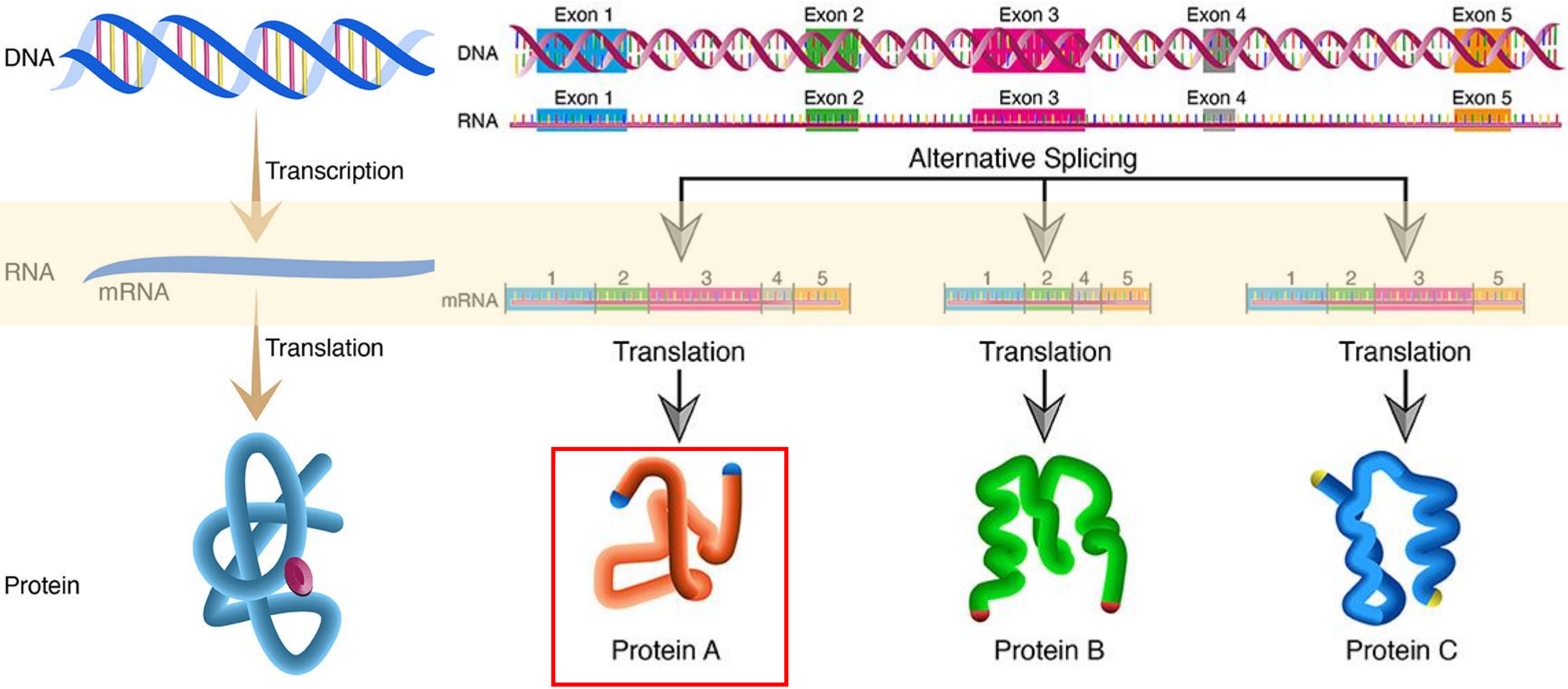
CITE-seq & TCR/BCR

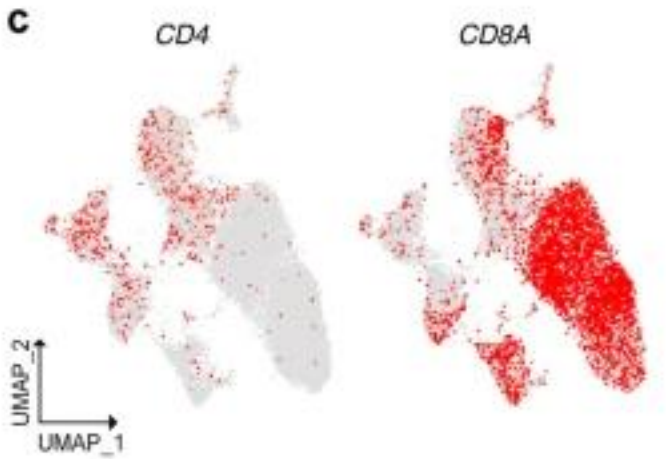
- CITE-seq

-Discrepancy between protein and RNA expression



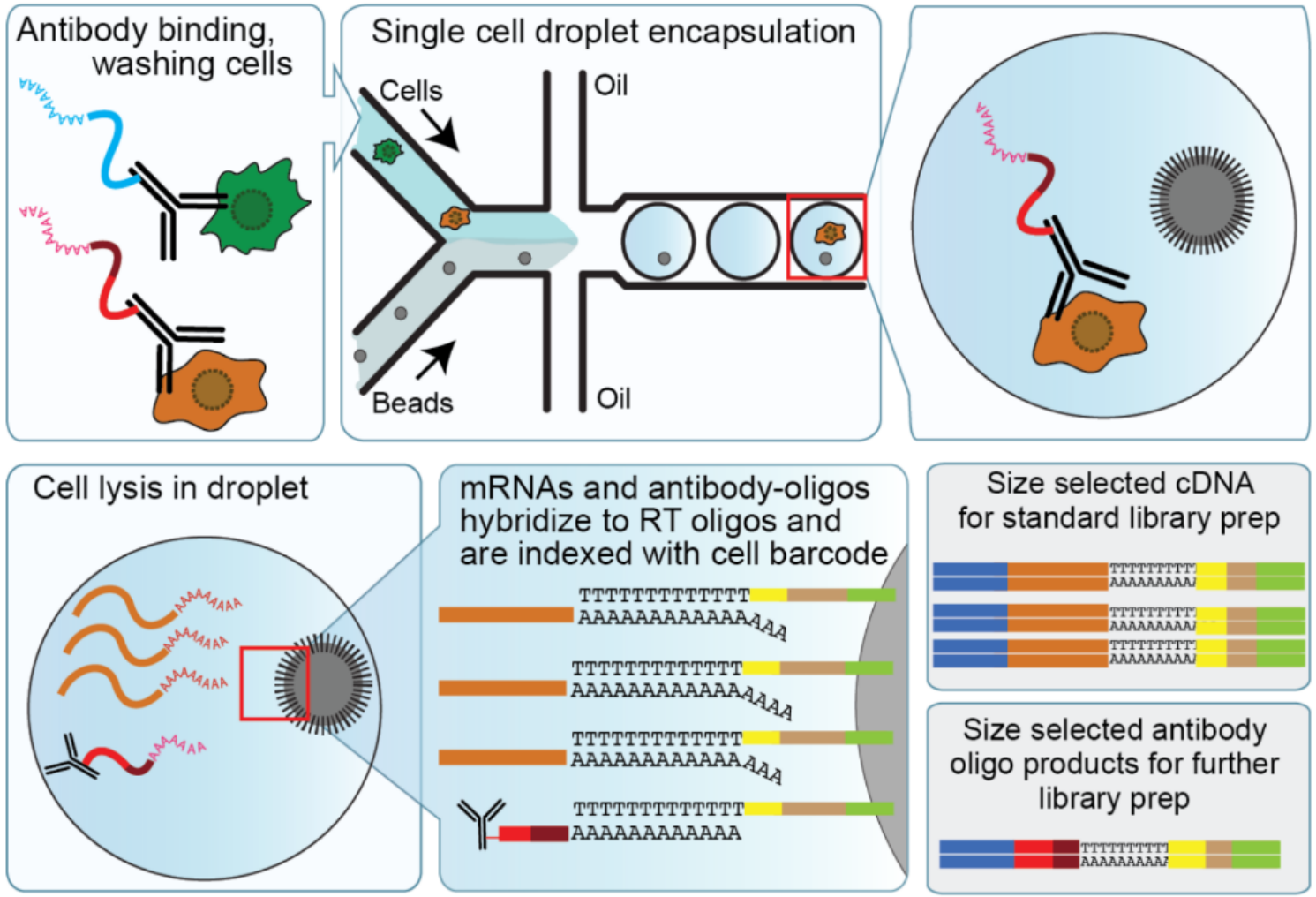
- CITE-seq

*Some bias in a certain category of proteins
Ex) receptor (ex: CD4)
→ Does not have to express RNA continuously



Single-cell analyses of Crohn's disease tissues reveal intestinal intraepithelial T cells heterogeneity and altered subset distributions

- CITE-seq



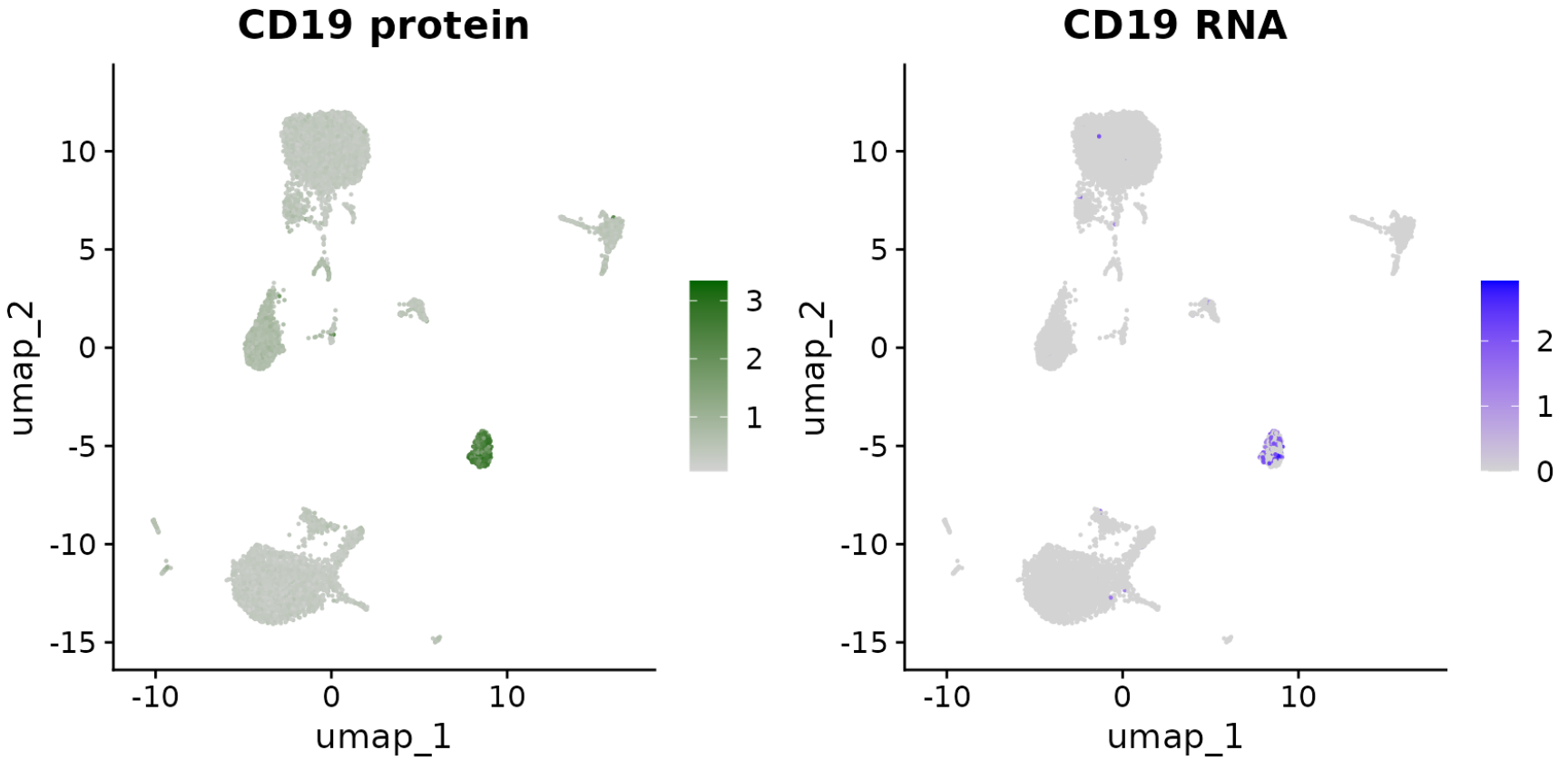
*Multi-modal approach: RNA + **Protein**!

