cells recognize antigens** and mount an immune response
- Antigen receptors (TCR, BCR) are **specific** for antigens they can recognize
- **Immense diversity** of antigen receptors to recognize billions of possible antigens
- This diversity of TCR, BCR is achieved **by VDJ gene recombination** during cell maturation

10x GENOMICS

VDJ recombination



VDJ pipeline summary

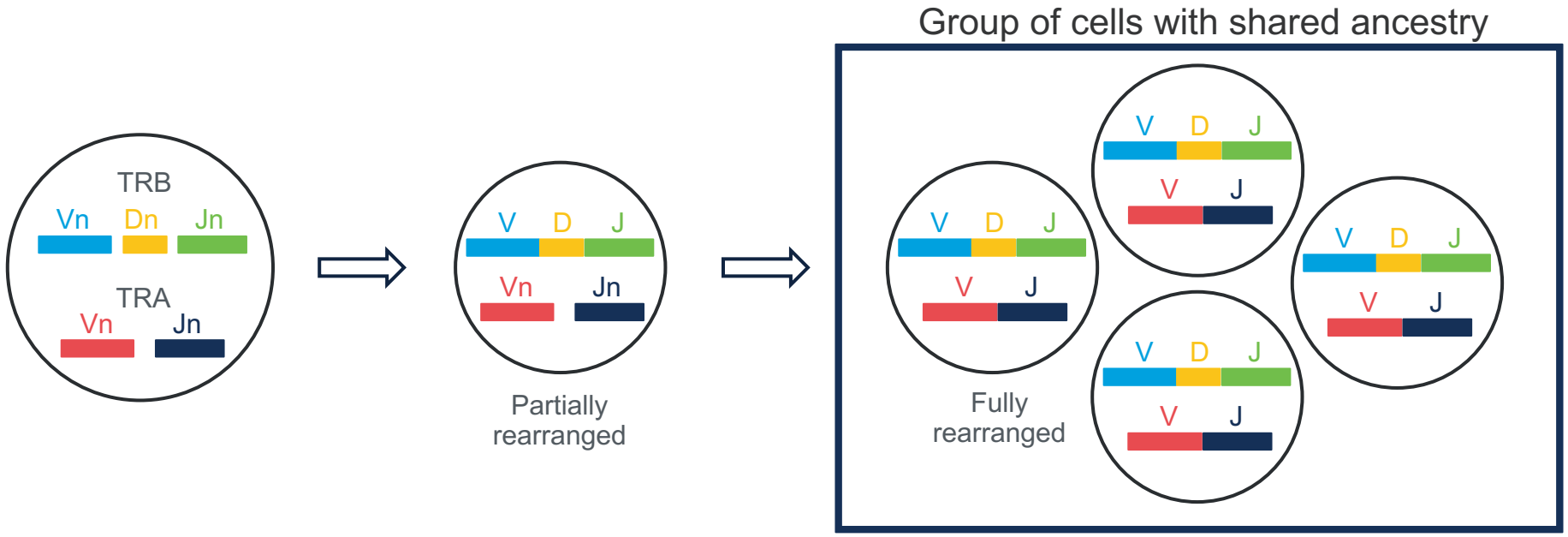


Clonotype Grouping

New clonotype grouping algorithm works correctly on B cells

What is a Clonotype ?

