

# *Immunoinformatics*

-

## Computational approaches to study the human immune system

Katharina Imkeller

*University Hospital Frankfurt and Frankfurt Cancer Institute*

*CSAMA 2022*



## *Goals for this lecture*

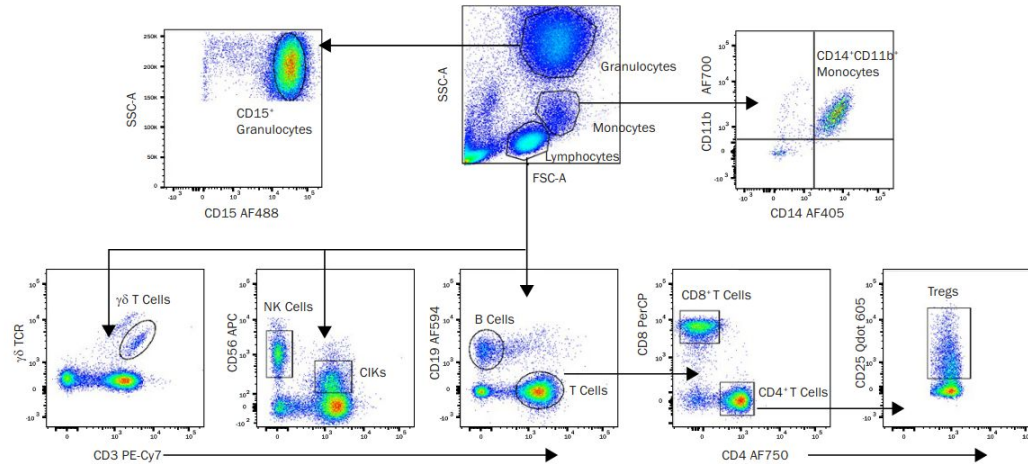
Get to know computational approaches to study:

1. Cell types and phenotypes
2. Interactions of immune cells
3. Antigen specificity

- 1. Cell types and phenotypes**
2. Interactions of immune cells
3. Antigen specificity

# How immunologists usually define cell types

## FACS gating for cell type identification



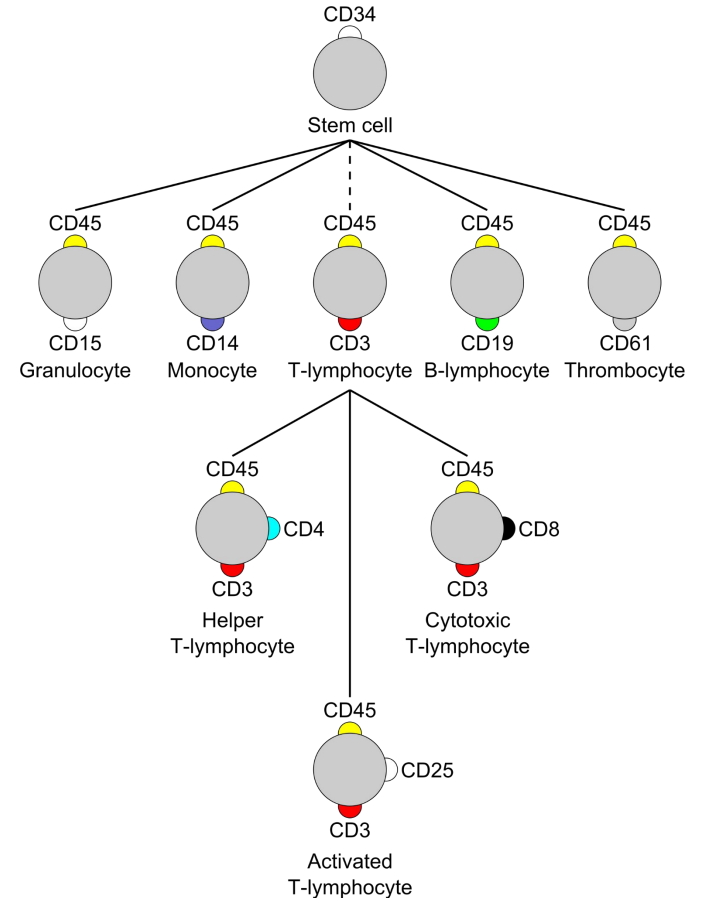
# Identification of cell types by clusters of differentiation (CD)

371 defined clusters of differentiation

Historical origin: grouping of antibodies that bind to the same cell surface antigen.