Leveraging sequencing strategy for protein detection

→ Unique “barcode” for antibody

→ Antibody captures protein

- CITE-seq



- CITE-seq (WNN)

*Weighted nearest neighbor analysis
-Mulitmodal integration analysis
-Incorporate (two) modalities (cell-specific weight)

(1) Constructing independent k-nearest neighbor (KNN) graphs for both modalities.
(from PCA embedding)
cf) protein: CLR (centered log ratio) normalization

(2) Performing within and across-modality prediction
Rknn: average of the low-dimensional profile from each neighbor set
R: RNA, P: protein

Within-modality prediction:

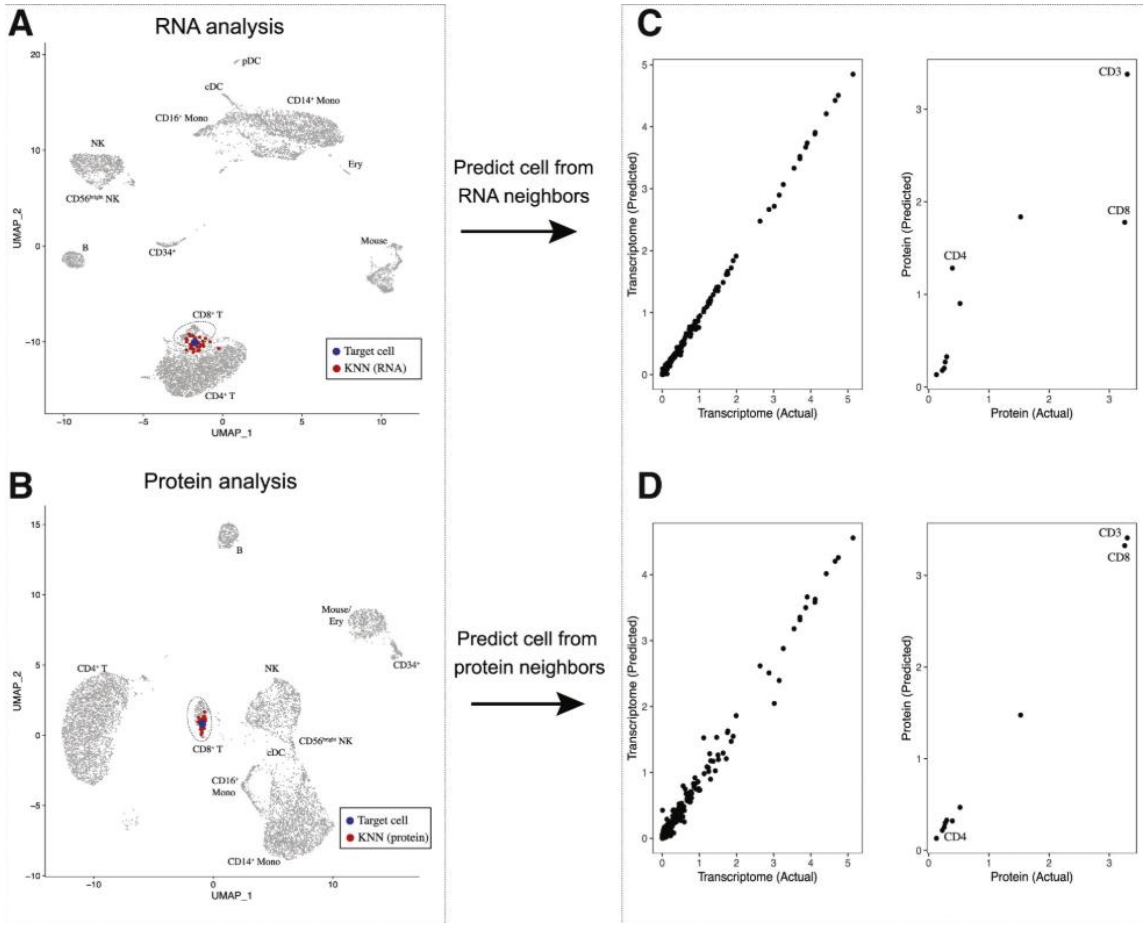
$$\hat{r}_{i,knn_r} = \frac{\sum_{j=1}^{k} r_{knn_r,i,j}}{k}$$

$$\hat{p}_{i,knn_p} = \frac{\sum_{j=1}^{k} p_{knn_p,i,j}}{k}$$

Cross-modality prediction:

$$\hat{r}_{i,knn_p} = \frac{\sum_{j=1}^{k} r_{knn_p,i,j}}{k}$$

$$\hat{p}_{i,knn_r} = \frac{\sum_{j=1}^{k} p_{knn_r,i,j}}{k}$$



- CITE-seq (WNN)

(3) Calculating cell-specific modality weights.
Affinity measurement between (by exponential kernel)
Ri and Rknn from RNA, Rknn from protein
Pi and Pknn from RNA, Pknn from protein
(Band width optimization)
→Weight: Soft max with affinity ratio

Closest Euclidian distance

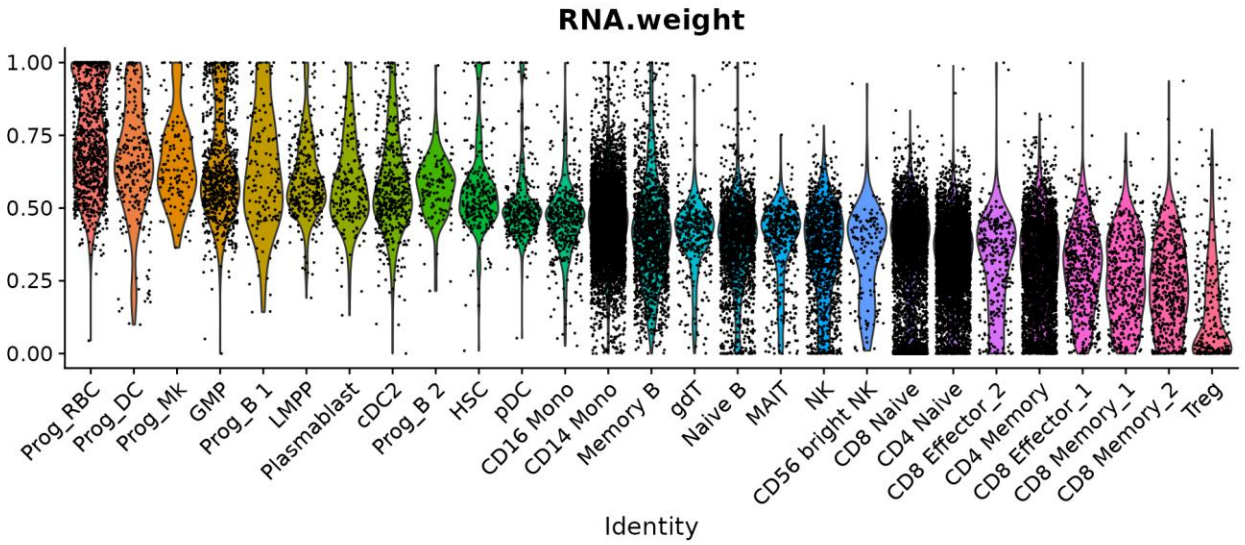
$$\theta_{rna}(r_i, \hat{r}_{i,knn_r}) = \exp\left(-\frac{\max\left(d(r_i,\hat{r}_{i,knn_r}) - d\left(r_i, r_{knn_{r,i,1}}\right), 0\right)}{\sigma_{r,i} - d\left(r_i, r_{knn_{r,i,1}}\right)}\right)$$

$$s_{rna}(i) = \frac{\theta_{rna}(r_i,\hat{r}_{i,knn_r})}{\theta_{rna}(r_i,\hat{r}_{i,knn_p})+\varepsilon}, \quad s_{protein}(i) = \frac{\theta_{protein}(p_i,\hat{p}_{i,knn_p})}{\theta_{protein}(p_i,\hat{p}_{i,knn_r})+\varepsilon}$$

$$w_{rna}(i) = \frac{e^{s_{rna}(i)}}{e^{s_{rna}(i)}+e^{s_{protein}(i)}}, \quad w_{protein}(i) = \frac{e^{s_{protein}(i)}}{e^{s_{rna}(i)}+e^{s_{protein}(i)}}$$

(4) Calculating a WNN graph (cell-cell relation)
→ Generate KNN graph by weighted similar matrix

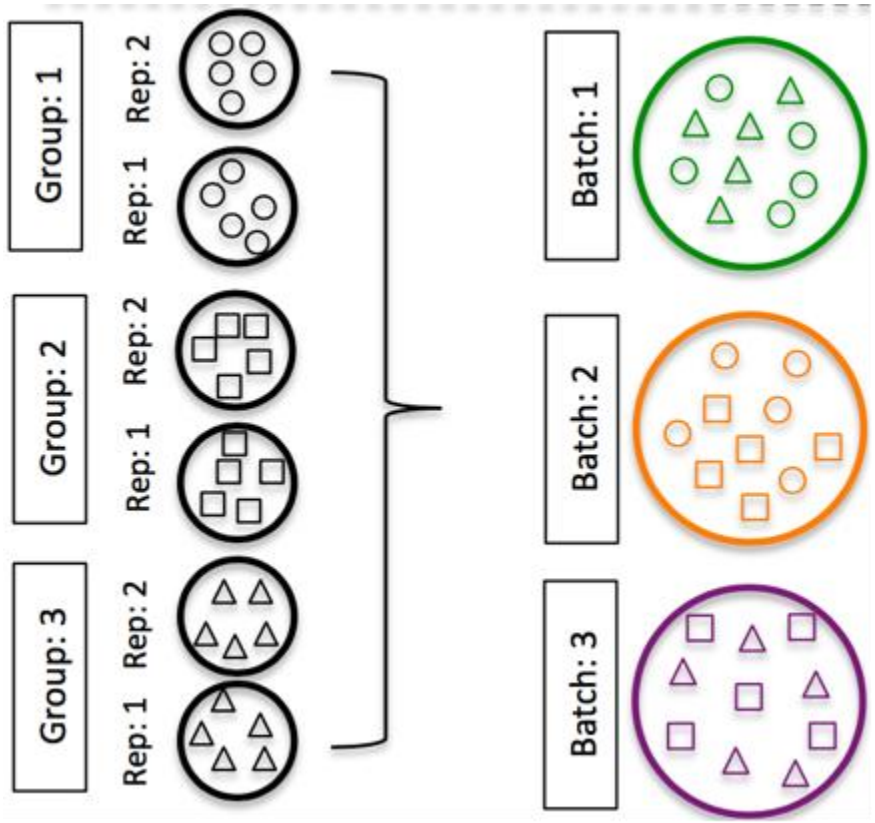
$$\theta_{weighted}(i,j) = w_{rna}(i)\,\theta_{rna}(r_i,r_j) + w_{protein}(i)\,\theta_{protein}(p_i,p_j)$$