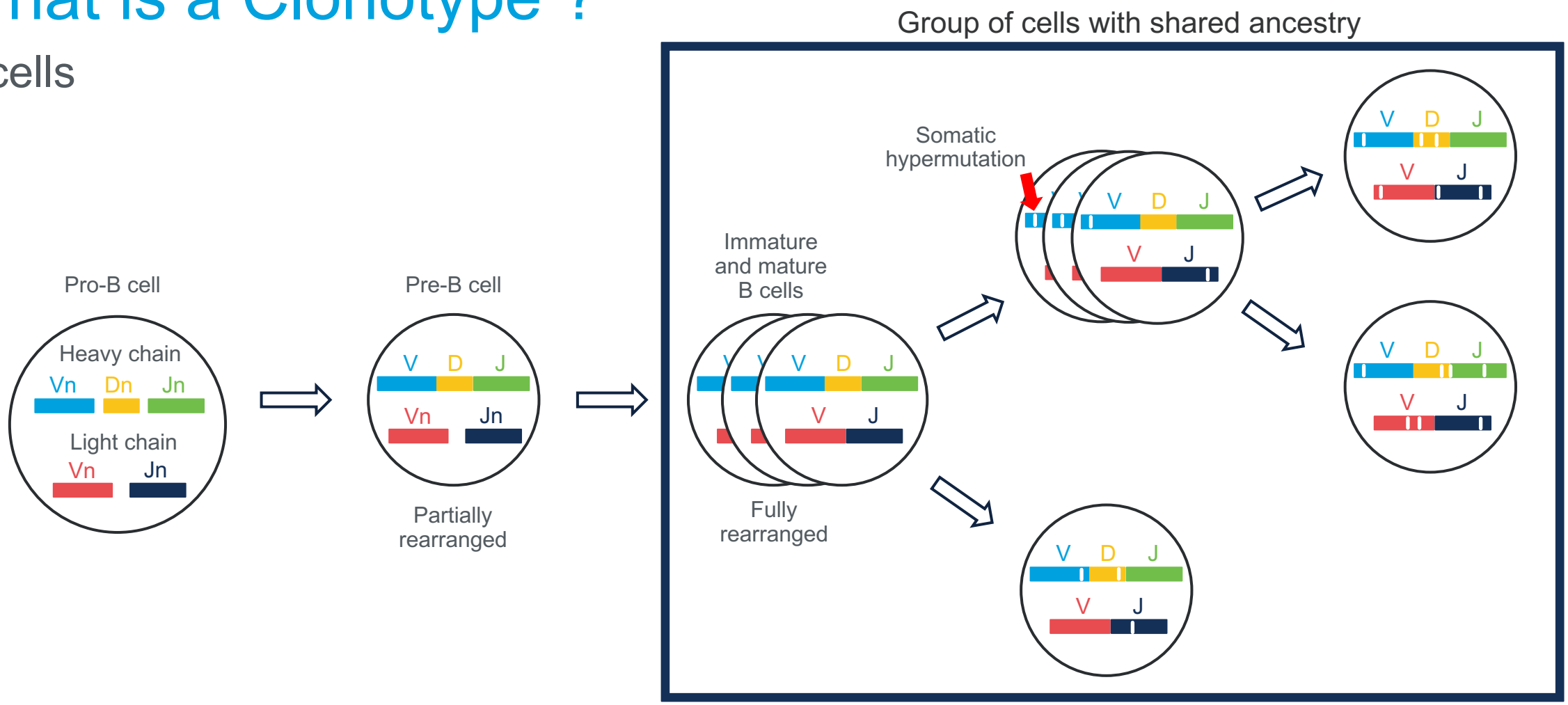
T cells



- The VDJ transcripts in cells in a clonotype are copies of each other
- Cell Ranger v4.0 merges cells based on exact match in CDR3 sequence
- This works okay for T cells