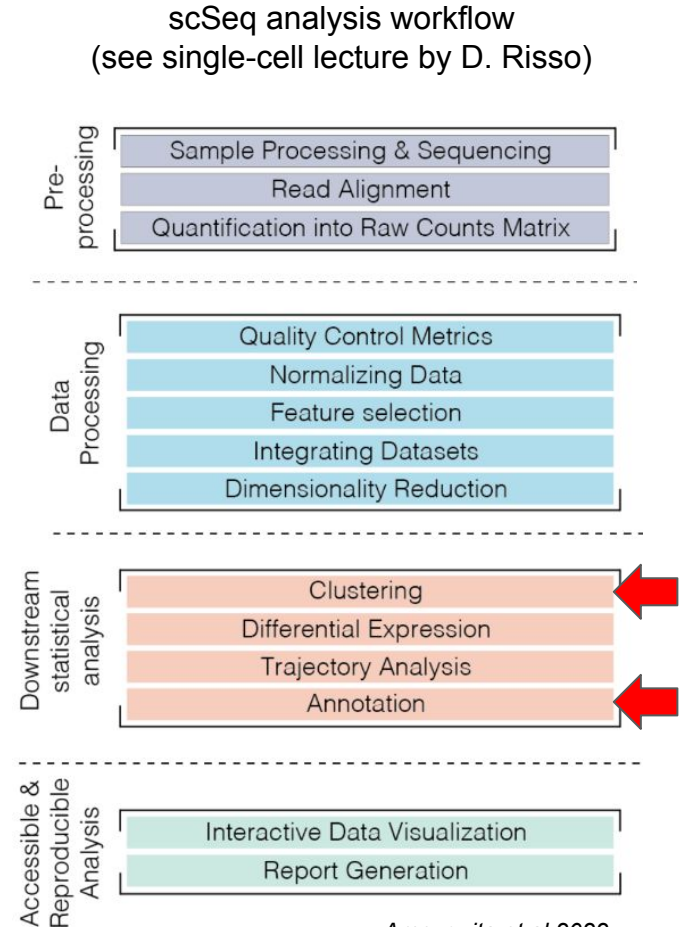
Challenges for immunoinformatics:

- consistent usage of CD nomenclature dependent on field of immunology
- CD nomenclature does not always correspond to protein name/gene name
- antibody binding != surface marker expression != gene expression