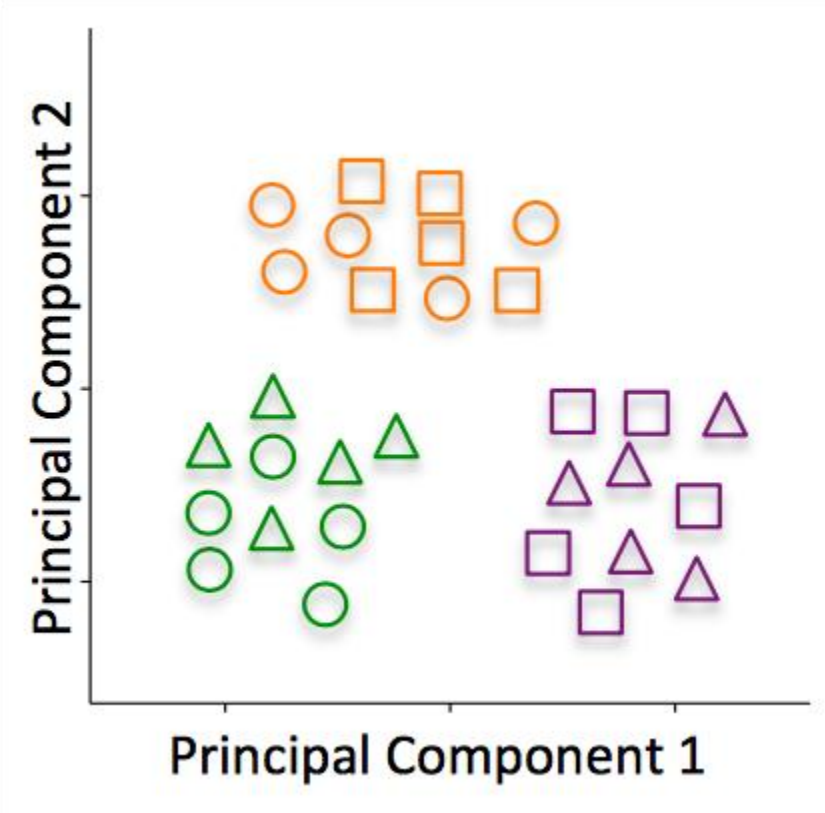


(5) More than 2 modalities:
all-pairwise processing above

- Hash-tagging

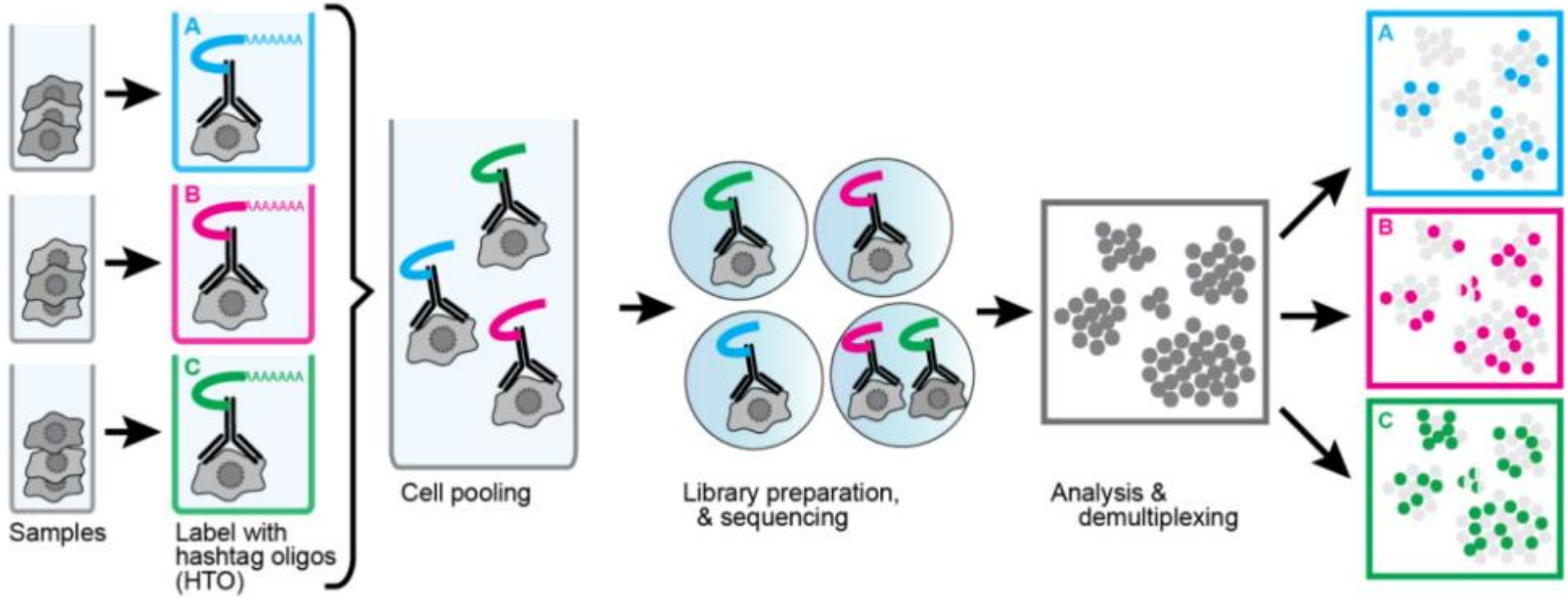
-Multiple samples → batch effect



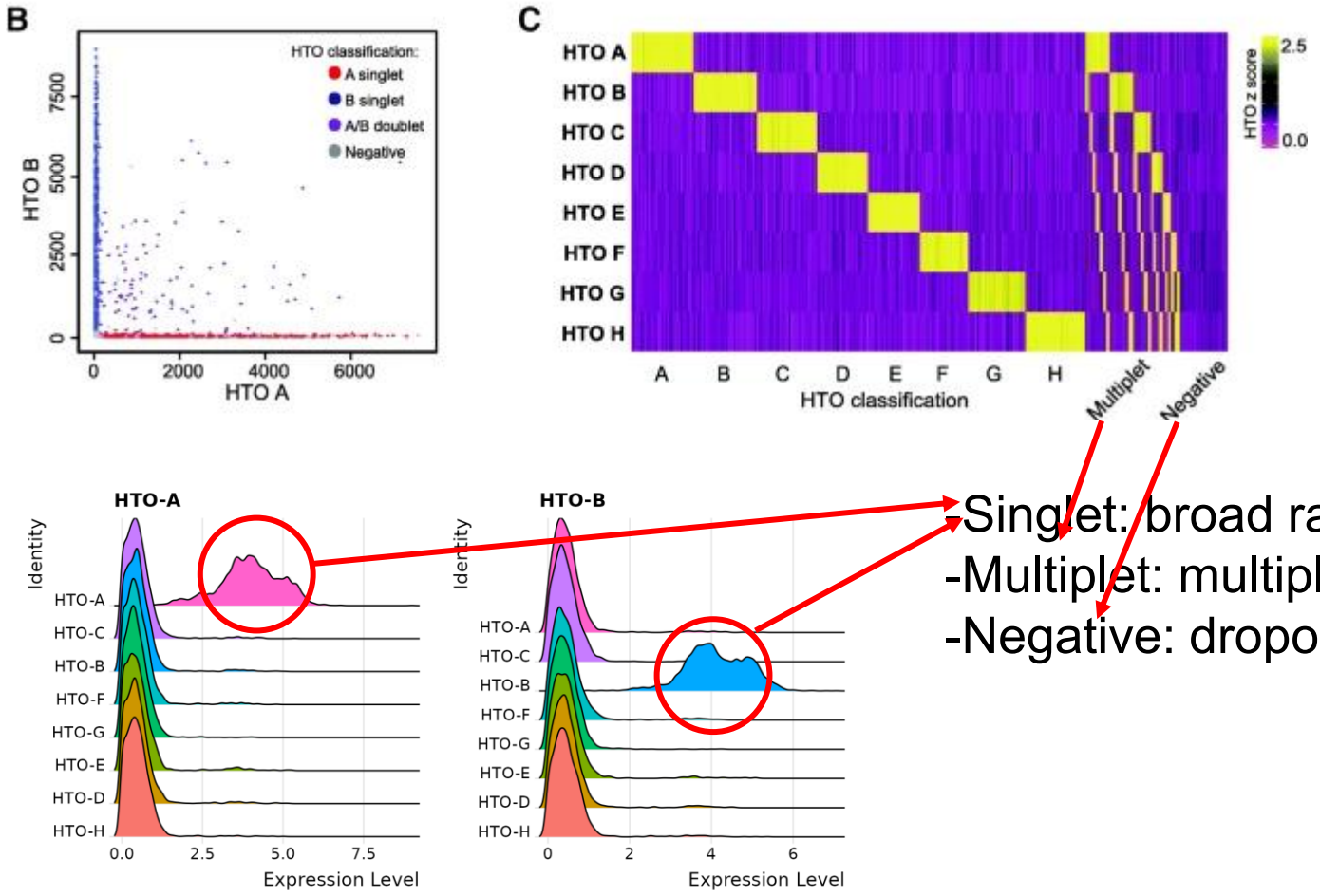


- Hash-tagging

-Multiple samples → hash tagging → one experiment → reduce technical batch effect
-Leverage the strategy from CITE-seq
-Oligo-tagged antibodies against ubiquitously expressed surface proteins

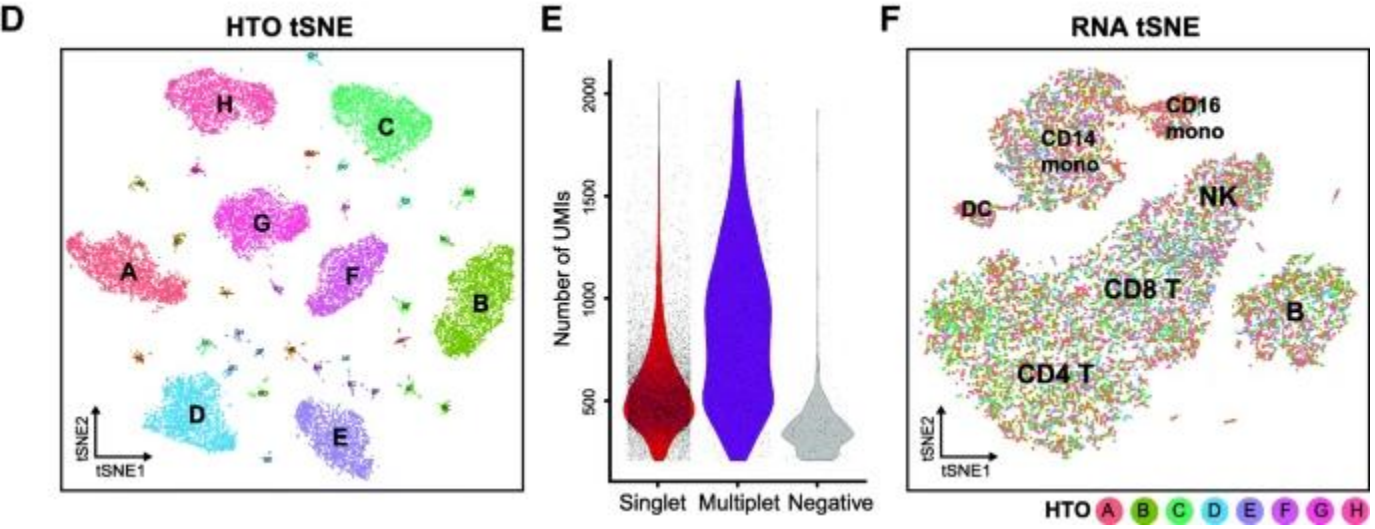


- Hash-tagging



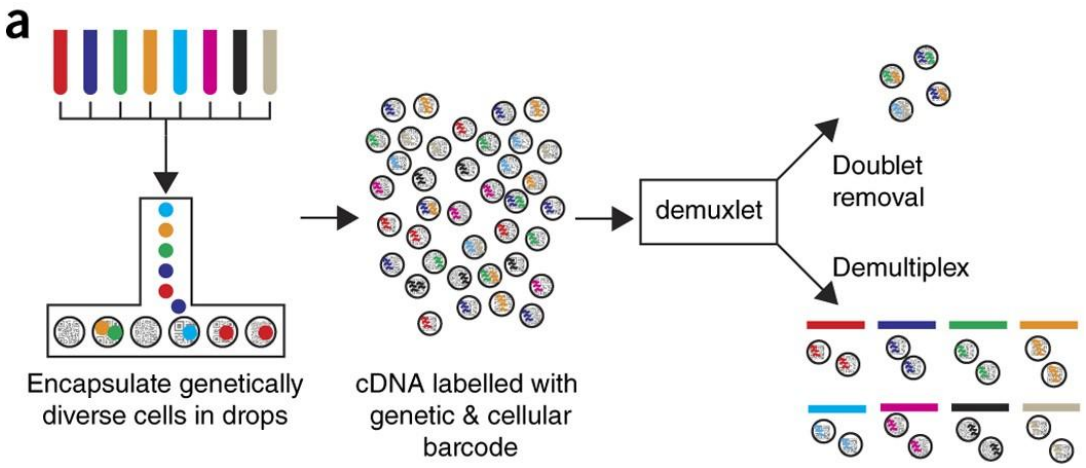
-Singlet: broad range
-Multiplet: multiple tagging → non-specific binding
-Negative: dropout

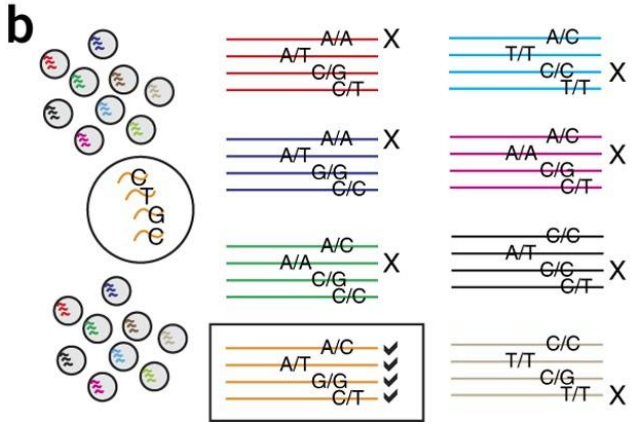
- Hash-tagging

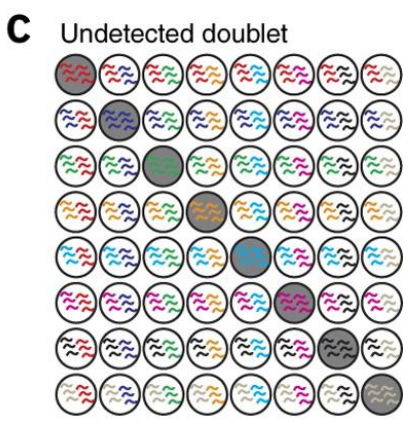


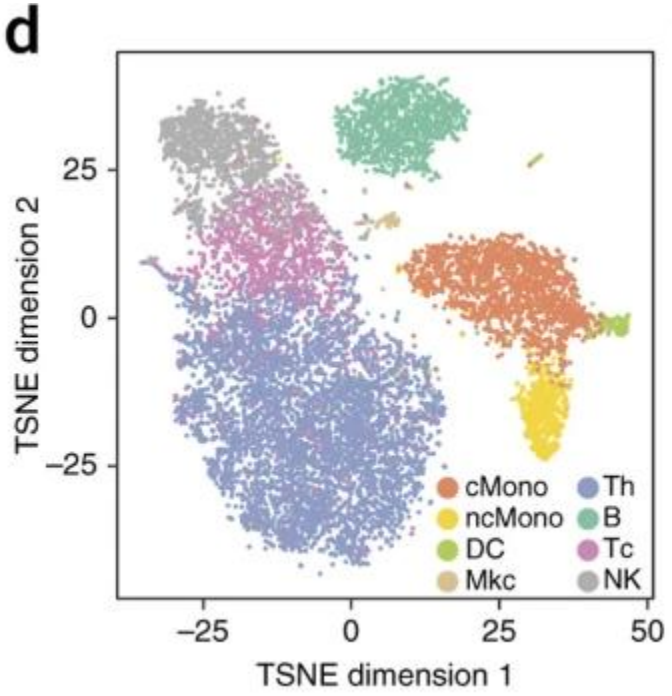
-Big island: Singlet → each sample
-Small island: multiplets or dropout → distinctive profile
-Well mixed from cell type level