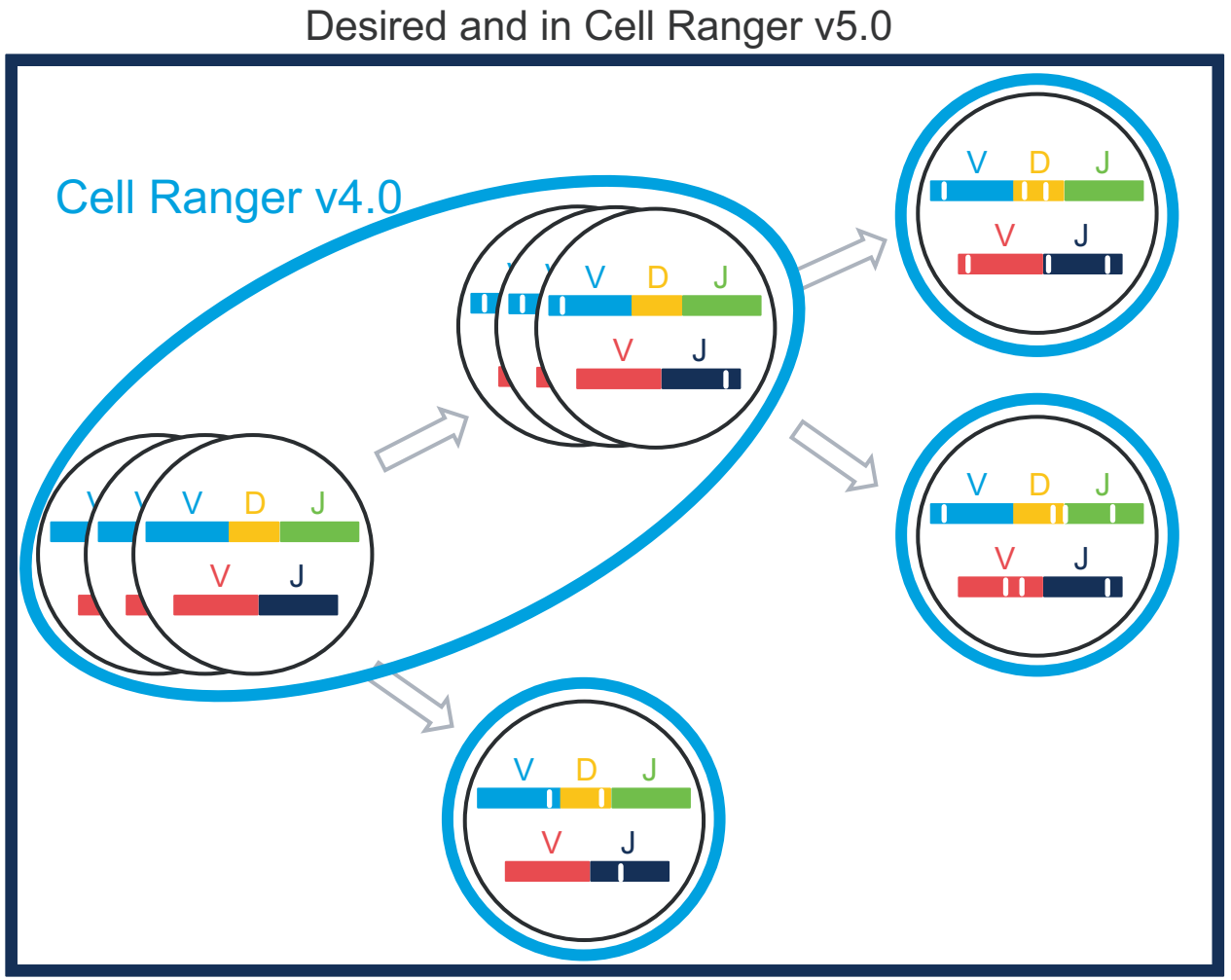
What is a Clonotype ?

B cells



Clonotypes in v4.0 vs v5.0

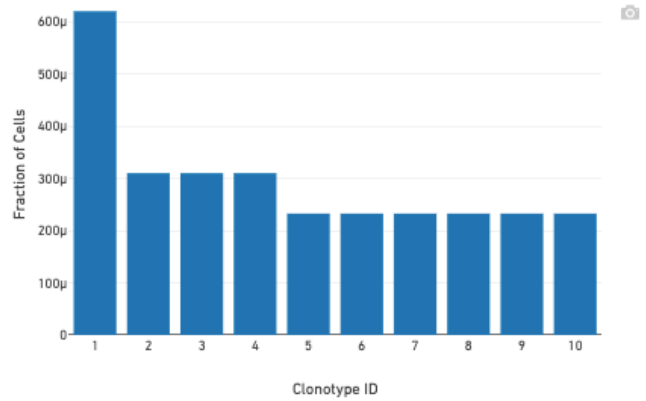
- Desired for B cells
  - Group cells, allowing for arbitrary somatic hypermutation, including in CDR3 regions.
- Cell Ranger v4.0 algorithm
  - Combine cells based on exact matching of CDR3.
  - Assume somatic hypermutations do not occur in CDR3.
    - Works okay for T cells
    - Incorrect for B cells. Somatic hypermutations in CDR3 will lead to cells grouped in different clonotypes