# Cell type assignment in the single cell transcriptomics analysis workflow

Very likely:  
scRNA cell type not exactly identical to  
FACS cell type

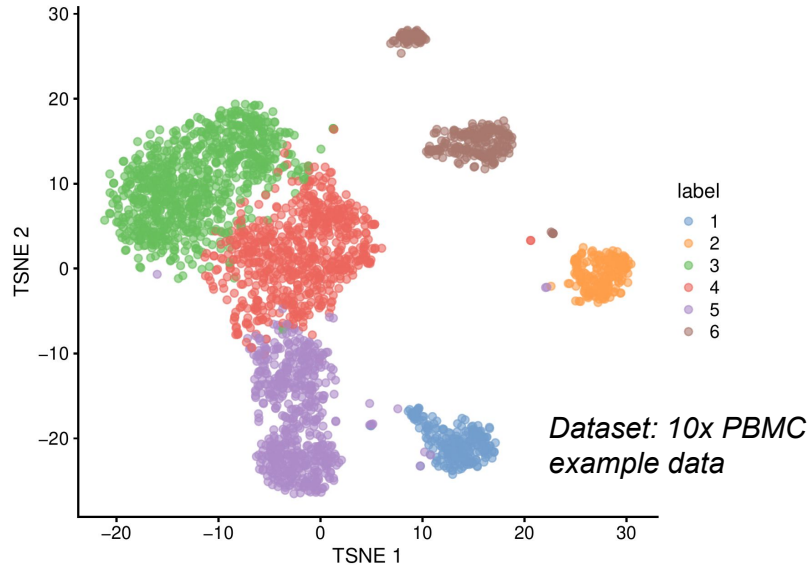


# Manual annotation using marker gene detection

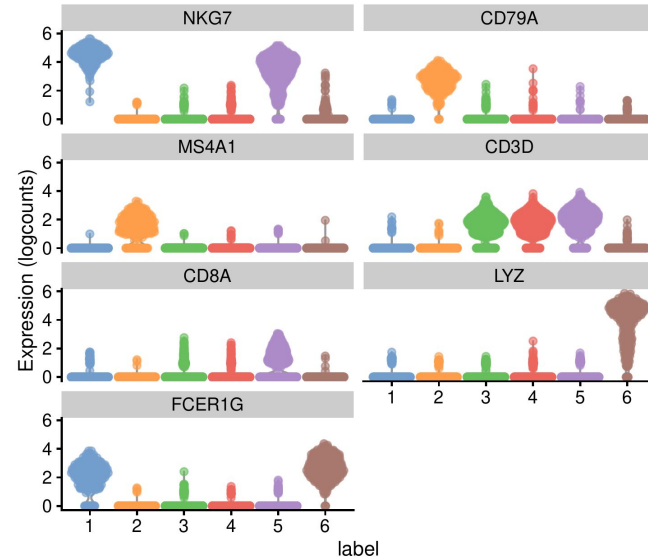
Bioconductor package: *scrn*

Functions: *scoreMarkers()*, *findMarkers()*

## T-SNE of cells coloured by cluster



## Expression of marker genes

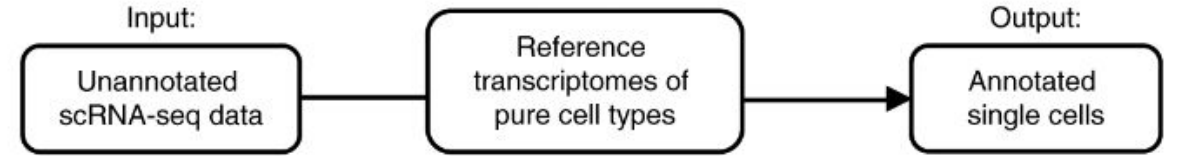


# Automated cell type annotation using Bioconductor

Bioconductor packages: *SingleR*, *cellDex*

Documentation: <http://bioconductor.org/books/release/SingleRBook/>

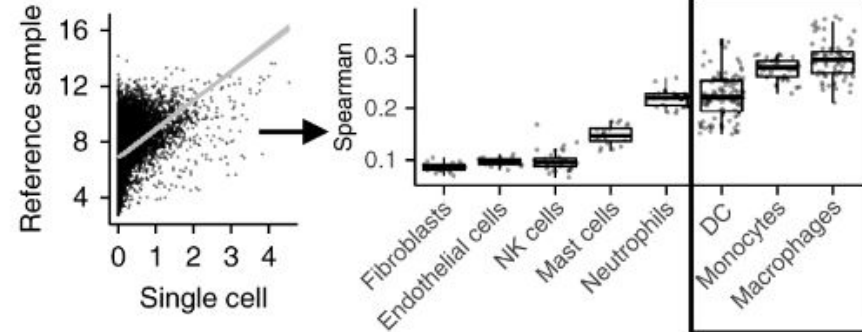
```
ref <- BlueprintEncodeData()  
pred <- SingleR(test = sce,  
  ref = ref,  
  labels = ref$label.main)
```



Step 1:  
Identifying variable genes among cell types in the reference set

Step 2:  
Correlating each single-cell transcriptome with each sample in the reference set

Step 3: Iterative fine-tuning—reducing the reference set to only top cell types

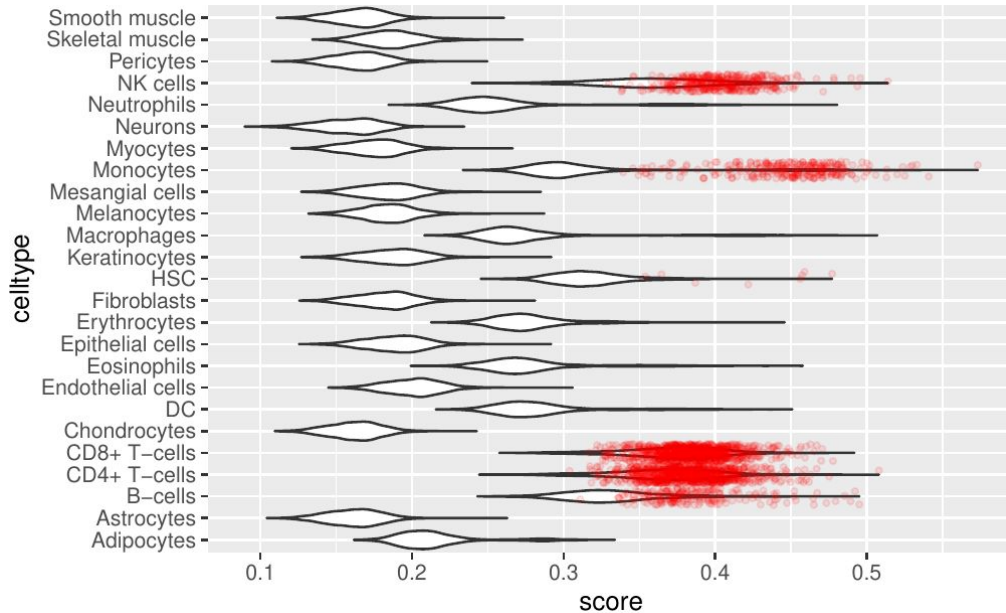




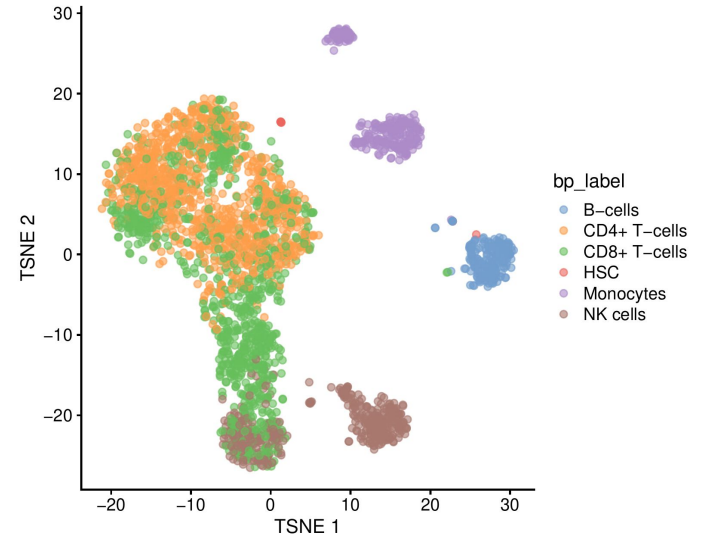
# SingleR returns prediction scores and cell type labels