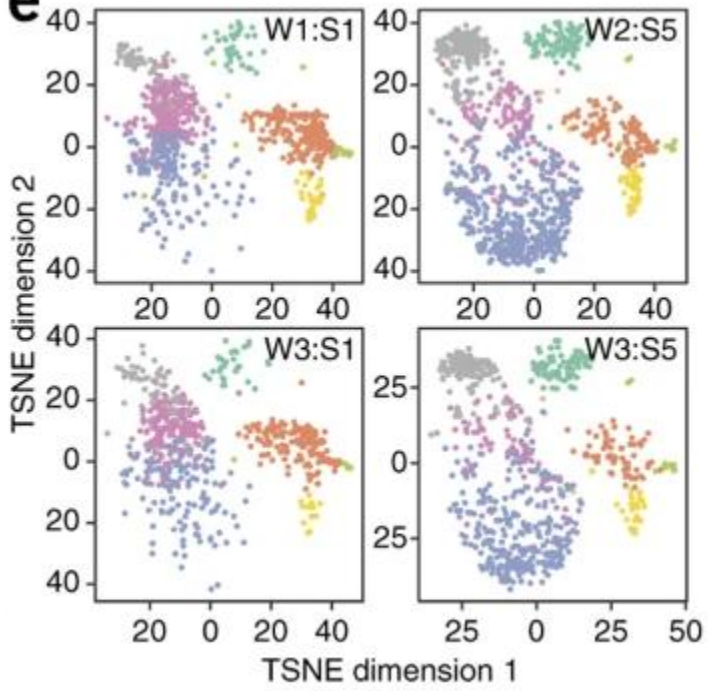
- Demuxlet, Souporcell





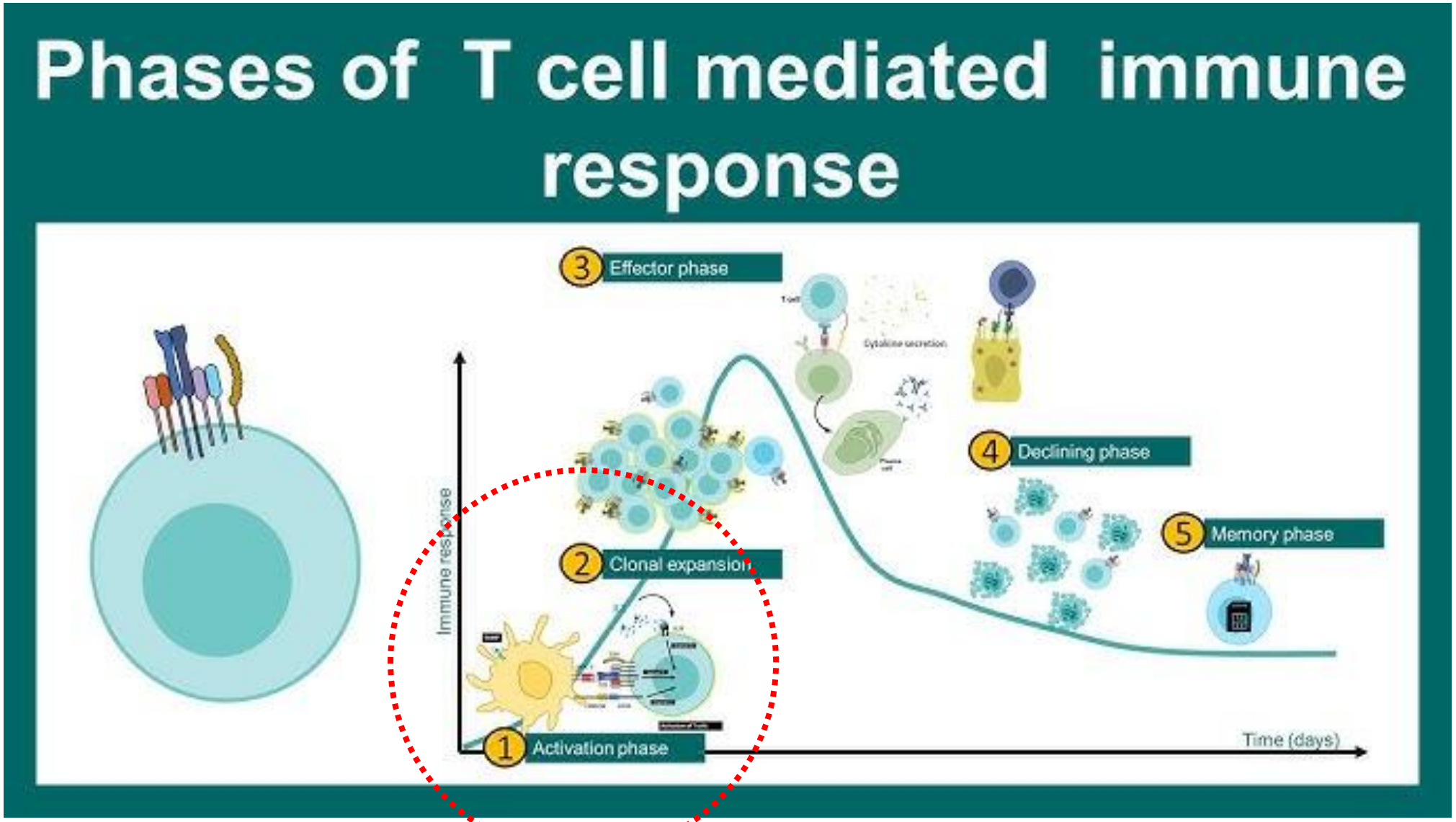




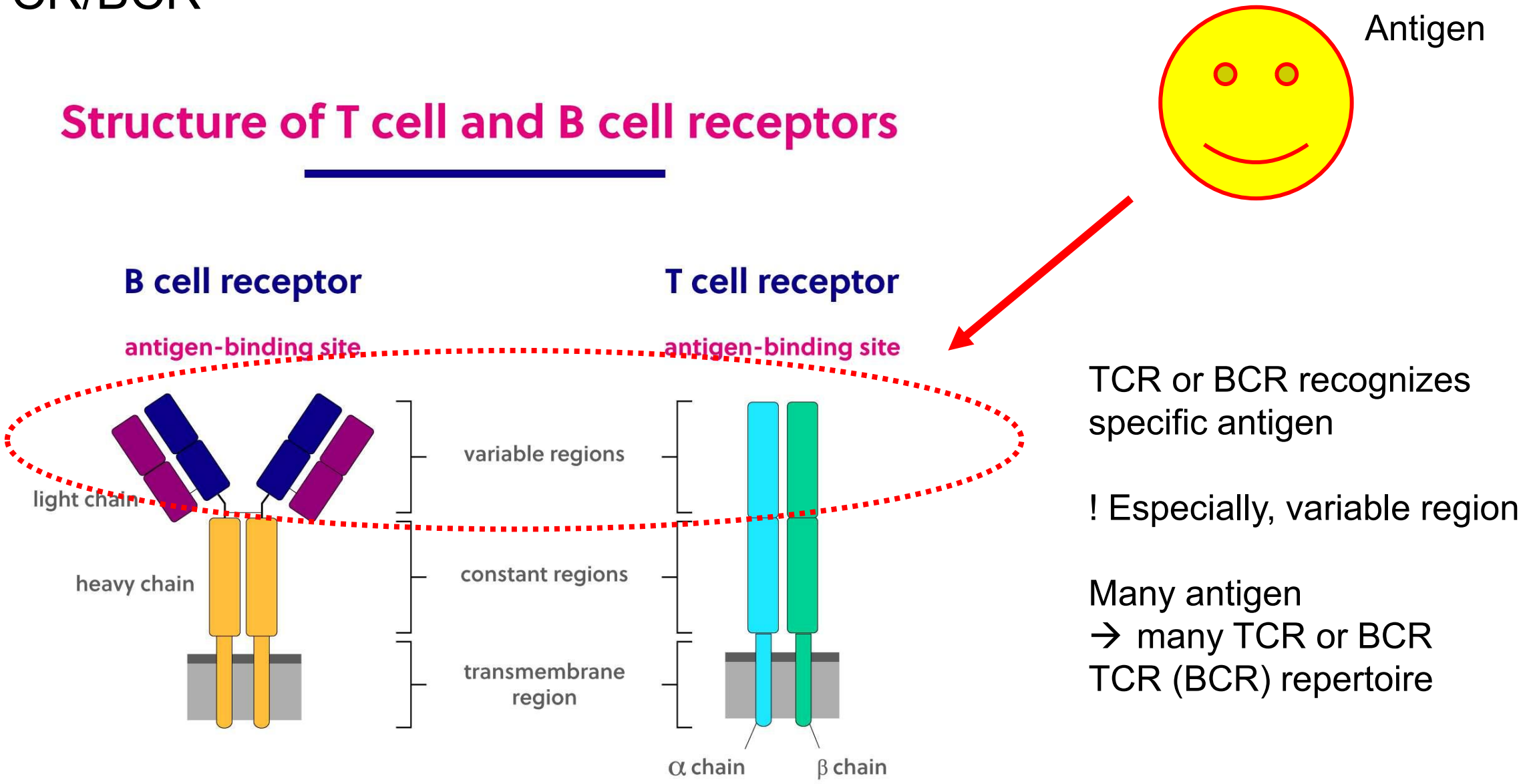


-Demultiplexing without hash-tagging!
-Assumption, different individual has different genome → SNP
-Each cell from each individual may contain distinctive SNP from transcriptome
-If one barcode contains multiple SNPs from different individual → multiplet!
(Cannot detect multiplet from the same individual)

- TCR/BCR


Phases of T cell mediated immune response
3 Effector phase
4 Declining phase
5 Memory phase
2 Clonal expansion
1 Activation phase
Immune response
Time (days)

- TCR/BCR

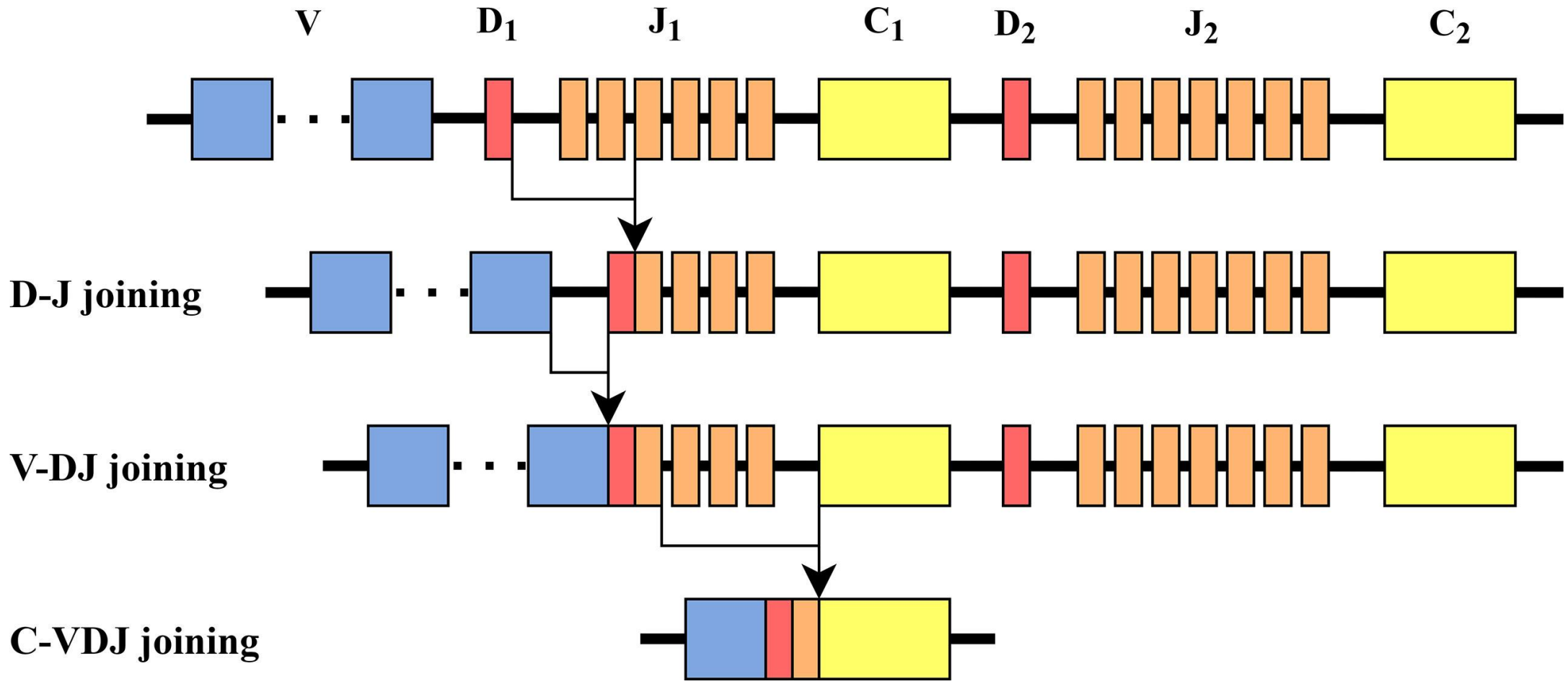


TCR or BCR recognizes specific antigen

! Especially, variable region

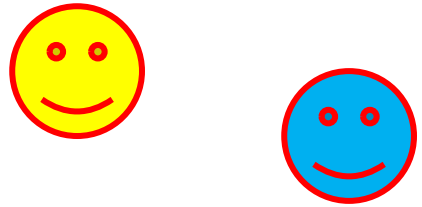
Many antigen
→ many TCR or BCR
TCR (BCR) repertoire

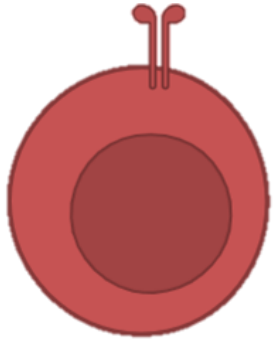
- TCR/BCR

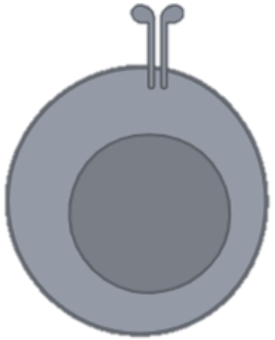


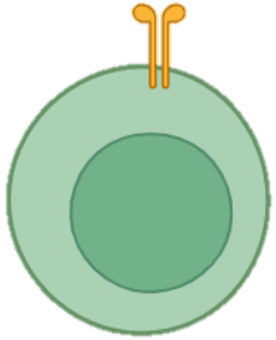
VDJ recombindation → various combination + hypermutation → Diversity ↑