Clonotypes in v4.0 vs v5.0

Cell Ranger v5.0

Top 10 Clonotype Frequencies

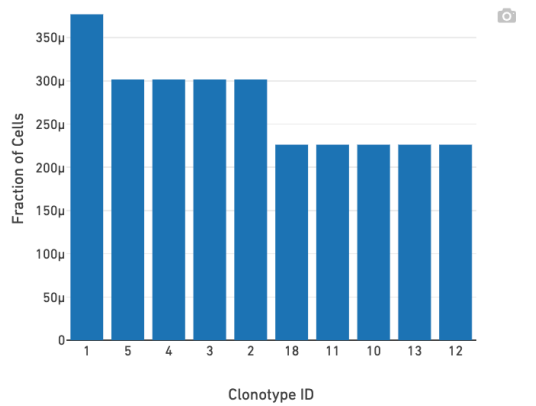


Top 10 Clonotype CDR3 Sequences

Clonotype ID	CDR3s	Frequency	Proportion
1	IGH:CARDGSNDDFWSGYFHWFDPW IGL:CGTWDISLSSVIF	8	0.06%
2	IGH:CARDDYGDYFDYW IGL:CNSRDSSGNHVVF	4	0.03%
3	IGH:CARDSSSSWYGGYSLDYW IGL:CGTWDSSLSAVVF	4	0.03%
4	IGH:CARDGYSGYDWSPDYYYYIDVW IGL:CGTWDISLSAGVF	4	0.03%
5	IGH:CARRGEDYYFDYW IGL:CCSYANIWSLVVL	3	0.02%

Cell Ranger v4.0

Top 10 Clonotype Frequencies

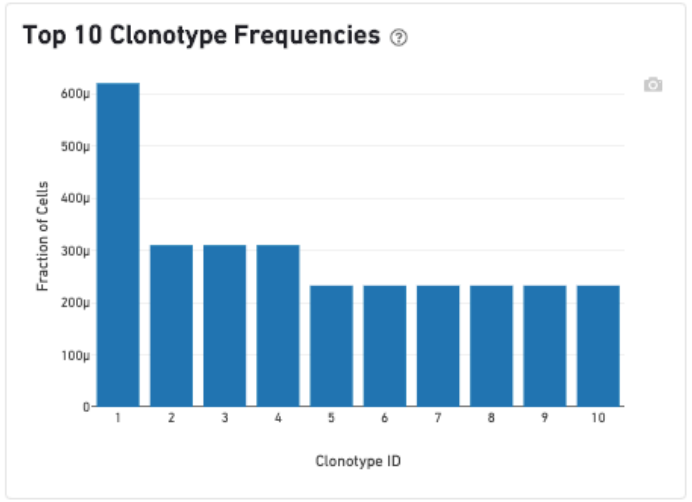


Top 10 Clonotype CDR3 Sequences

Clonotype ID	CDR3s	Frequency	Proportion
1	IGK:CQQYNNWPPLTF	5	0.04%
5	IGH:CAKDGAGPWSPERYYYYYGMDVW IGK:CMQALQTRLTF	4	0.03%
4	IGL:CQAWDSSTVVF	4	0.03%
3	IGH:CARDGSDDDFWNAFFNWFDPW IGL:CGTWDISLSSVIF	4	0.03%
2	IGH:CARDDYGDYFDYW IGL:CNSRDSSGNHVVF	4	0.03%

Biologically correct clonotype in v5.0

Cell Ranger v5.0



Top 10 Clonotype CDR3 Sequences

Clonotype ID	CDR3s	Frequency	Proportion
1	IGH:CARDGSNDDFWSGYFHWFDPW IGL:CGTWDISLSSVIF	8	0.06%
2	IGH:CARDDYGDYFDYW IGL:CNSRDSSGNHVVF	4	0.03%
3	IGH:CARDSSSSWYGGYSLDYW IGL:CGTWDSSLSAVVF	4	0.03%
4	IGH:CARDGYSGYDWSPDYYYYIDVW IGL:CGTWDISLSAGVF	4	0.03%
5	IGH:CARRGEDYYFDYW IGL:CCSYANIWSLVVL	3	0.02%