Scores for cell type assignment

Scores for assigned labels are indicated in red

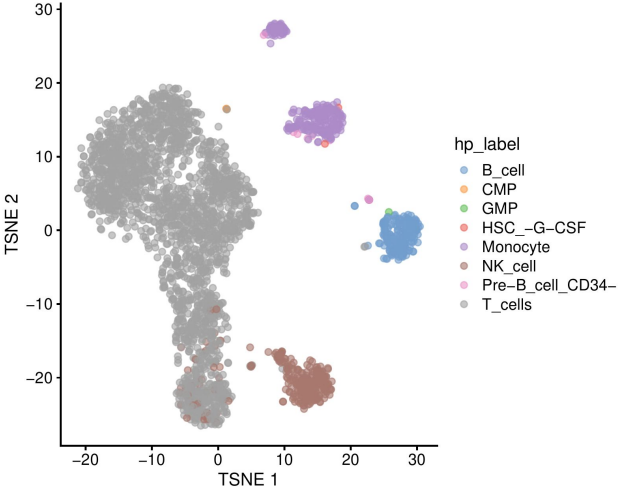


Result of automated cell type assignment

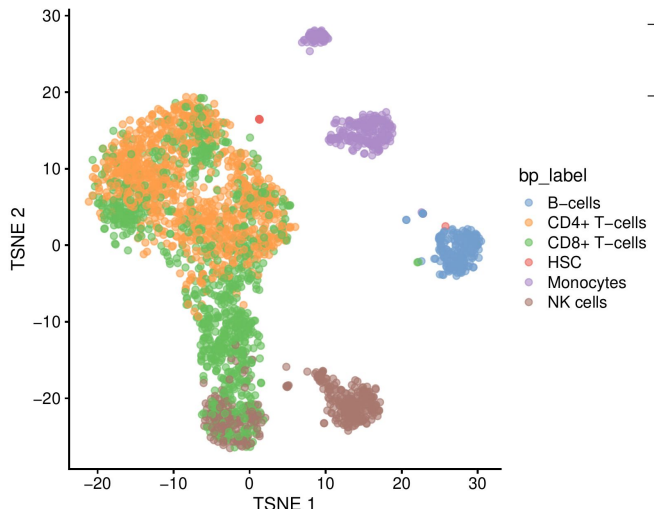


# Dependent on the reference, the predictions may change

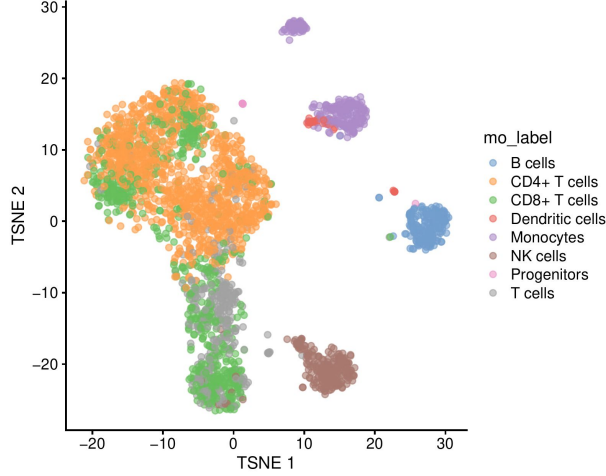
## HumanPrimaryCellAtlasData()



## BlueprintEncodeData()



## MonacoImmuneData()

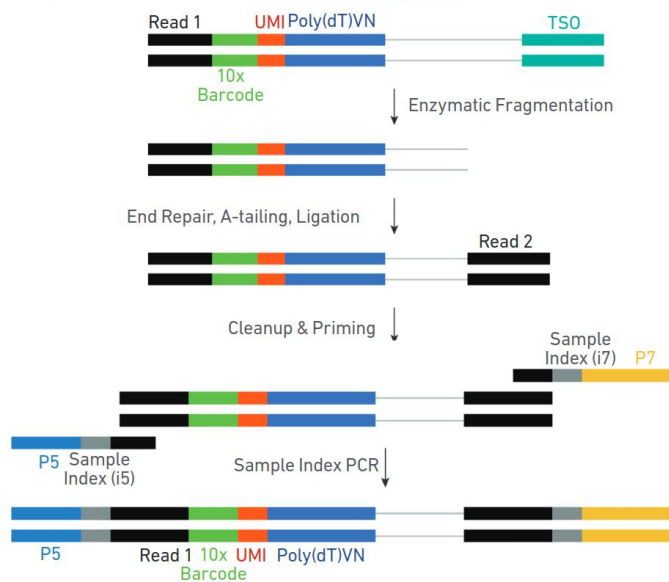


# sc-seq datasets generated using different experimental method

## 10x genomics 3'

Whole transcriptome, poly(A)-enrichment

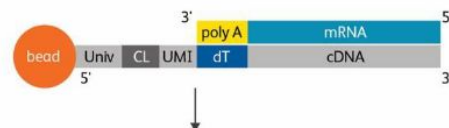
Amplified cDNA processing (dual index)



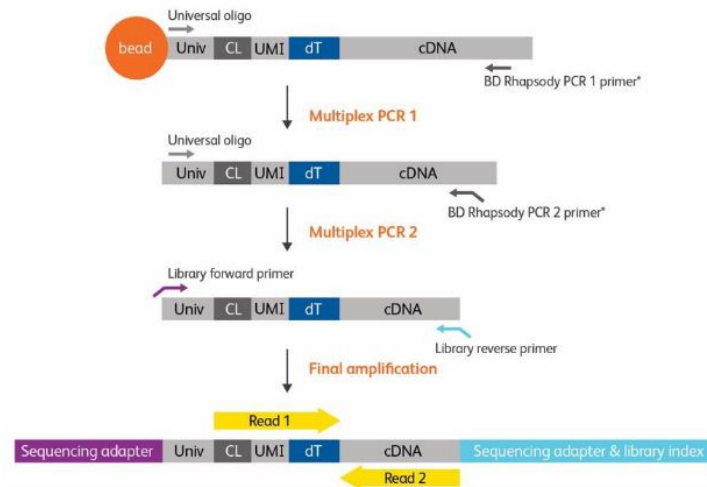
## BD Rhapsody targeted

Panel sequencing (primers for ~ 4000 genes)