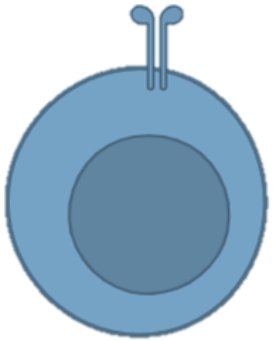
- TCR/BCR

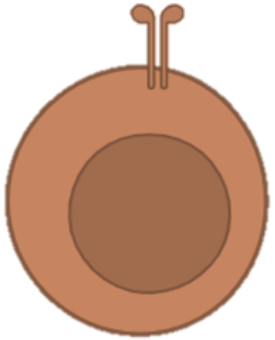




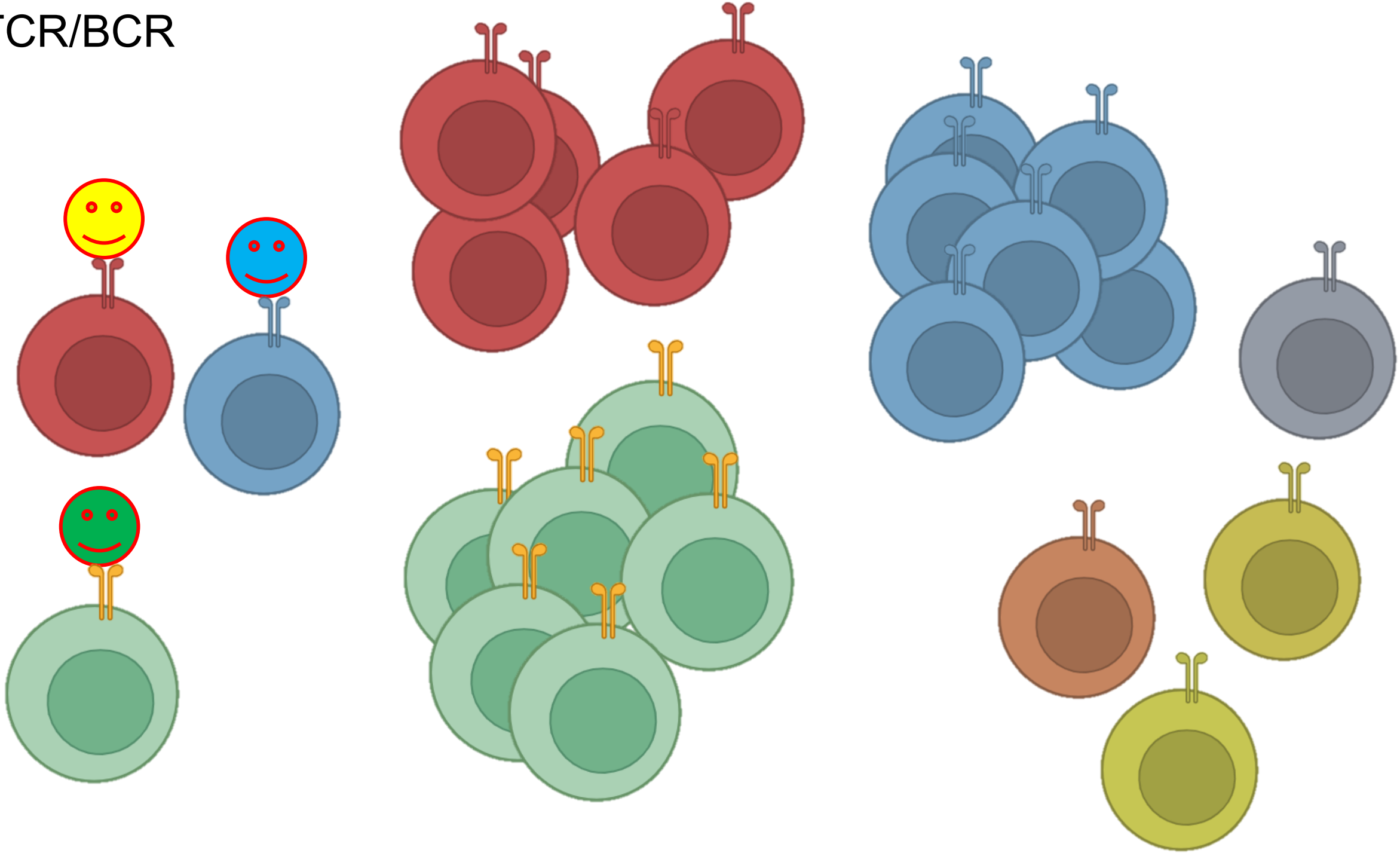




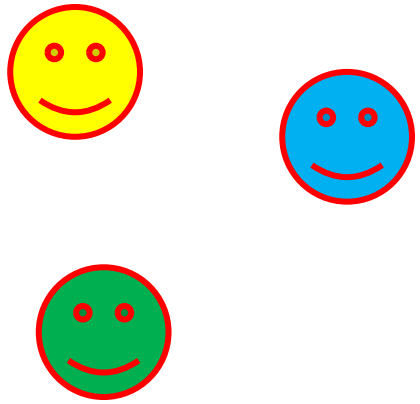




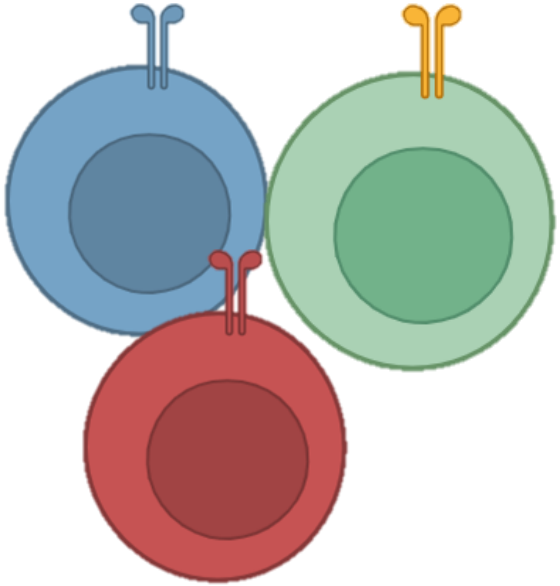
- TCR/BCR



- TCR/BCR

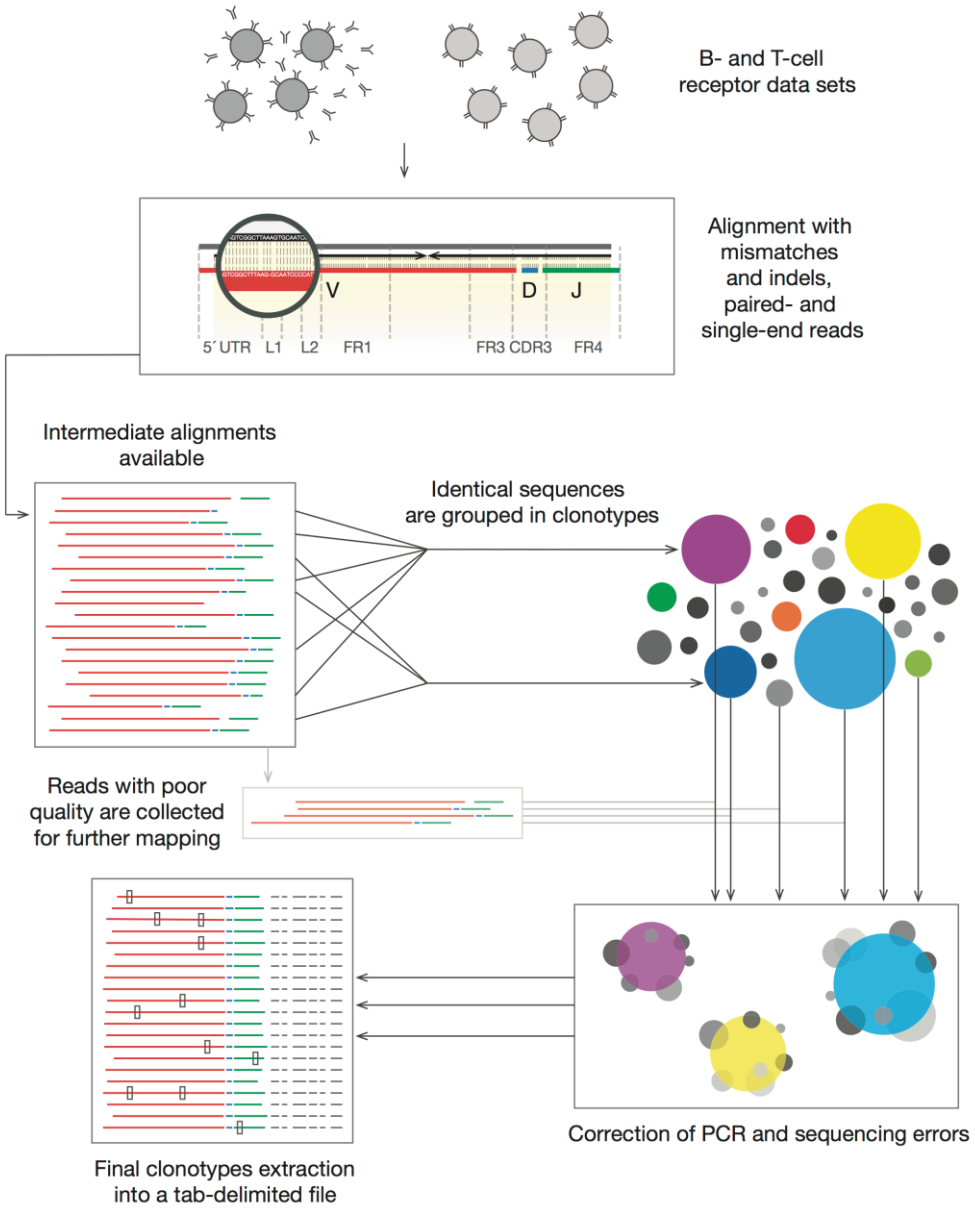


Which antigen??
→ Drug design, target



Which TCR or BCR ??
Antibody → vaccine

- TCR/BCR

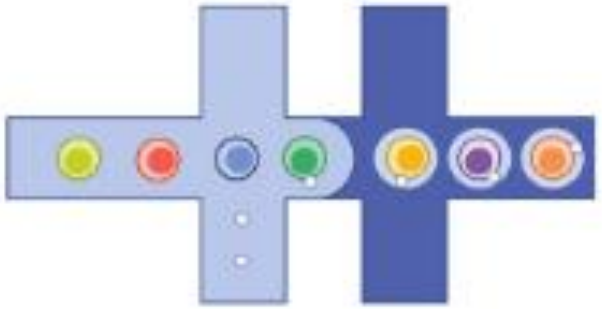


Bulk data
-Genome seq: VDJ region
→ Clonotyping
→ Abundance: clonal expansion

-Transcriptome data
→ VDJ region capture

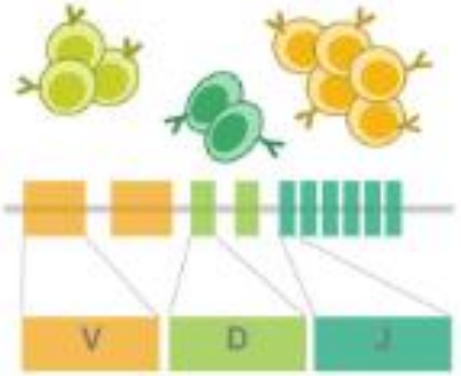
Ex) MiXCR

- TCR/BCR



Drop Isolation of single cell

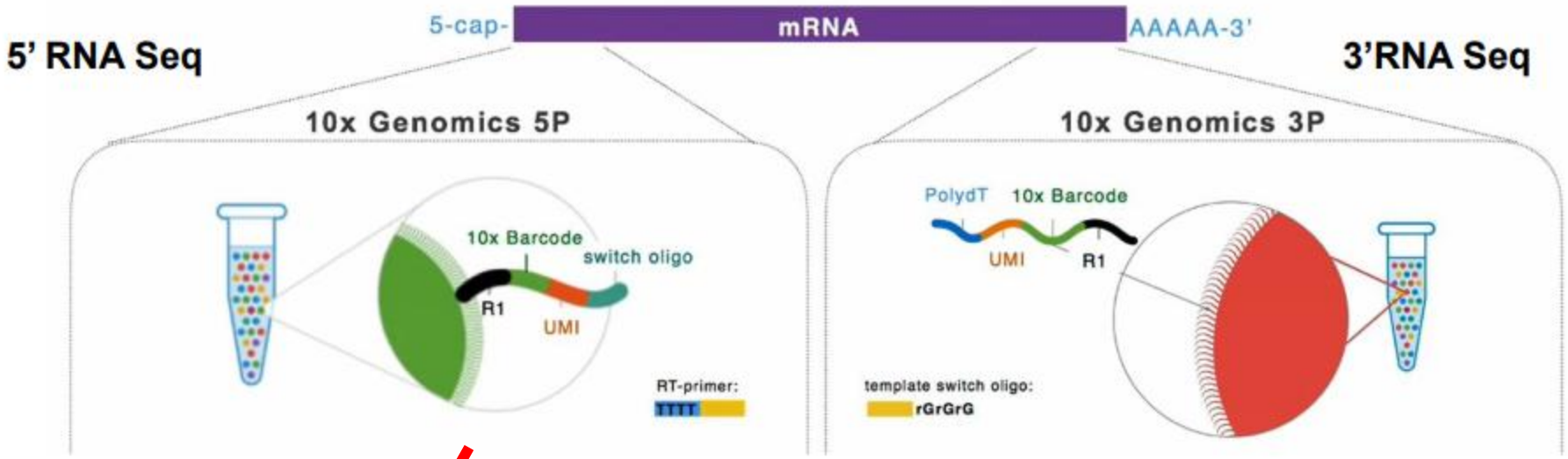
Visualization of single cell sequencing



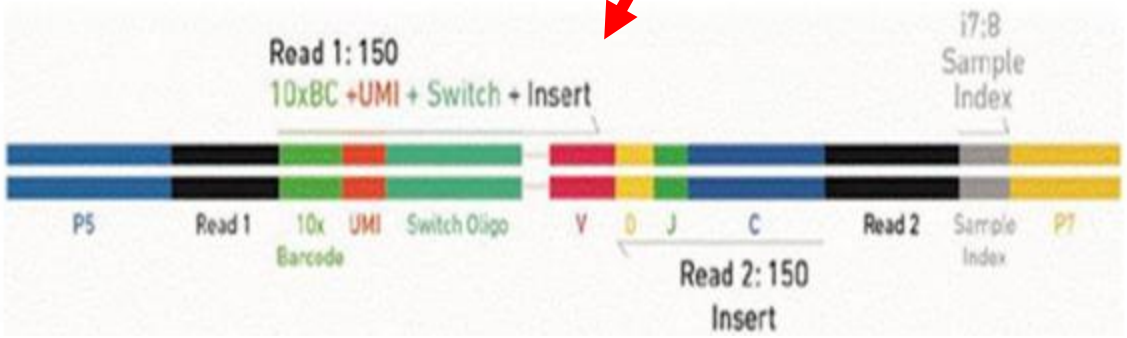
Detection of VDJ recombination

-Obtain TCR (BCR) information for each cell
-Inferring clonal expansion by abundance of sequence: indirect
→ Directly counts the abundance of clonal cell
+ obtains individual gene expression from each clone