Clonotypes in Cell Ranger v4.0

Number of Cells	Chain Type	CDR3
4	IGH	CARDGSDDDFWNAFFNWFDPW
	IGL	CGTWDISLSSVIF
2	IGH	CARDGSDDDFWSGYFNWFDPW
	IGL	CGTWDSSLSAVIF
1	IGH	CARDGSNDDFWSGYFHWFDPW
	IGL	CGTWDISLSSVIF
1	IGH	CARDGSDYDFWSGYFNWLDPW
	IGL	CGTWDSSLSAVVF

v5.0 combines cells with somatic hypermutation in the CDR3 unlike v4.0

Clonotype grouping: Take home

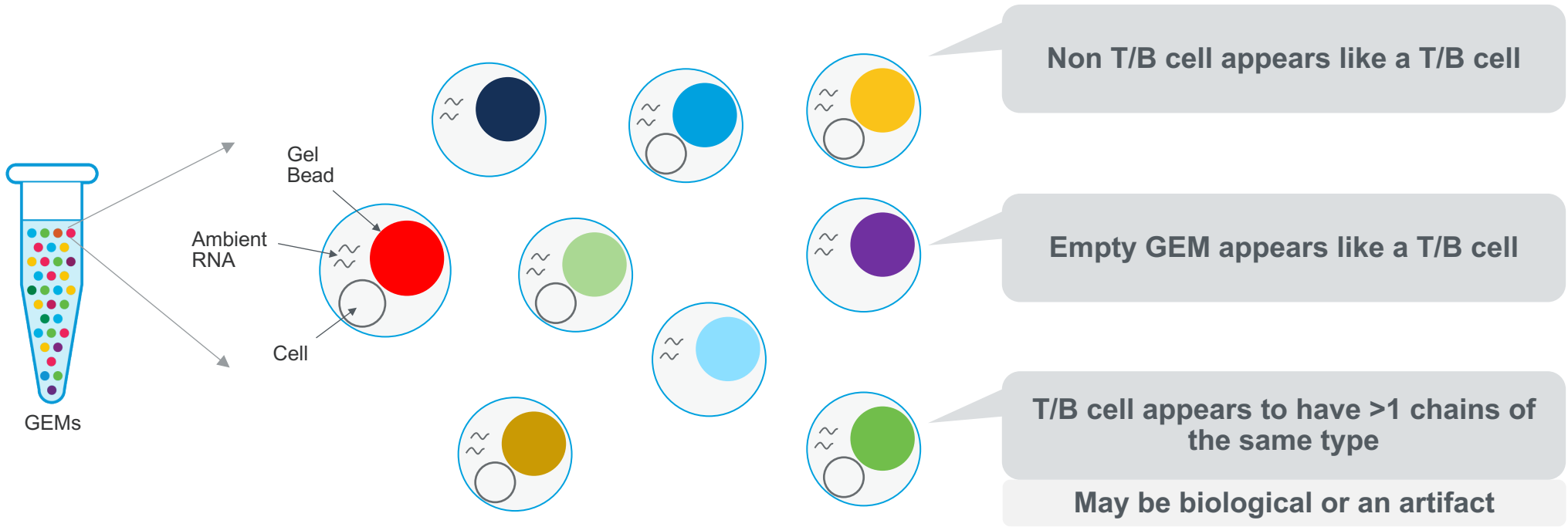
- The B cells clonotypes generated by pipeline are now biologically meaningful.
- Customers working with B cell clonotypes should upgrade to Cell Ranger 5.0.