cDNA archived on bead and tagged with cell label and molecular index

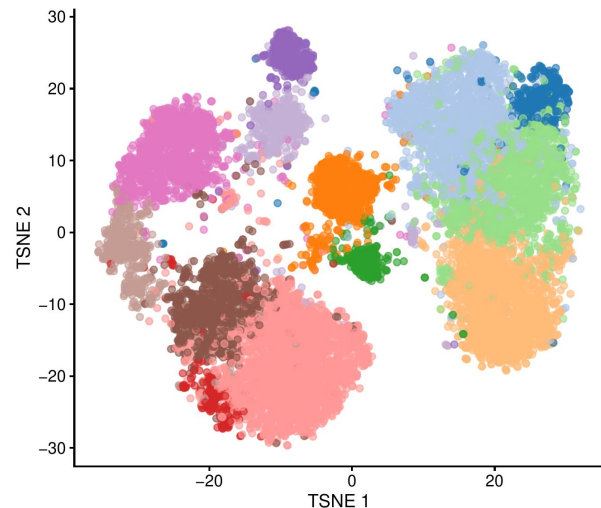


Amplification of cDNA using targeted primer panels



# Whole transcriptome reference to annotate a targeted sequencing dataset

## Manual annotation

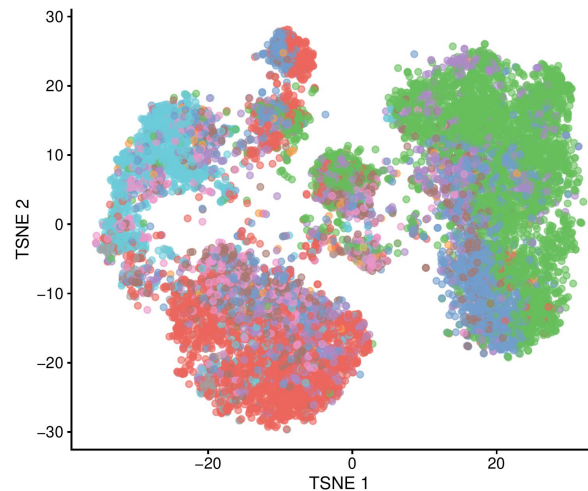


cell\_type

- CD4+ CD25+ FOXP3+ Tregs
- CD4+ CD26+ CD45RO+ KLRB1+ memory
- CD4+ CD5+ CD6+ cytotoxic
- CD4+ naive
- CD4+ NK-like FCGR3A+
- CD4+ RGS1+ ICOS+ memory
- CD8+ CD45RA+ FCGR3A+ effector memory
- CD8+ effector memory
- CD8+ naive
- CD8+ tissue resident memory
- CD8+ TNF+ IFNG+ effector memory
- gamma/delta T cell
- NKT cells



## SingleR annotation



mo\_label

- B cells
- Basophils
- CD4+ T cells
- CD8+ T cells
- Dendritic cells
- Monocytes
- Neutrophils
- NK cells
- Progenitors
- T cells

# Take home messages for cell type assignment

## **Automated cell type assignment:**

- Works well for common cell populations sequenced with whole transcriptome sequencing.
- Does not work well if you enrich for rare cell populations (NKT cells, atypical B cells)
- Does not work well for other sequencing approaches.

## **Recommendations from my own experience:**

- Check that the markers you expect are also expressed in the clusters.
- If you have many different cell types, split the data into subpopulations (B cells, T cells, tumor cells...). Independent subsequent analysis.
- After annotating, save an annotated intermediate object for downstream analysis.

## *Outline of the lecture*

1. Cell types and phenotypes
- 2. Interactions of immune cells**
3. Antigen specificity

# Ligand - Receptor interactions

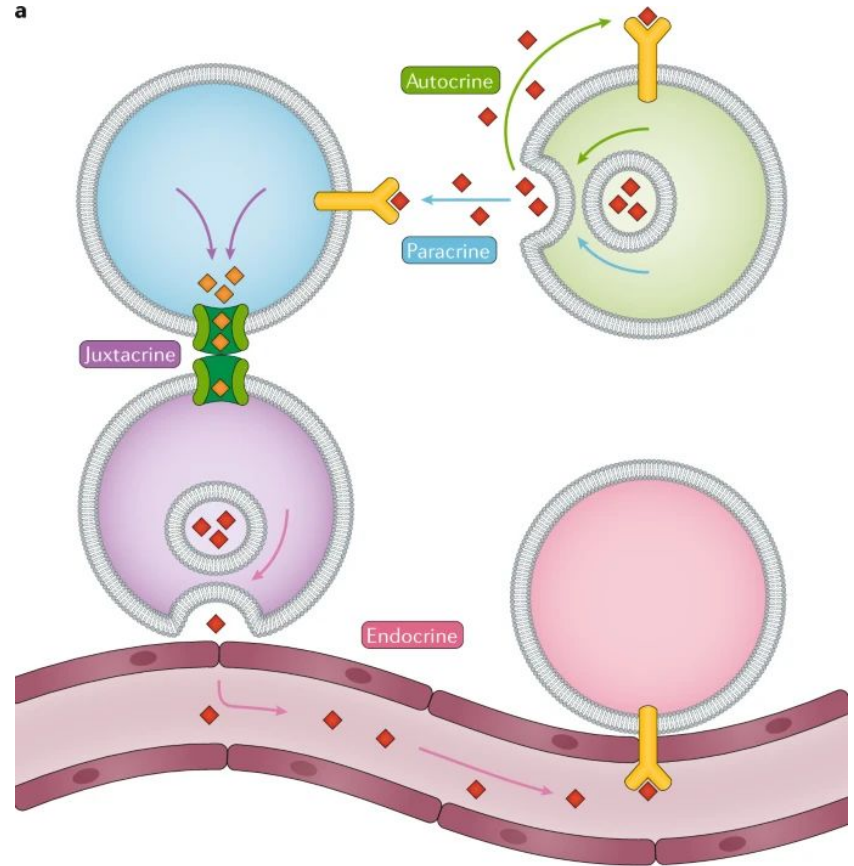
Here: Receptors and ligands encoded in the germline

