16th Annual Midsouth Computational Biology & Bioinformatics Society Conference MCBIOS 2019

Single-cell RNA sequencing in Immunology

Christopher Fucile MS

Informatics Institute

University of Alabama at Birmingham

Objective

- **Understand the gene product targeted in VDJ analysis and why a special approach is needed**

- **Understand initial processing of sequence data to generate gene-X-cell and VDJ clonotype data**
  - Be able to run the Cellranger pipeline using the basic command from the terminal
  - Know what output files are needed for additional analysis
  - Identify key QC metrics to check before moving on with further analysis

- **Identify reasons to combine gene expression and VDJ clonotype data using single cell platforms**
  - Know how to navigate the VDJ & Cell Loupe browser functions
  - Be able to integrate the 5' gene expression data with VDJ clonotype data
  - Export sequence data from Cellranger pipeline for analysis in other software packages
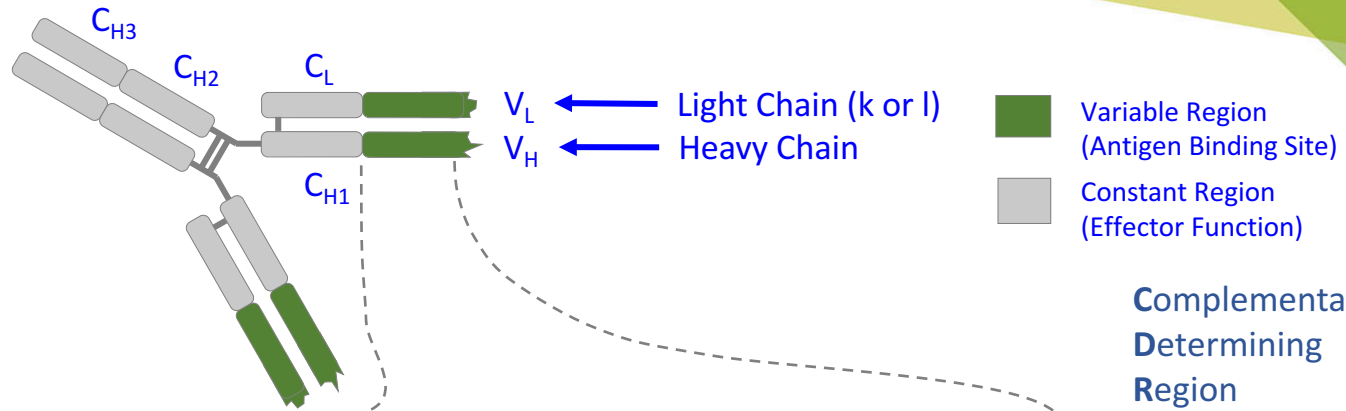
Repertoire Sequencing Applications

- Broad applications in immunology: allergy, autoimmunity, infectious disease, vax
- Establish baseline data of "typical" immune repertoires to understand variability
- Repertoire diversity profile as a readout of immunological status
- Potential readout of antigenic exposure history
- Immune response dynamics
- Lineage questions involving multiple B cell populations
- Identification of potentially convergent Abs in response to vaccine or infection
- Characterizing MRD in B-cell cancers, monitoring toxicity/efficacy of chemo
- Diagnostics – e.g. characterization in autoimmune diseases
- Designing vaccine dose schedules
- Identification of sequences related to particular Ab from single-cell Ab screens
- High throughput mAb identification (via paired V_H:V_L sequencing)
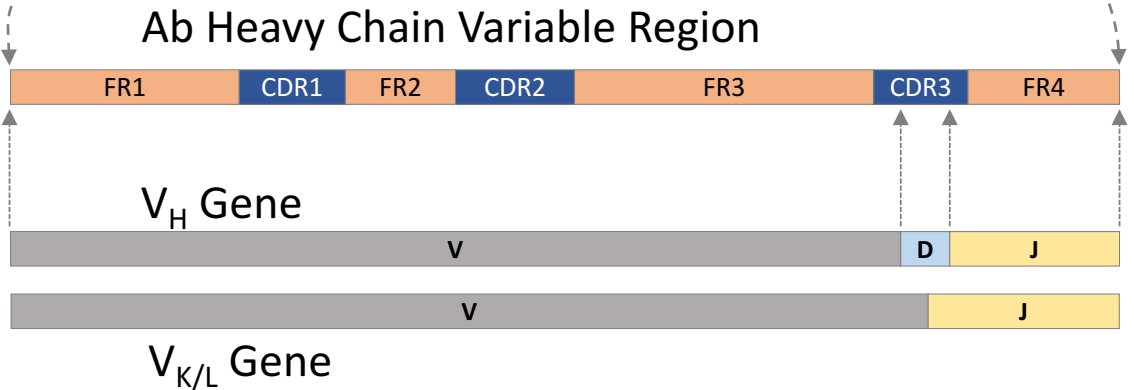- Searching repertoires against databases of sequences with known specificities