Improved filters

Improved heuristics to filter out potentially false cell calls and clonotypes

Sources of noise

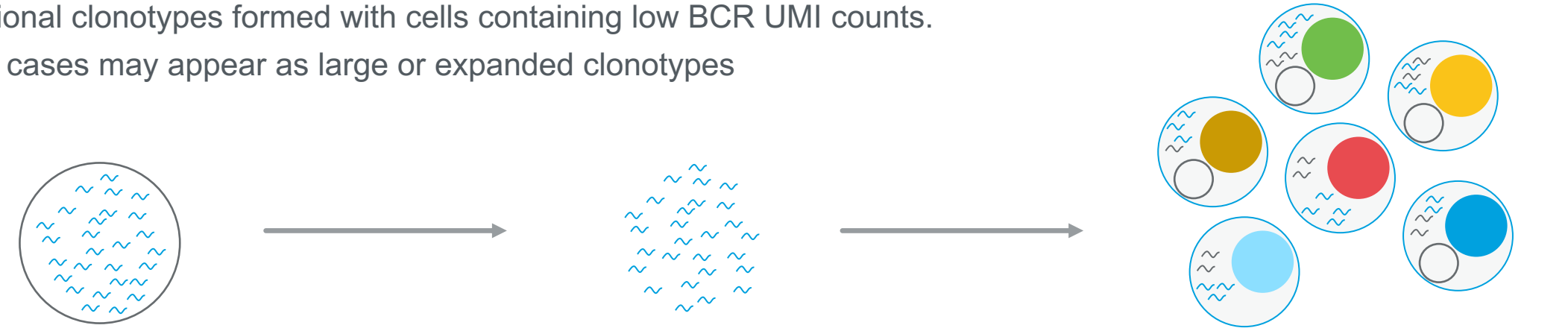
- Background noise in the data, e.g. from ambient mRNA, cell doublets etc.



Effect of ambient mRNA

UMI based filter

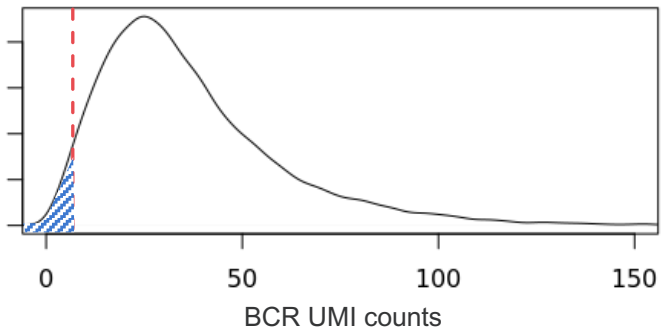
- One of the new filters implemented in the VDJ pipeline to reduce false positives
- What: Filter out low UMI B cells grouped in clonotypes
- Why:
  - Occasional clonotypes formed with cells containing low BCR UMI counts.
  - In rare cases may appear as large or expanded clonotypes



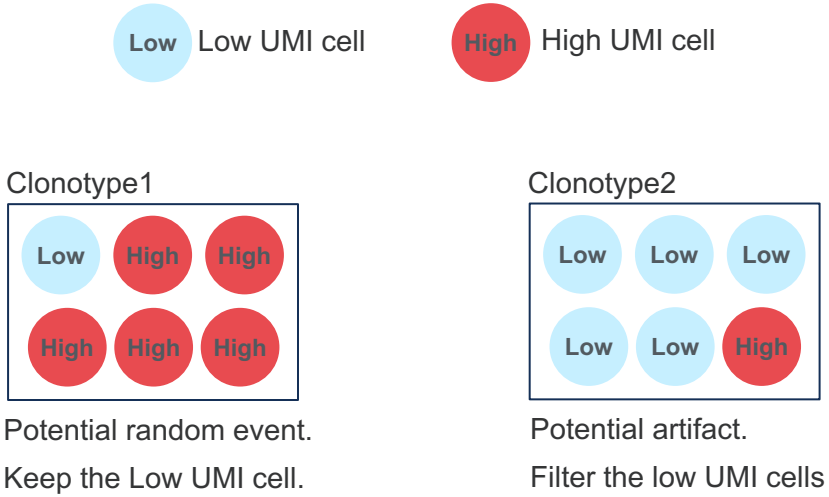
Potential mechanism of low UMI clonotypes

Advanced: UMI based filter

Objective: Filter out low UMI B cells grouped together



Look at clonotypes for set of cells



- Step 1
  - Use distribution of UMI counts in a dataset to determine low UMI threshold
  - Find out low UMI cells based on above threshold
  - These are candidates for removal

- Step 2:
  - Find how many cells with low UMI in a clonotype.
  - Calculate probability that the low UMI cells in the clonotype are a random event
  - More cells with low UMI indicate low probability of random event
  - Filter low UMI cells in the clonotype if p < threshold

Example results

Cell Ranger v4.0

9,245		
Estimated Number of Cells		
9,891	**8,963**	
Mean Read Pairs per Cell	Number of Cells With Productive V-J Spanning Pair	