Experimental measurements:

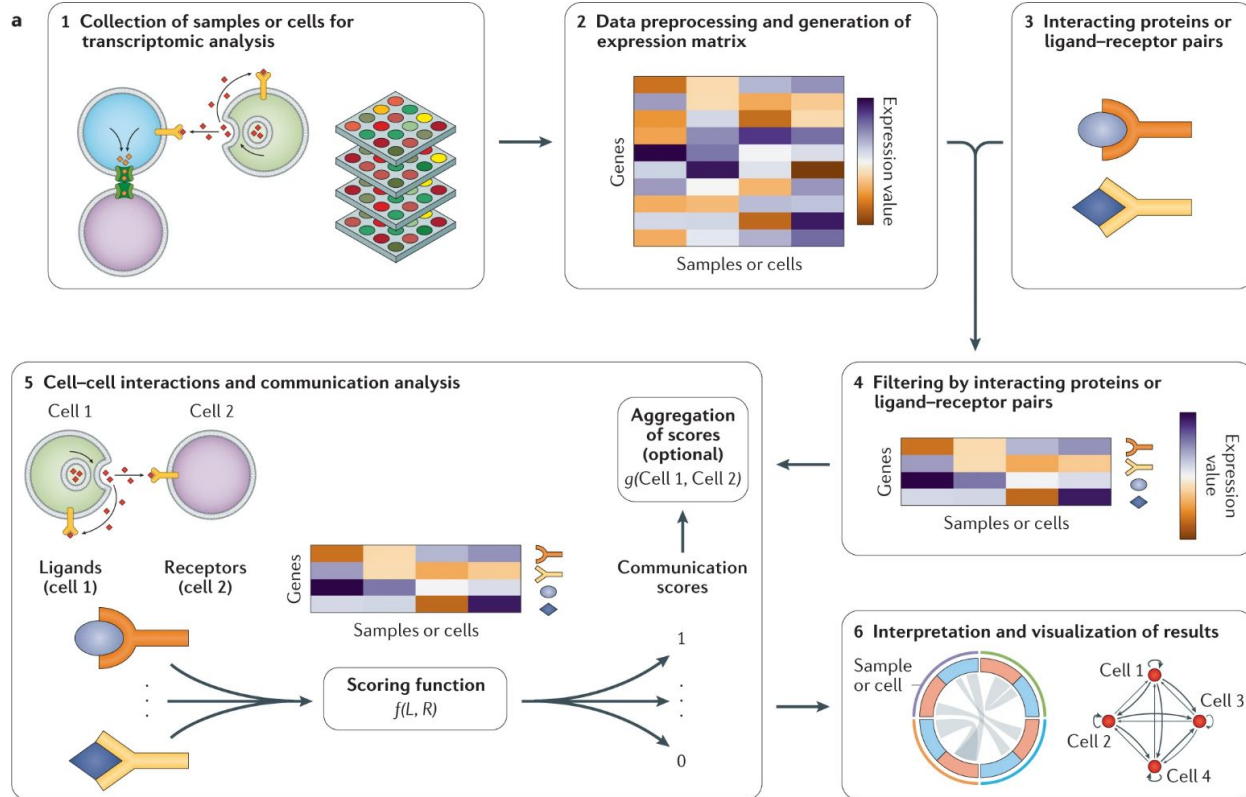
- Mass spectrometry of complexes
- Binding assays
- Affinity measurements

Databases of ligands and receptors:

- Cellinker
- CellChat
- CellPhoneDB
- iCELLNET
- ...



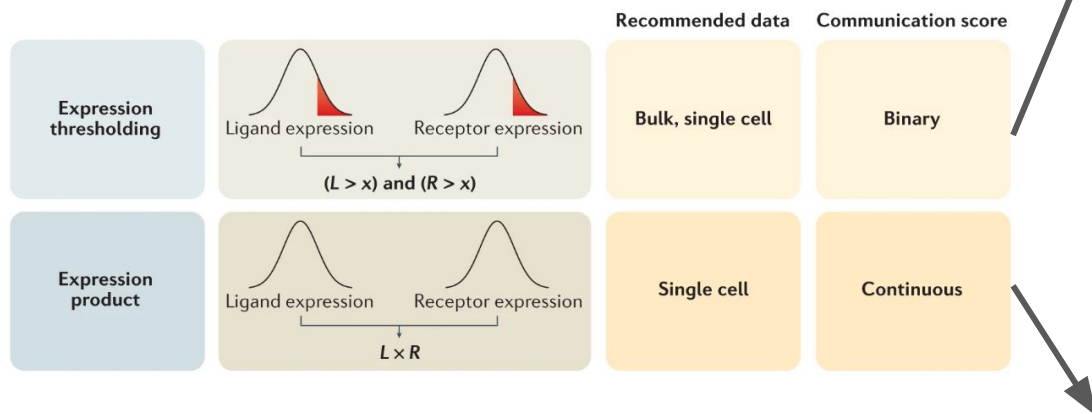
# Workflow for scoring receptor-ligand interactions