- TCR/BCR

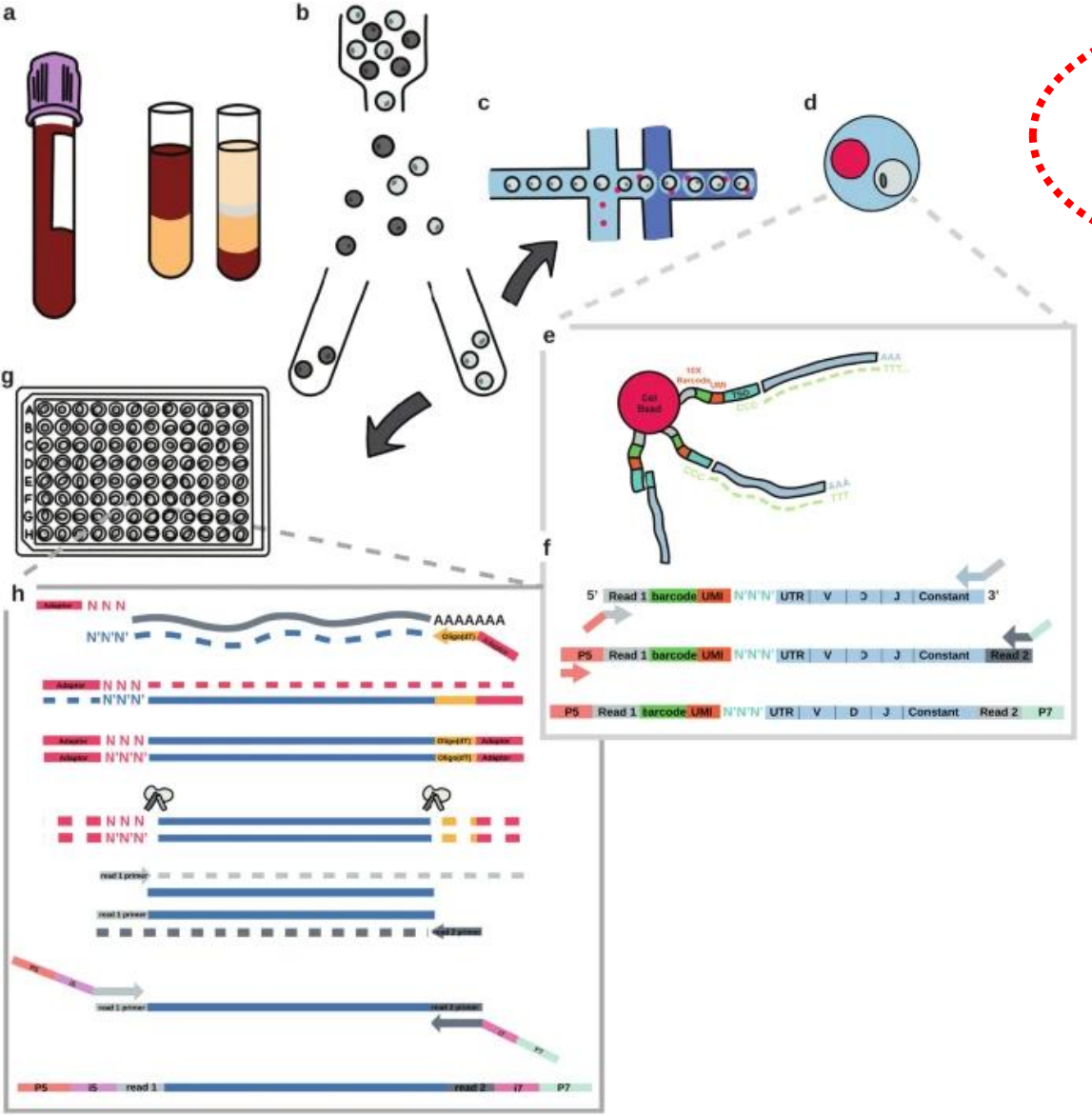


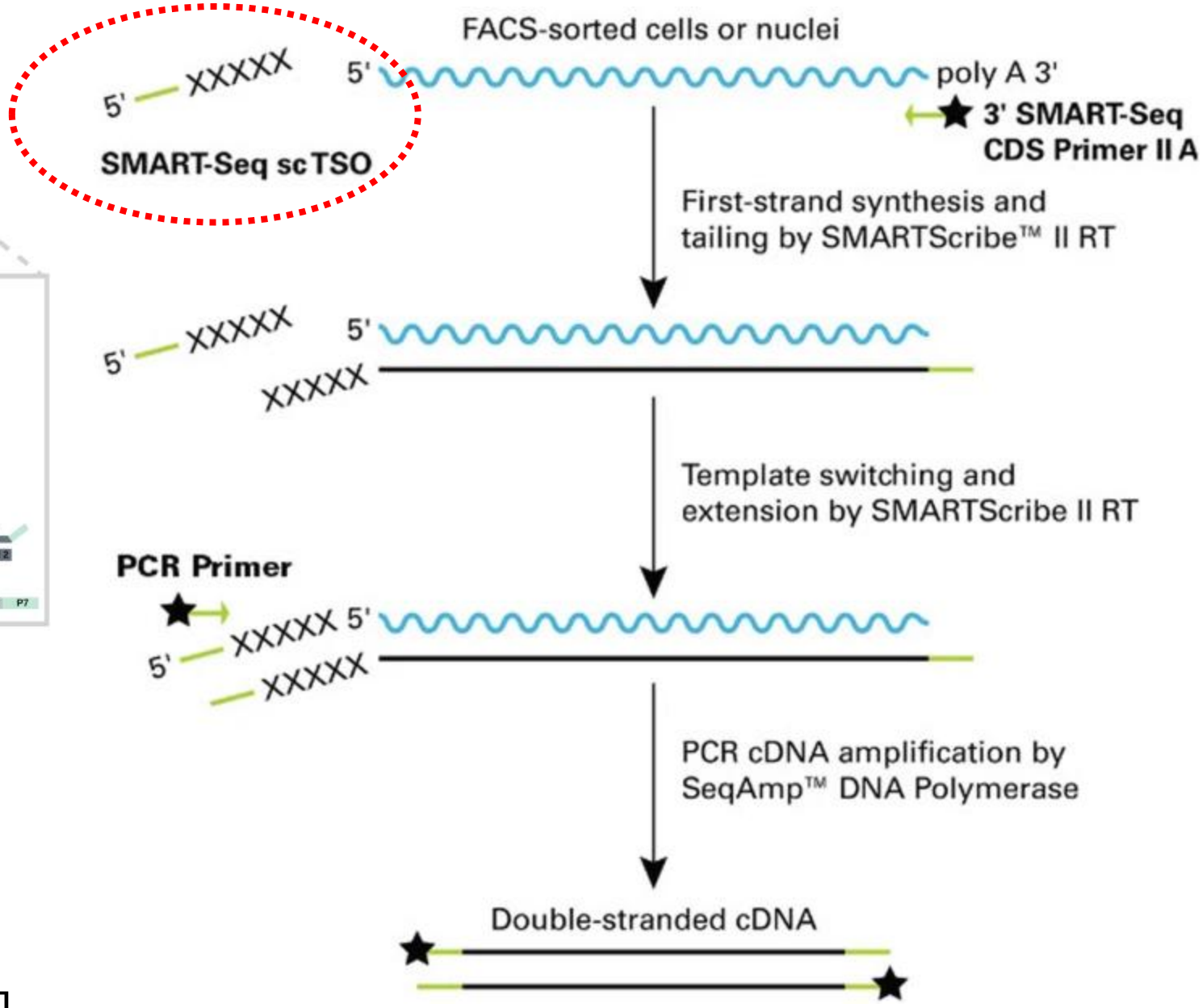
Immune Profiling



Using the Chromium Single Cell VDJ profiling, one can simultaneously examine the cellular context of the adaptive immune response and immune repertoires of hundreds to tens of thousands of T and B cells on a cell-by-cell basis.

- TCR/BCR





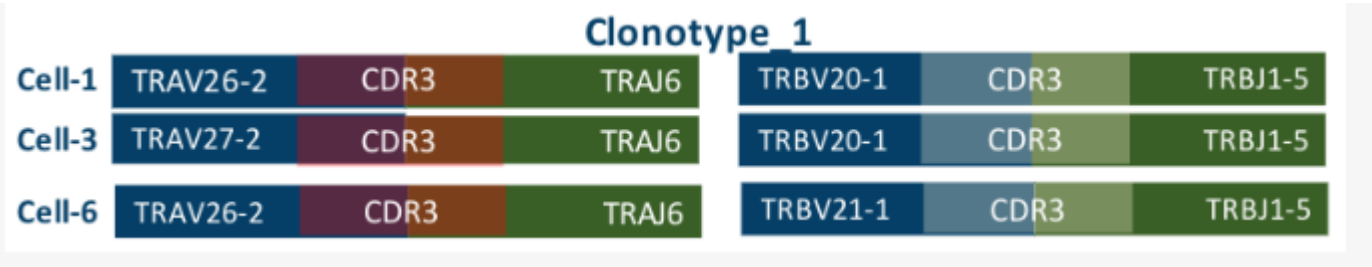
5’ sequencing [template switch oligo (TSO)]

- TCR/BCR

barcode	is_cell	contig_id	high_confi	length	chain	v_gene	d_gene	j_gene	c_gene	full_length	productive	cdr3	cdr3_nt	reads	umis	raw_clonotype_id	raw_consensus_id
AAACCTG	TRUE	AAACCTG	TRUE	574	IGH	IGHV3-21		IGHJ4	IGHM	TRUE	TRUE	CARGSRFL	TGTGCGA	2252	37	clonotype3669	clonotype3669_consensus_1
AAACCTG	TRUE	AAACCTG	TRUE	568	IGK	IGKV2-28		IGKJ5	IGKC	TRUE	TRUE	CMQALQT	TGCATGC	1247	15	clonotype3669	clonotype3669_consensus_2
AAACCTG	TRUE	AAACCTG	TRUE	551	IGK	IGKV1-5		IGKJ1	IGKC	TRUE	TRUE	CQHYNGY	TGCCAAC	6289	77	clonotype1358	clonotype1358_consensus_2
AAACCTG	TRUE	AAACCTG	TRUE	565	IGH	IGHV3-7		IGHJ4	IGHM	TRUE	TRUE	CARDWRE	TGTGCGC	2220	30	clonotype1358	clonotype1358_consensus_1

Barcode: cell barcode (same as scRNA-seq)
Is_cell, high_confidence, full_length, productive → QC
cf) partial CDR3: not fully sequenced, Out_of_frame: stop codon: no protein product (Trust4)

Clonotype_id: clonotype (same CDR3 seq)
Consensus_id: representative clonotype id



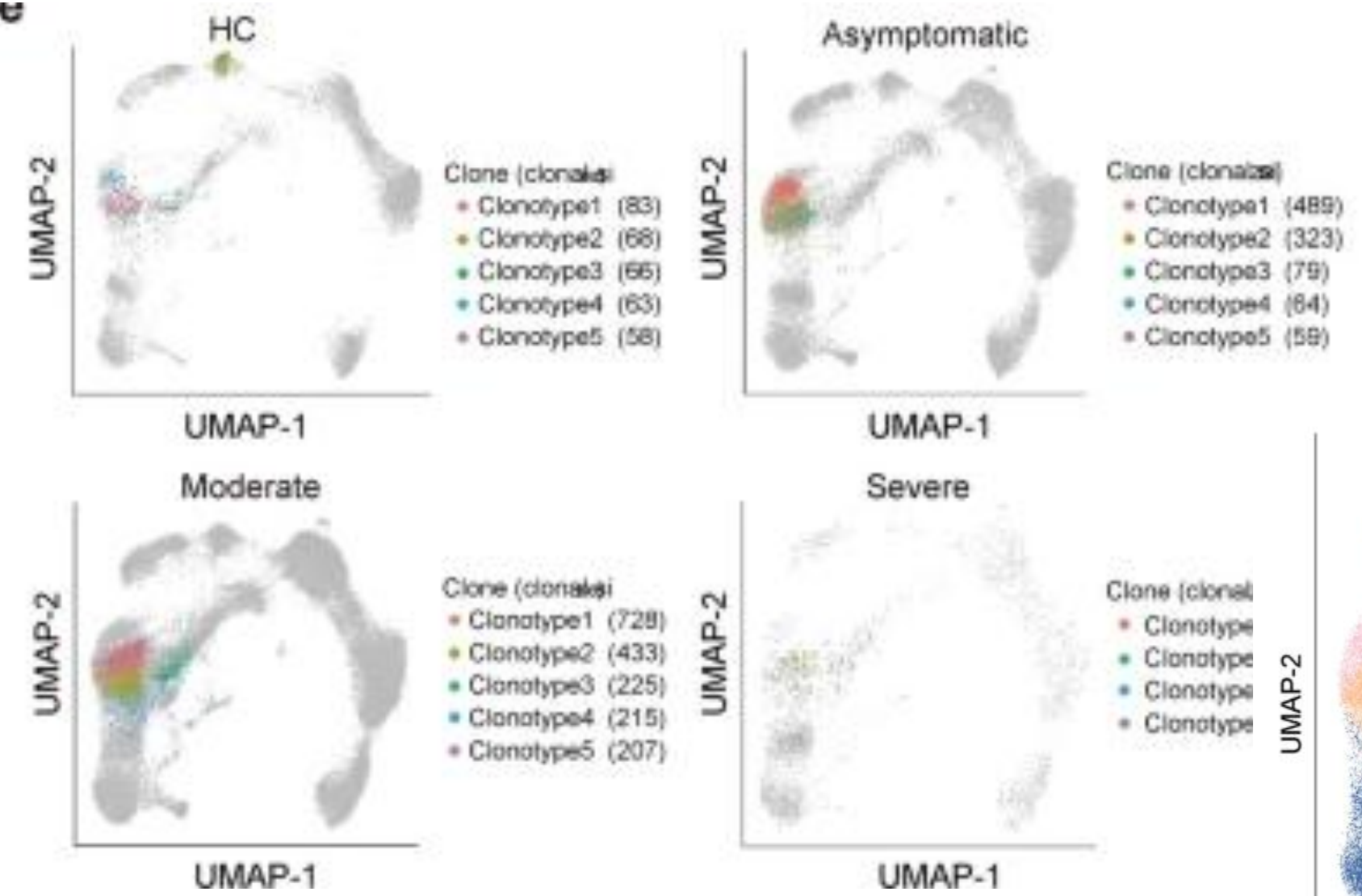
Same clonotype: Not only CDR3 but also other regions should be the same

- TCR/BCR

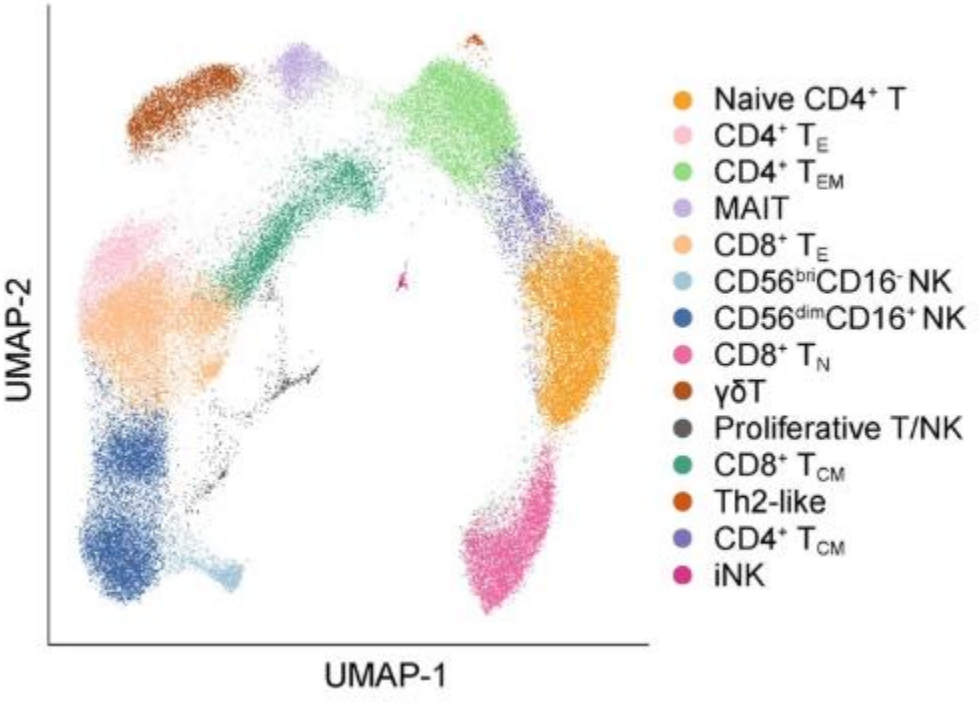
*Multiple TCR (BCR) per single cell
-TCR: 2 alpha + 1 beta or 1 alpha + 2 beta
-BCR: more than 1 heavy or light chain
+haplotype