Top 10 Clonotype CDR3 Sequences

Clonotype ID	CDR3s	Frequency	Proportion
1	IGH:CARDSAFFYGSGSWHDYW IGK:CQQYNNWPRTF	5	0.05%
11	IGH:CTTDRPEGDYIWGSYRTYYYYYGMDVW IGL:CGTWDSSLSGNWVF	3	0.03%
10	IGH:CAKTIGGDSSGRYYYYYGMDVW IGL:CAAWDDSLSGRVVF	3	0.03%
13	IGH:CARQTTSGYSSGWLVVGNWFDPW IGH:CARSPRGSGGPDYW IGK:CQQSYHGGTF IGL:CQVWDSSSDHPGVF	3	0.03%
12	IGH:CATSVDAPGNYW IGL:CSAWDSSLSAWVF	3	0.03%
14	IGH:CARDPDEVLGVAFDYW IGK:CQQYINWPPITF	3	0.03%
9	IGH:CVRGRFKSDYW IGL:CSSYTNINILF	3	0.03%
5	IGH:CARSQTTLTFLDAFDIW IGK:CMQRIEFPCSF	3	0.03%
4	IGH:CARDIISGSYYQGPDYW IGL:CCSYVGSSTLVF	3	0.03%
7	IGH:CARGHHSMDVW IGK:CQQYNNSPGSF	3	0.03%

Reduction in cells reported

- 4 of the 5 cells are filtered out because of the UMI filter.
- Cell with relatively higher UMIs is kept.
- Filtered cells had low total UMI counts in the GEX data and low CD79 expression indicating poor quality signal.

Cell Ranger v5.0

9,048		
Estimated Number of Cells		
10,106	**8,804**	
Mean Read Pairs per Cell	Number of Cells With Productive V-J Spanning Pair	

Top 10 Clonotype CDR3 Sequences

Clonotype ID	CDR3s	Frequency	Proportion
1	IGH:CSRSLNYW IGK:CQQYNSYPRTF	5	0.06%
2	IGH:CATSVAAPGNYW IGL:CSAWDSSLSAWVF	5	0.06%
3	IGH:CARVREGDYFGPLEYW IGK:CQQSYTAPGTF	4	0.04%
4	IGH:CARQAIFAYYYYYYGMDVW IGL:CGTWDSSLSAAVF	3	0.03%
5	IGH:CVRGRFKSDYW IGL:CSSYTNINILF	3	0.03%
6	IGH:CARGHHSMDVW IGK:CQQYNNSPGSF	3	0.03%
7	IGH:CARVFHASSGYFQCW IGK:CQQYNSWPPTF	3	0.03%
8	IGH:CARDPDEVLGVAFDYW IGK:CQQYINWPPITF	3	0.03%
9	IGH:CARTPNYYGTW IGK:CQQYGSSPWTF	3	0.03%
10	IGH:CARDIISGSYYQGPDYW IGL:CCSYVGSSTLVF	3	0.03%

Improved filters: Take home

- Cell Ranger v5.0 comes with improved heuristics that eliminate various false positives
- The filters are especially helpful in improving BCR data quality

multi pipeline

New pipeline to analyze VDJ and GEX data together for streamlined cell calls.

multi pipeline

- Motivation
  - **Simplify analysis** of VDJ data with Gene Expression and/or Feature Barcoding data.
  - Enable more **consistent cell calling** between various library types.