# Different scoring methods

b



**Expression thresholding**

	L (cell 1)	R (cell 4)	Communication scores
	1	1	1
	0	0	0
	0	1	0
	1	1	1
	1	0	0

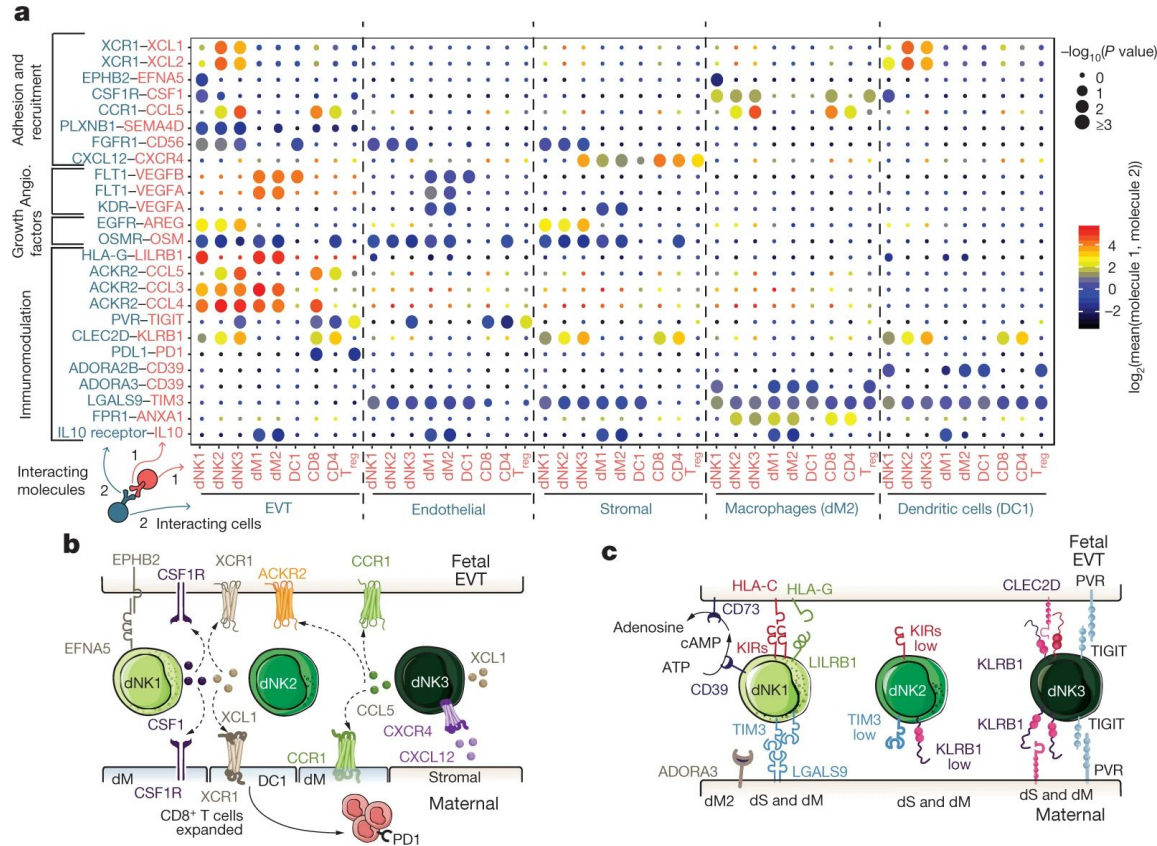
**Expression product**

	L (cell 1)	R (cell 4)	Communication scores
	3.81	3.46	13.17
	3.17	3.17	10.05
	3.00	6.91	20.72
	4.00	6.91	27.63
	5.32	3.17	16.87

Please see reference for complete list of tools and methods...

*Armingol et al. Nat Rev Genetics (2021)*

# Application: Identifying receptor-ligand interactions between cell types



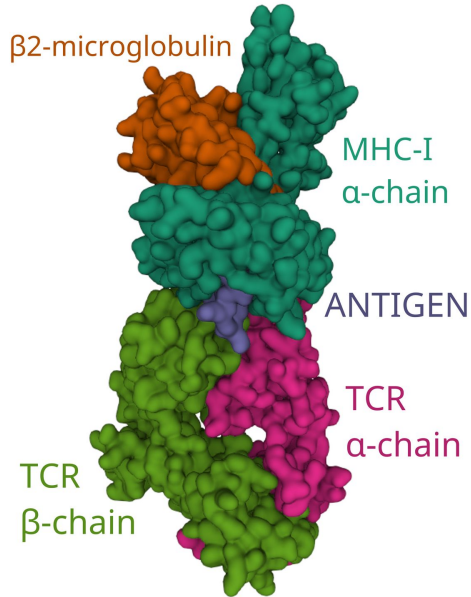
Vento-Tormo (2018)