→ Theoretically, more than 4 is possible
→ Cellranger: exclude > 4

- TCR/BCR

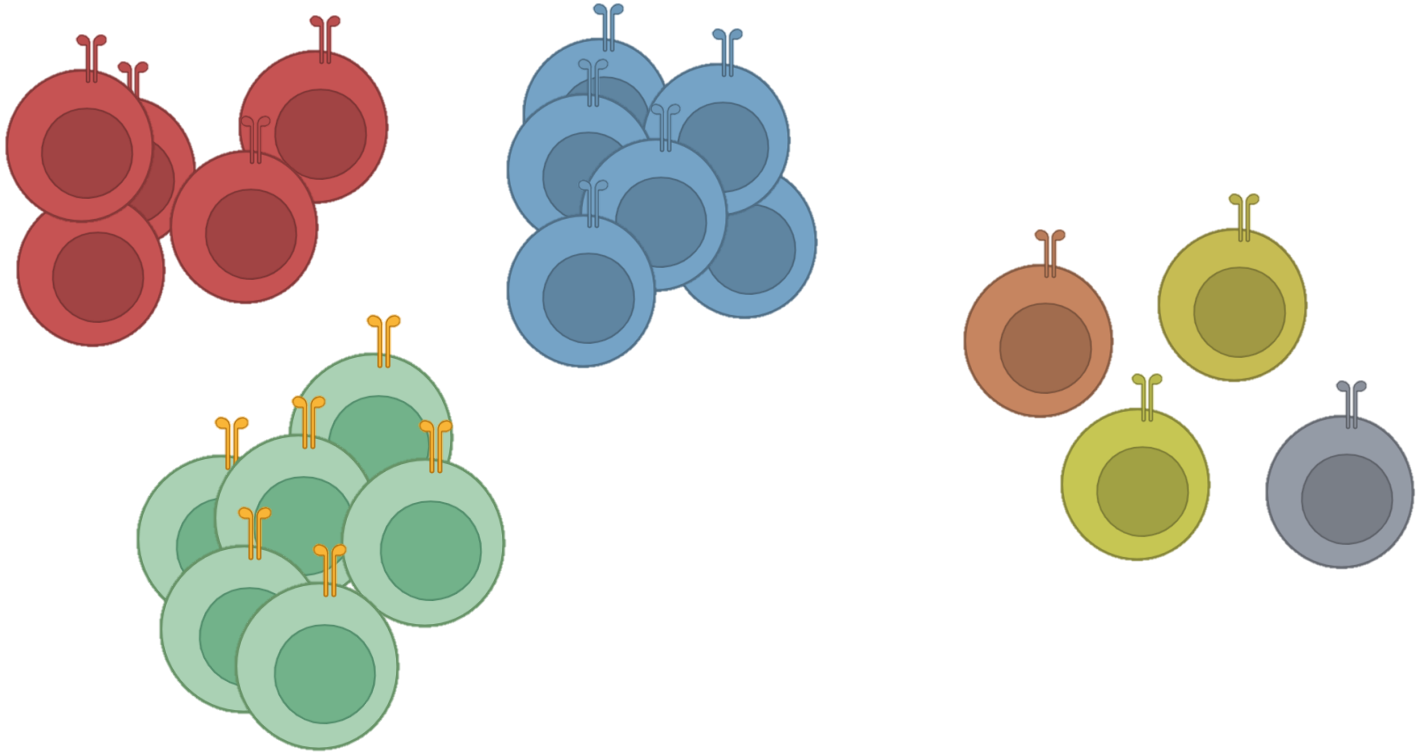


-Clonotpying in single-cell level
-Different clonotype across different cell state
-Different clonotype across different disease state

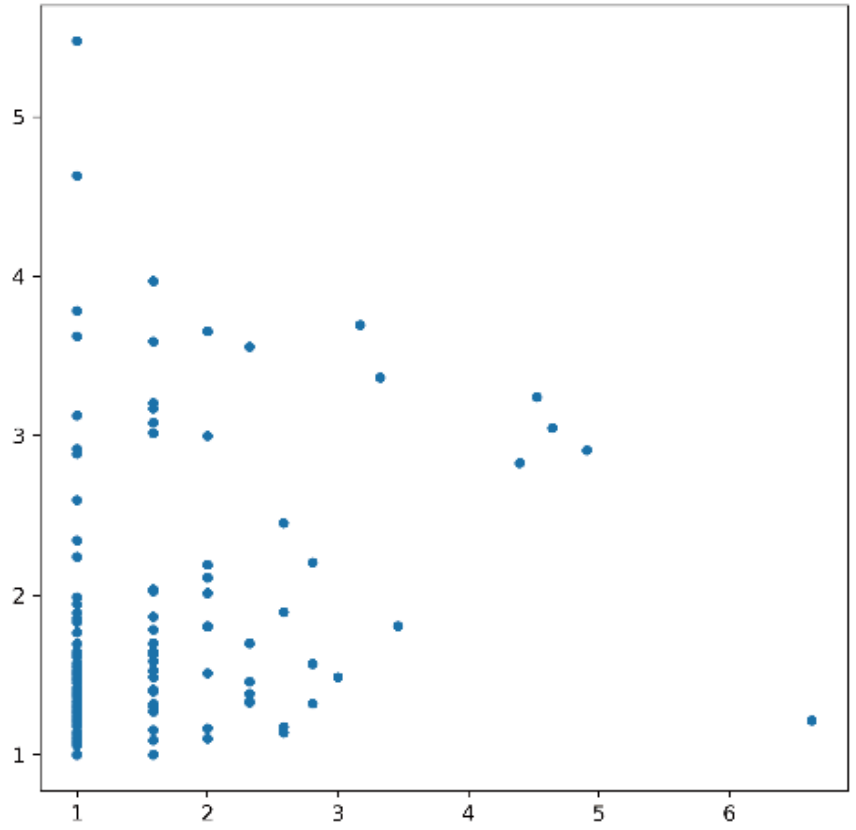


Single-cell immune profiling reveals distinct immune response in asymptomatic COVID-19 patients

- TCR/BCR



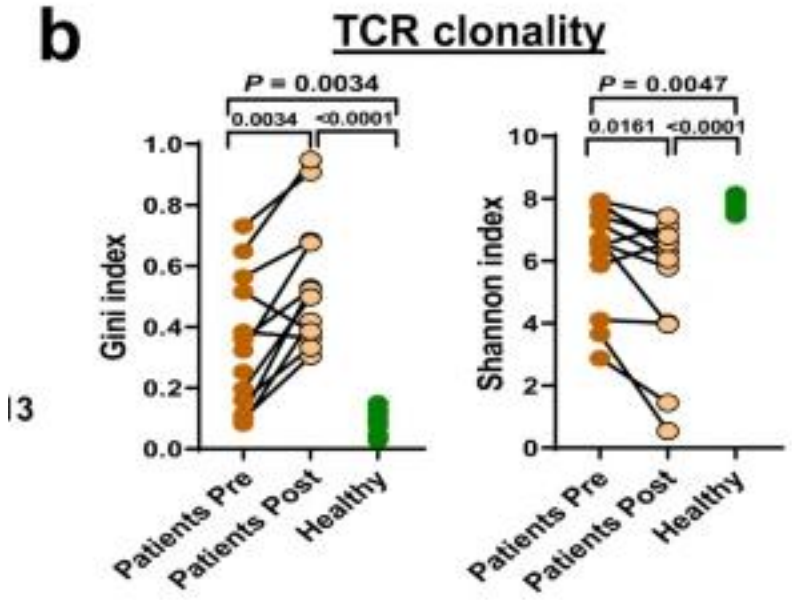
-Comparison between clonally expanded Tcell and singlets
-Association between clonal expansion and transcriptome

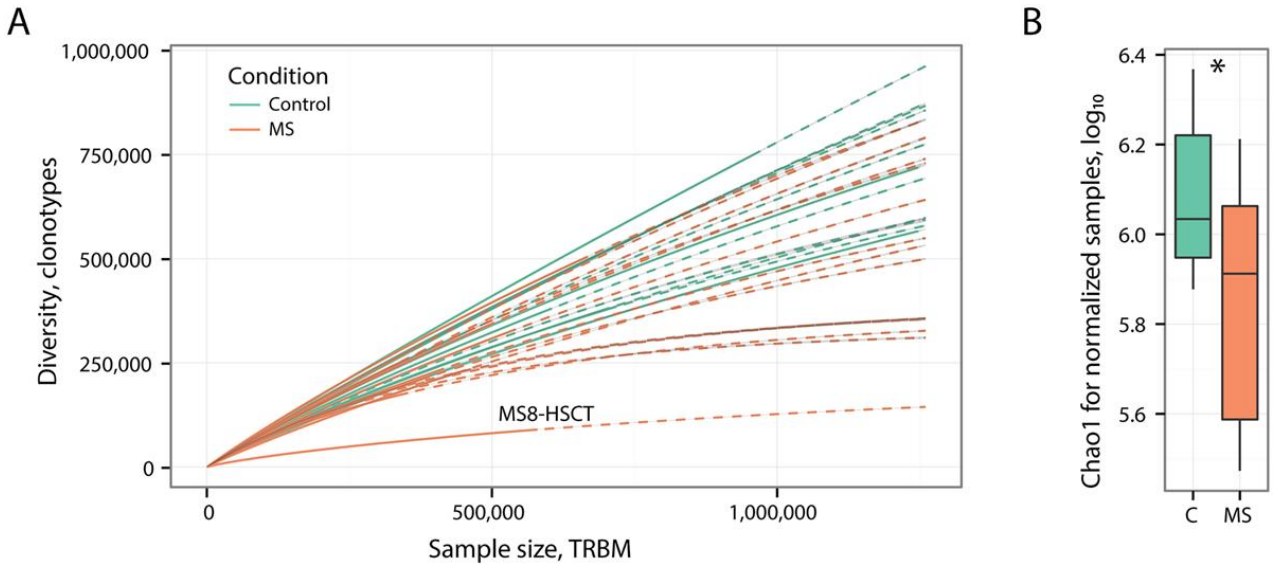


- TCR/BCR

Diversity measurement: quantification of repertoire
→ Gini index, Shannon index, diversity index

Diversity is correlated with sample size
→ down-sampling to adjust across samples





Single-cell RNA sequencing coupled to TCR proflifing of large granular lymphosute leukemia T cells

VDJtools: Unifying Post-analysis of T cell receptor repertoires

- TCR/BCR

Simpson's Diversity index (ecology based)
-Richness: how many species → large sample size → always big
-Evenness: relative abundance of the different species

→ Considering both measurement

Simpson's Diversity Indices

The term 'Simpson's Diversity Index' can actually refer to any one of 3 closely related indices.

Simpson's Index (**D**) measures the probability that two individuals randomly selected from a sample will belong to the same species (or some category other than species). There are two versions of the formula for calculating **D**. Either is acceptable, but be consistent.

$D = \sum (n / N)^2$	$D = \frac{\sum n(n-1)}{N(N-1)}$
n = the total number of organisms of a particular species N = the total number of organisms of all species	

The value of **D** ranges between 0 and 1

Bigger → low diversity
Simpson's index of diversity: 1-D
Simpson's reciprocal index: 1/D

- STARTRAC

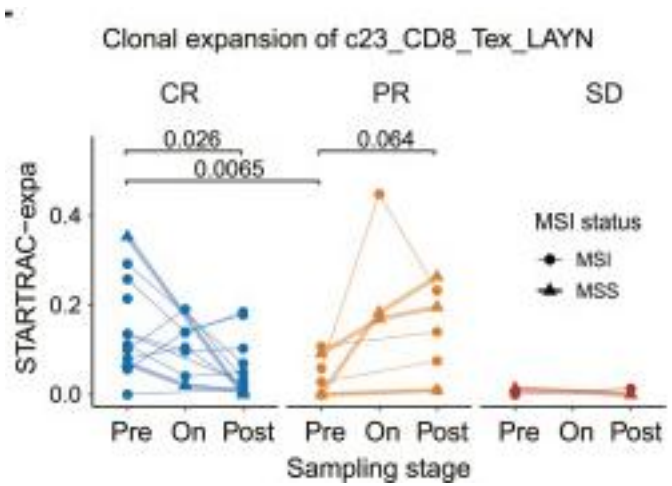
Clonality measurement

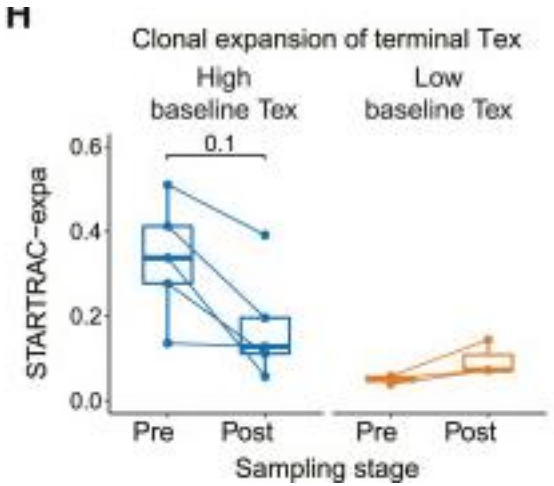
$$\text{Clonality} = 1 - \frac{H}{\log(n)}$$

H: shannon entropy
N: number of clonality
~1: one dominant clone → clonal expansion
<<1: diverse