Alex Rosenberg, PhD
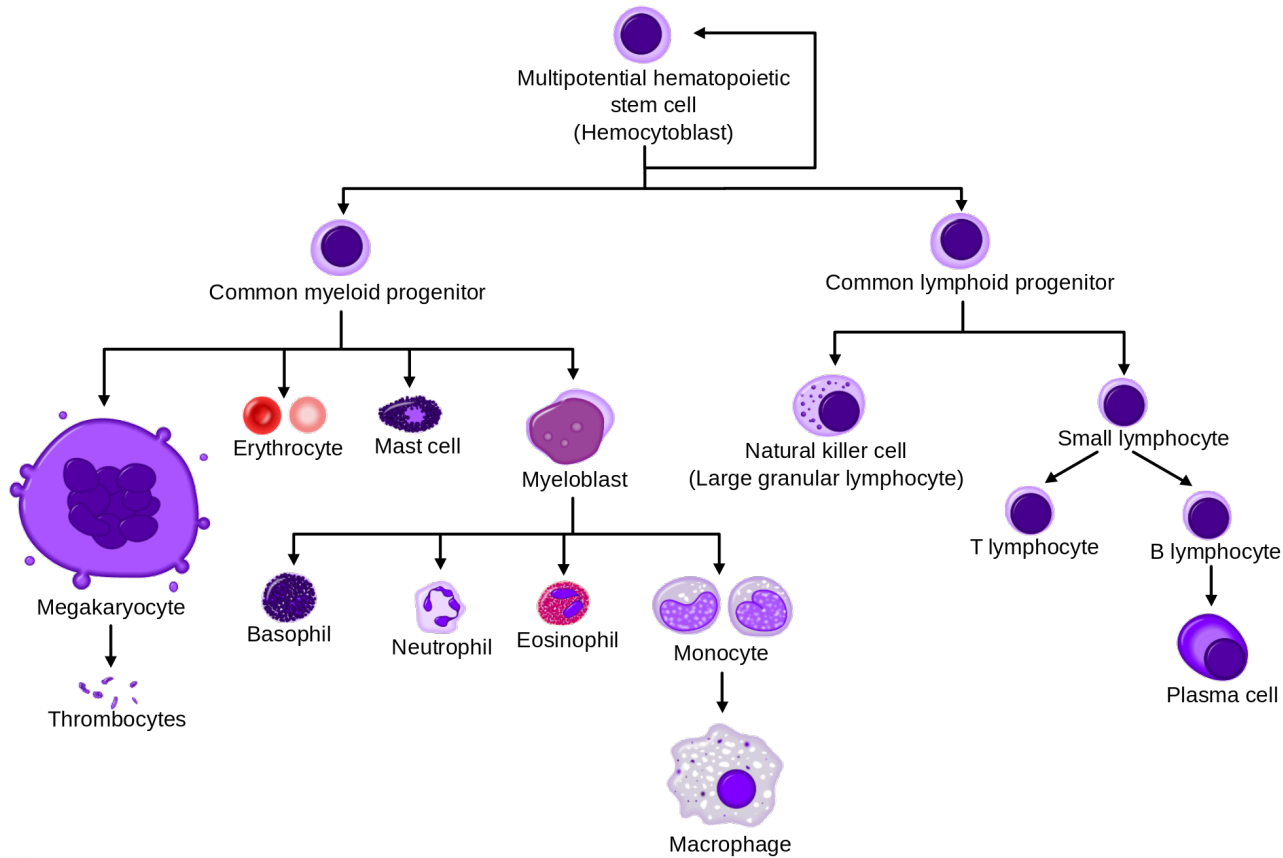
B cell Genes	T cell Genes
CD20	LEF1
CD19	CD3G
PAX5	TCF7
SYK	CD2
CD200	CD28
SPIB	STAT4
IRF4	GATA3
HLA-DMA	MAL
CD72	