- Note: Detailed slides on *multi* are in separate section outside of Appendix

VDJ aggr

Combine VDJ data from multiple libraries

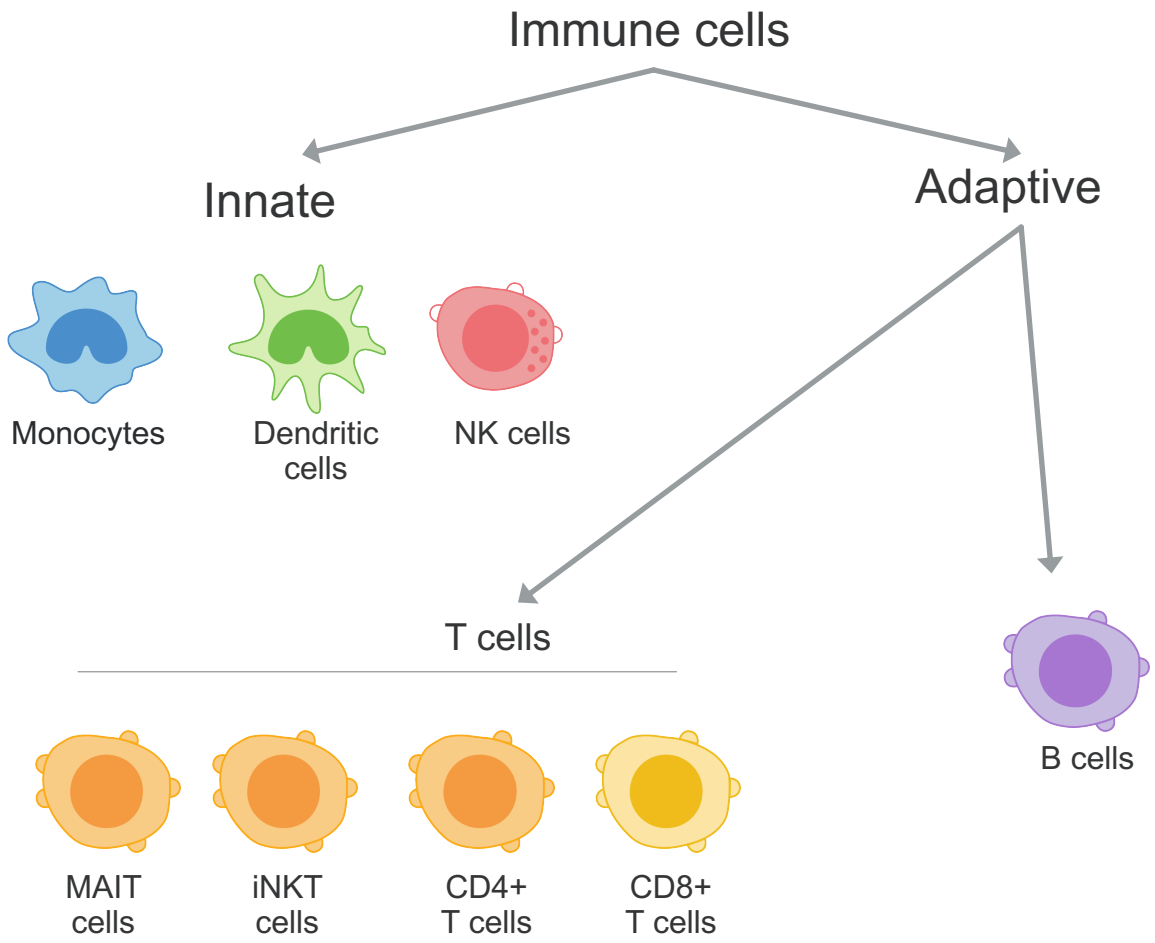
VDJ *aggr*

- "*cellranger aggr*" command supports combining VDJ data from Cell Ranger v5
  - Combine clonotypes
  - Generate a combined vloupe file
- Why ?
  - Analyze clonotypes between experimental conditions
  - Boost the power of clonotype grouping by combining more cells
  - Interactive analysis of combined data in vloupe
- Note: Detailed slides on *aggr* are in separate section outside of Appendix

Other exciting updates

iNKT and MAIT cell types from TCR data

- Annotate iNKT and MAIT cell types from TCR data.
  - iNKT cells (Invariant Natural Killer T cells)
  - MAIT cells (Mucosal Associated Invariant T cells)
- Express near-invariant TRA chain and limited number of TRB chains.
  - Detected based on the gene and junction (CDR3)
  - Example: Human iNKT cells express TRAV10,TRAJ18 genes
- Annotations output in clonotypes.csv file



Recommendations for upgrading to Cell Ranger 5.0

Area of interest	Recommendation	Why
BCR sequencing	Strongly recommend to upgrade	Clonotypes in-line with B cell biology
TCR sequencing	May upgrade	Improved clonotypes, INKT-MAIT annotations
5’ GEX only	Upgrade if using version less than v4.0	Performance improvements since v4.0

Should I re-run my analysis?

- If analysis still pending, strongly recommend analyzing using 5.0
- If partial analysis completed, run all the samples on the same version, preferably 5.0

10x GENOMICS