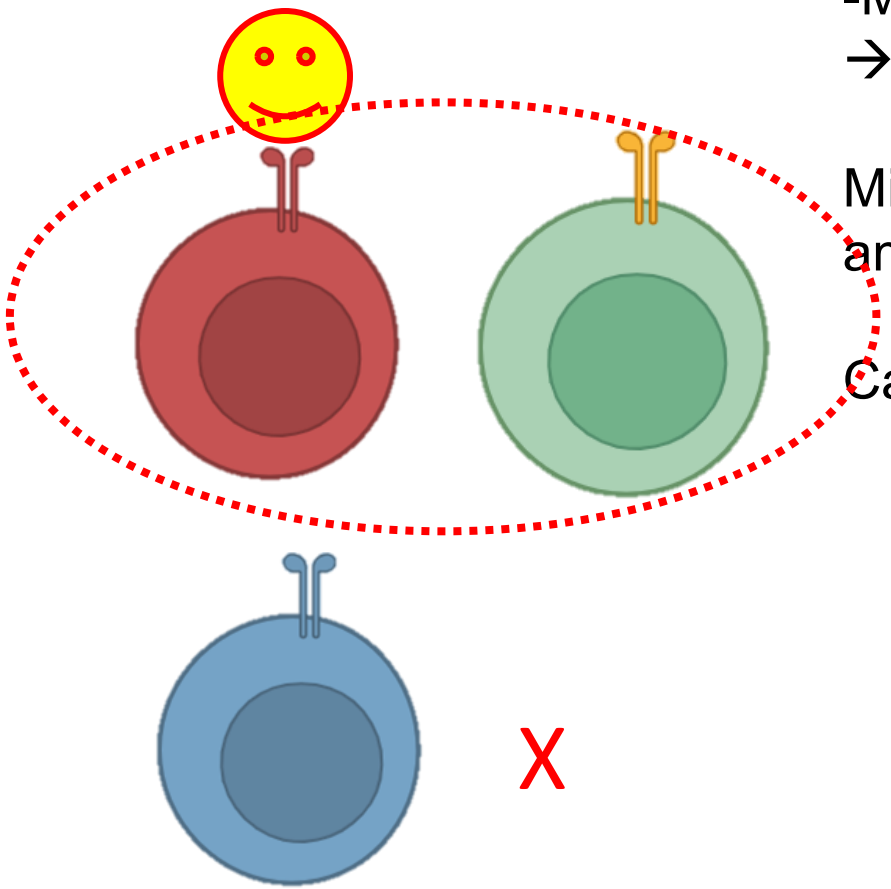
- TCR/BCR

-Diversity or clonality can be only measure in a sample level





- TCRdist



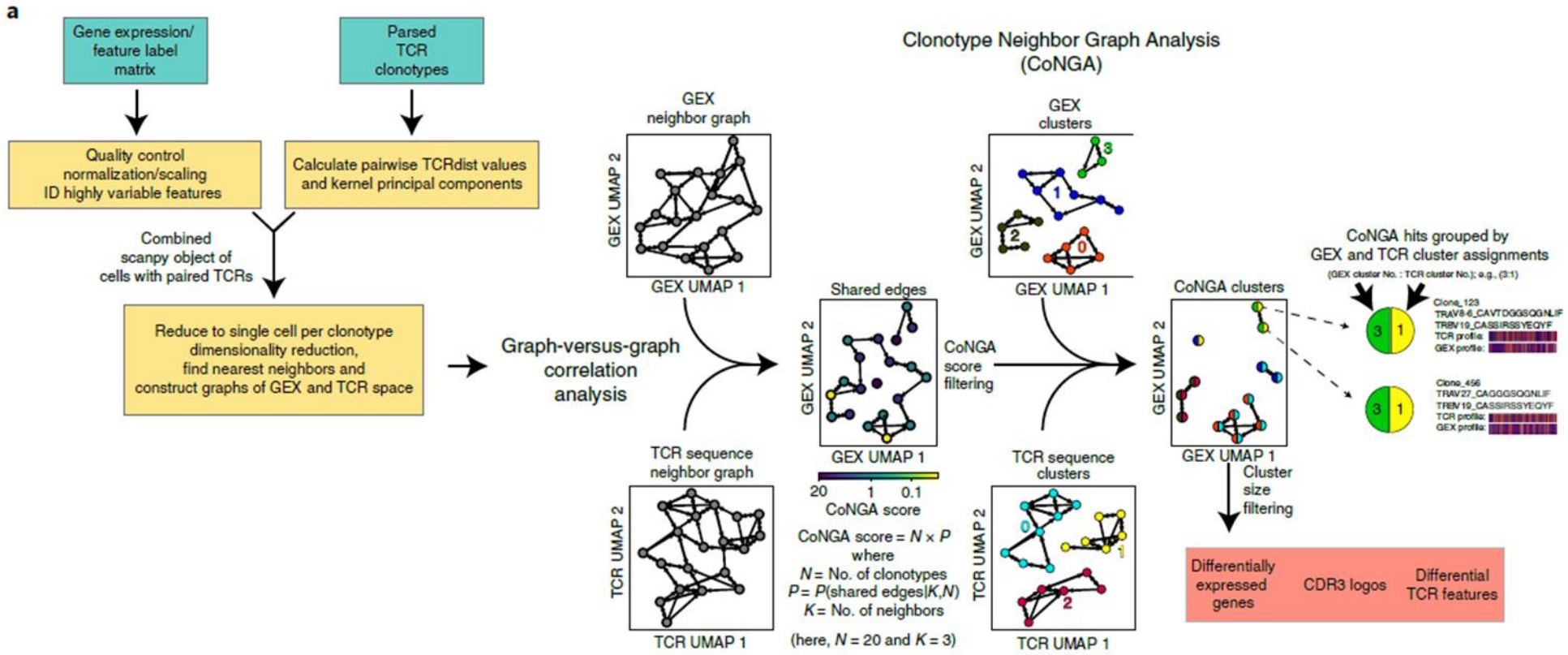
-Maybe similar TCR sequence can bind the same pMHC (epitope)
→ Let measure the distance across different TCR seq

Mismatch region → BLOSUM62 (distance measurement for amino-acid)

Can be grouped together!

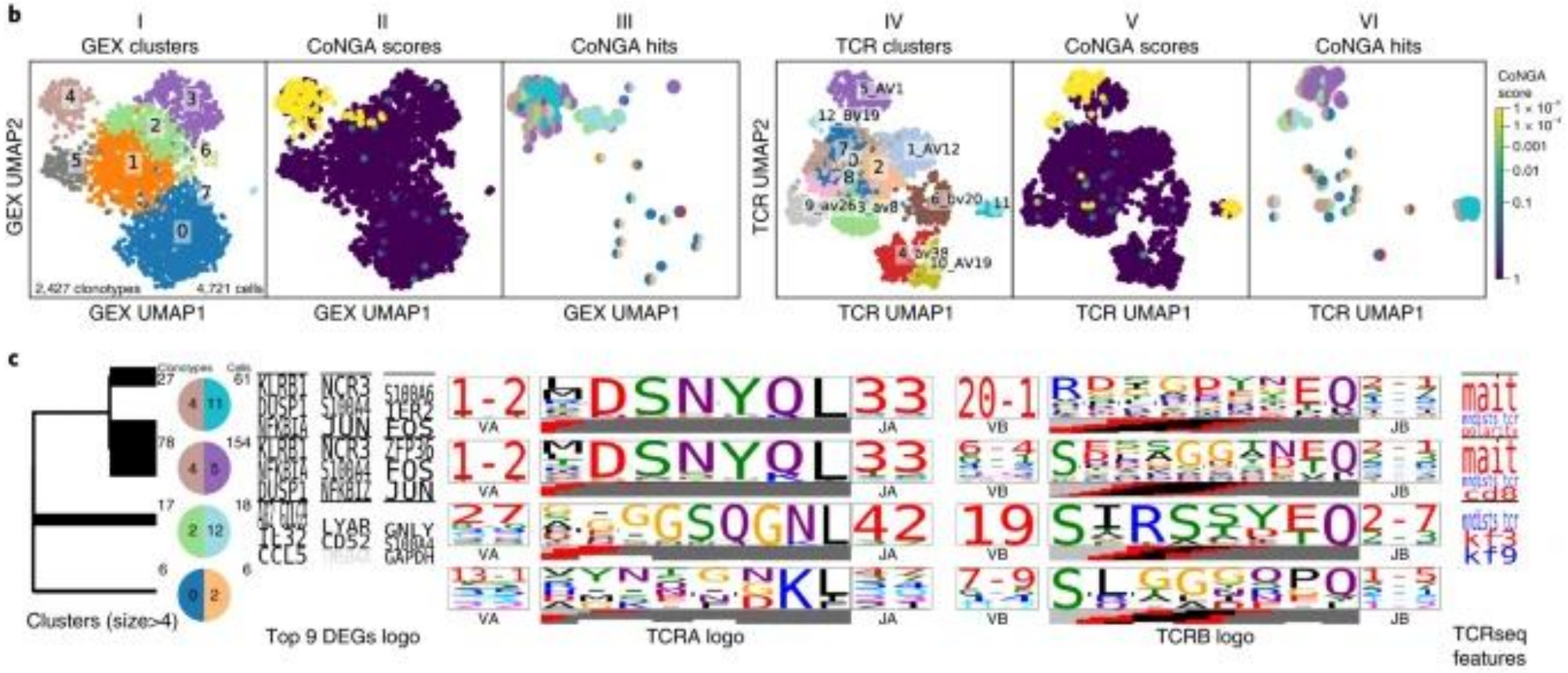
- CONGA

-Can we integrate between clonality and gene expression ?



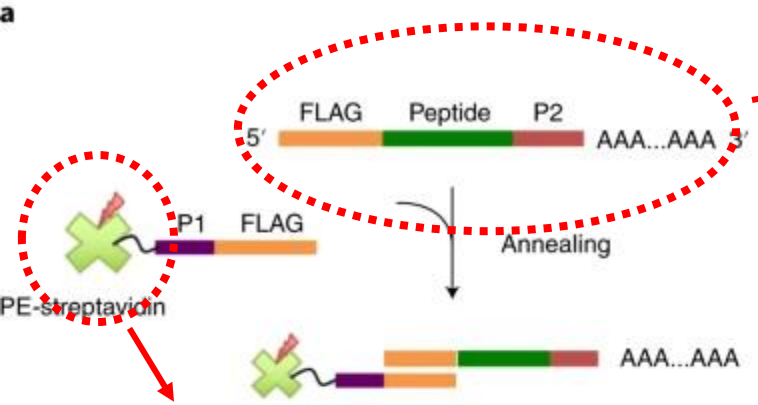
-Find shared edges (of cell-graph) between GEX and TCR → cluster

- CONGA

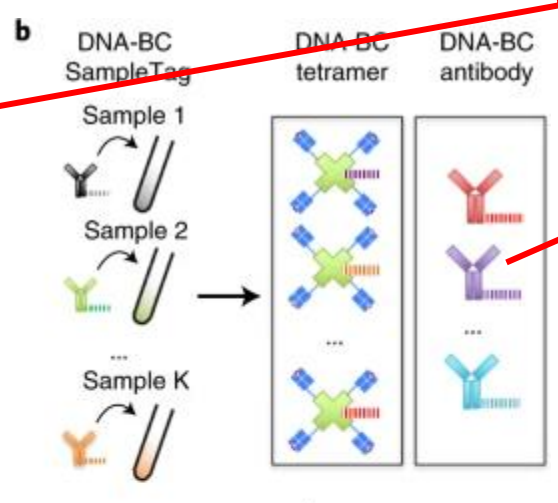


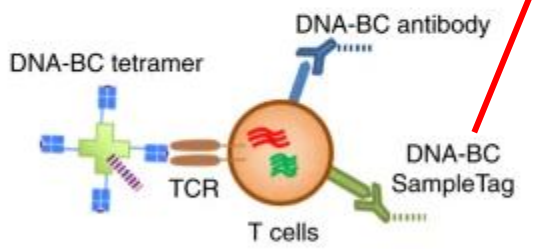
-Find which cluster (gex) has an enrichment of clonal expansion

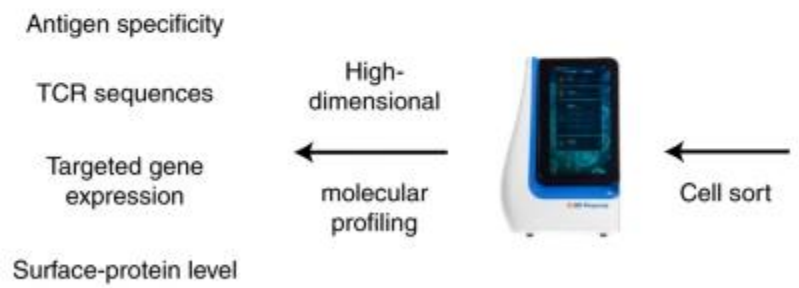
- Antigen-specific T cell profiling (TetTCR-SeqHD)

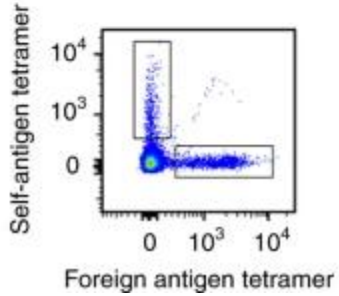


MHC tetramer









1: Peptide → loaded onto MHC tetramer (pMHC) DNA barcoded with poly A

2: similar to CITE-seq

3: sample-tag: Hash-tagging

FACS sorting → BD Rhapsody (scRNA-seq

→ T-cells with transcriptome + antigen-specificity

High-throughput and high-dimensional single-cell analysis of antigen-specific CD8+ T cells

- Antigen-specific T cell profiling (Indirect approach)

IEDB Analysis Resource

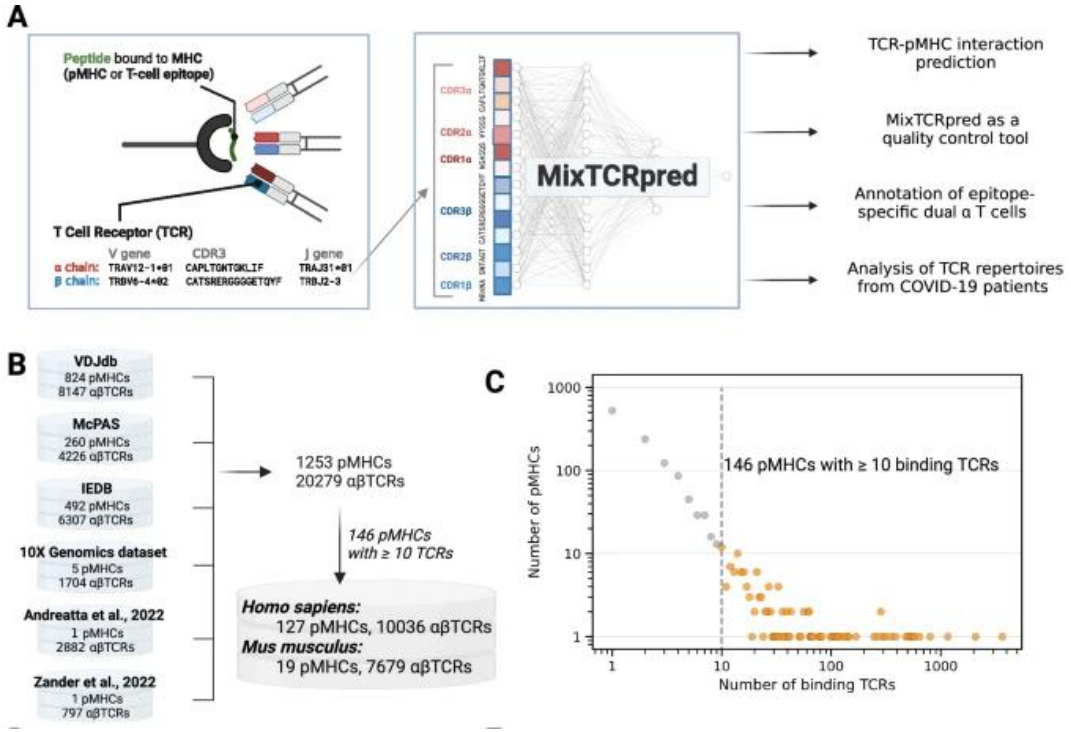
Overview | T Cell Tools | B Cell Tools | Analysis Tools | T

T Cell Epitope Prediction Tools

T Cell Epitopes - MHC Binding Prediction

NetTCR-2.0 enables accurate prediction of TCR-peptide binding by using paired TCRα and β sequence data

Deep learning predictions of TCR-epitope interactions reveal epitope-specific chains in dual alpha T cells



Predict the epitope by TCR sequence
(mostly based on deep learning)

→ Specific virus epitope
→ Cancer study: if there are exome-seq
→ mutation calling
→ candidate neoantigen
→ discard virus-bacteria epitope
→ Extract neoantigen-specific T cells

- 2-dimensional group set

Experimental design
A vs B or treated vs control

Now …

	Clonal expansion	Singlets
A group		
B group		
… group		

Every analysis from scRNA-seq
-DEG
-Geneset analysis
-Cell abundance
-Network (coexpression or GRN)
-Cell-Cell interaction
-Trajectory analysis