Multipotential hematopoietic stem cell (Hemocytoblast)

Common myeloid progenitor

Common lymphoid progenitor

Megakaryocyte

Erythrocyte

Mast cell

Myeloblast

Natural killer cell (Large granular lymphocyte)

Small lymphocyte

Thrombocytes

Basophil

Neutrophil

Eosinophil

Monocyte

T lymphocyte

B lymphocyte

Plasma cell

Macrophage

THE UNIVERSITY OF ALABAMA AT BIRMINGHAM
Knowledge that will change your world

C_{H3}

C_{H2}

C_L

C_{H1}

V_L ← Light Chain (k or l)

V_H ← Heavy Chain

Variable Region
(Antigen Binding Site)

Constant Region
(Effector Function)

Complementarity
Determining
Region
Framework
Region

Ab Heavy Chain Variable Region

| FR1 | CDR1 | FR2 | CDR2 | FR3 | CDR3 | FR4 |

~120 residues

V_H Gene

| V | D | J |

Encoding gene segments from IGH/K/L@

$V_{K/L}$ Gene

| V | J |

CDRs are hypervariable regions, with CDR3 having the most variability

Antigen Receptor Diversity

1. Combinatorial Diversity

   somatic recombination during B cell development

2. Junctional Diversity

   joining variability during somatic recombination (CDR3)

3. Combinatorial Association of Heavy and Light Chains

4. Somatic Hypermutation (B-cells)

   following activation, during affinity maturation

10x Genomics © Cellranger © & Loupe Browser © Overview

Dataset:

      PBMCs of healthy donor – 5' gene expression &
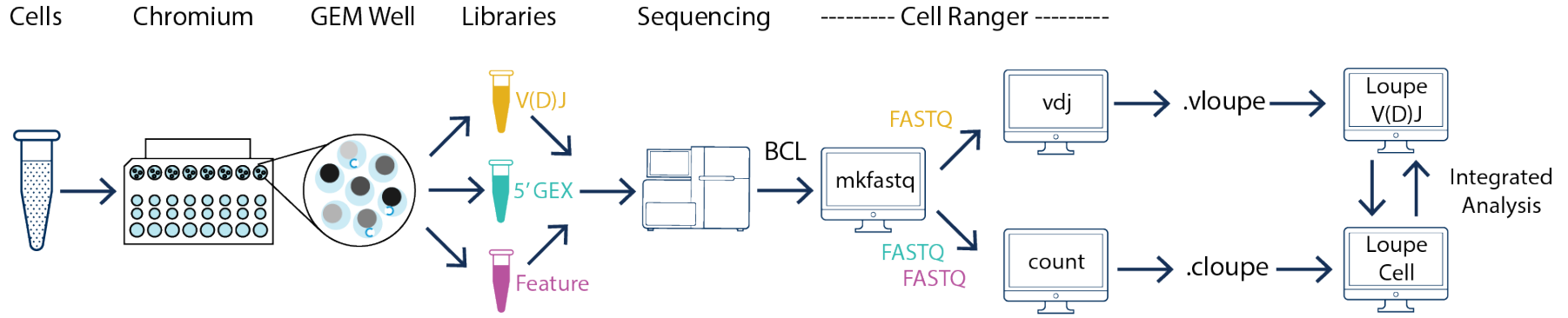
      Ig enrichment from amplified cDNA

      https://www.10xgenomics.com/solutions/vdj/#datasets

Software:

      Cellranger©

      Loupe Cell Browser©

      Loupe V(D)J Browser©

      https://support.10xgenomics.com/single-cell-vdj/software/downloads/latest

THE UNIVERSITY OF
ALABAMA AT BIRMINGHAM
Knowledge that will change your world

Cellranger ©

Cellranger © mkfastq

```
cellranger mkfastq --run=/path/to/bcl/files --csv=sample_layout.csv --qc
```

sample_layout.csv

```
Lane,Sample,Index
1,Sample1,SI-GA-A3
```

outs/qc_summary.json

```
"sample_qc": {
  "Sample1": {
    "3": {
      "barcode_exact_match_ratio": 0.9336158258904611,
      "barcode_q30_base_ratio": 0.9611993091728814,
      "bc_on_whitelist": 0.9447542078230667,
      "mean_barcode_qscore": 37.770630795934,
      "number_reads": 2748155,
      "read1_q30_base_ratio": 0.8947676653366835,
      "read2_q30_base_ratio": 0.7771883245304577
    }
  }
}
```

THE UNIVERSITY OF
ALABAMA AT BIRMINGHAM
Knowledge that will change your world

Cellranger © count

```
cellranger count --id=sample345 \
        --transcriptome=/opt/refdata-cellranger-GRCh38-3.0.0 \
        --fastqs=/home/jdoe/runs/HAWT7ADXX/outs/fastq_path \
        --sample=mysample  --expect-cells=1000
```

File Name	Description
web_summary.html	Run summary metrics and charts in HTML format
metrics_summary.csv	Run summary metrics in CSV format
possorted_genome_bam.bam	Reads aligned to the genome and transcriptome annotated with barcode information
possorted_genome_bam.bam.bai	Index for possorted_genome_bam.bam
filtered_feature_bc_matrix	Filtered feature-barcode matrices containing only cellular barcodes in MEX format
filtered_feature_bc_matrix_h5.h5	Filtered feature-barcode matrices containing only cellular barcodes in HDF5 format
raw_feature_bc_matrices	Unfiltered feature-barcode matrices containing all barcodes in MEX format
raw_feature_bc_matrix_h5.h5	Unfiltered feature-barcode matrices containing all barcodes in HDF5 format
analysis	Secondary analysis data including dimensionality reduction, cell clustering, and differential expression
molecule_info.h5	Molecule-level information used by cellranger aggr to aggregate samples into larger datasets.
cloupe.cloupe	Loupe Cell Browser visualization and analysis file

THE UNIVERSITY OF
ALABAMA AT BIRMINGHAM
Knowledge that will change your world

Cellranger © VDJ

```
cellranger vdj --id=sample345
         --transcriptome=/opt/refdata-cellranger-GRCh38-3.0.0
         --fastqs=/home/jdoe/runs/HAWT7ADXX/outs/fastq_path
         --sample=mysample  --expect-cells=1000
```

File Name	Description
concat_ref.fasta	Concatenated reference sequences
concat_ref.fasta.fai	Concatenated reference index
consensus.bam	Contig-consensus alignments
consensus.bam.bai	Contig-consensus alignment index
concat_ref.bam	Contig-reference alignments
concat_ref.bam.bai	Contig-reference alignment index
consensus_annotations.json	Clonotype consensus annotations (JSON)
consensus_annotations.csv	Clonotype consensus annotations (CSV)
clonotypes.csv	Clonotype info
vloupe.vloupe	Loupe V(D)J Browser file

File Name	Description
web_summary.html	Run summary HTML
metrics_summary.csv	Run summary CSV
all_contig.fasta	All-contig FASTA
all_contig.fasta.fai	All-contig FASTA index
all_contig.fastq	All-contig FASTQ
all_contig.bam	Read-contig alignments
all_contig.bam.bai	Read-contig alignment index
all_contig_annotations.json	All contig annotations (JSON)
all_contig_annotations.bed	All contig annotations (BED)
all_contig_annotations.csv	All contig annotations (CSV)
filtered_contig.fasta	Filtered contig sequences FASTA
filtered_contig.fastq	Filtered contig sequences FASTQ
filtered_contig_annotations.csv	Filtered contigs (CSV)
consensus.fasta	Clonotype consensus FASTA
consensus.fasta.fai	Clonotype consensus FASTA index
consensus.fastq	Clonotype consensus FASTQ

THE UNIVERSITY OF
ALABAMA AT BIRMINGHAM
Knowledge that will change your world