## *Outline of the lecture*

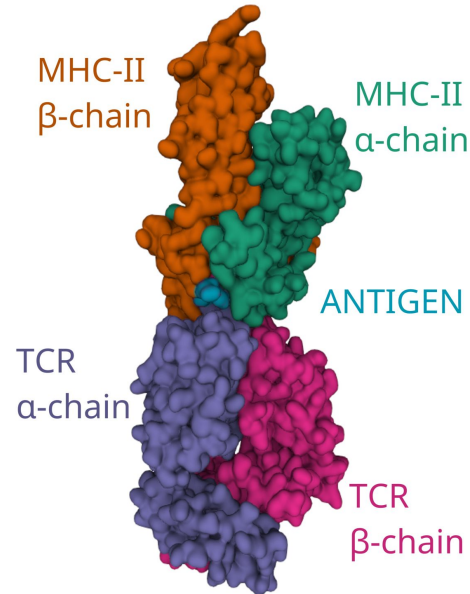
1. Cell types and phenotypes
2. Interactions of immune cells
- 3. Antigen specificity**

# Adaptive immune receptors are specific for a particular antigen

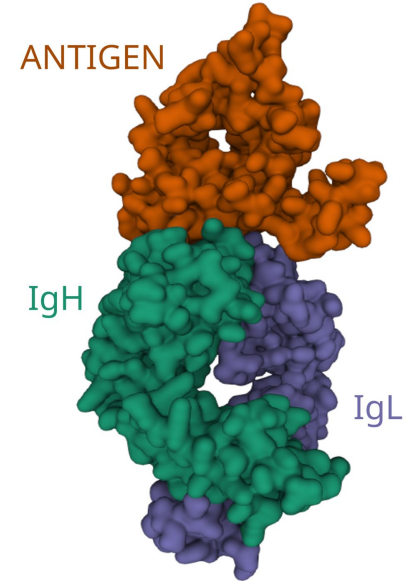
## MHC-I-antigen-TCR binding



## MHC-II-antigen-TCR binding



## Antibody-antigen binding



# Annotating immunoglobulin sequences

				-----><-----	-----CDR3-IMGT----->	
V	94.9% (280/295)	Query_1 <a href="#">IGHV1-3*04</a>	362 271	ACGGCTGTTTATTTCTGTGCGAGAGATTTGAGTTGTACTAGTACTACCACCTGCCATAGGCCGTTGAAGACAAACTACGGTATGGACGTC	T A V Y F C A R D L S C T S T T T C H R P L K T N Y G M D V	451 295
D	85.0% (17/20)	<a href="#">IGHD2-2*02</a>	10	.....G....A.....	T A V Y Y C A R	
J	100.0% (50/50)	<a href="#">IGHJ6*02</a>	13	.....G.....G....T.....		29 29

## Features:

- V,D,J usage
- CDR/FWR
- Somatic hypermutations

# AIRR exchange format

Standard format for annotating adaptive immune receptor sequences

## AIRR common repositories

Archives for AIRR sequences and metadata  
e.g. iReceptor public archive



## AIRR Rearrangement Schema

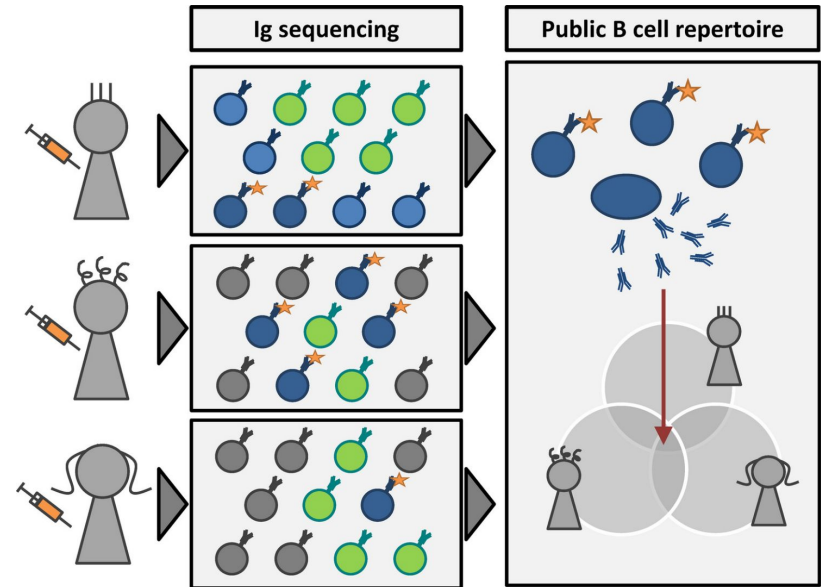
Input	Alignment Annotations	Alignment Positions	Region Sequence
<ul style="list-style-type: none"><li>• <b>sequence</b></li><li>• <b>sequence_aa</b></li></ul>	<ul style="list-style-type: none"><li>• <b>sequence_alignment</b></li><li>• <b>sequence_alignment_aa</b></li><li>• <b>germline_alignment</b></li><li>• <b>germline_alignment_aa</b></li><li>• <b>v_cigar</b></li><li>• <b>v_identity</b></li><li>• <b>v_score</b></li><li>• <b>v_support</b></li><li>• <b>d_cigar</b></li><li>• <b>d_identity</b></li><li>• <b>d_score</b></li><li>• <b>d_support</b></li><li>• <b>j_cigar</b></li><li>• <b>j_identity</b></li><li>• <b>j_score</b></li><li>• <b>j_support</b></li><li>• <b>c_cigar</b></li><li>• <b>c_identity</b></li><li>• <b>c_score</b></li><li>• <b>c_support</b></li><li>• <b>v_sequence_alignment</b></li><li>• <b>v_sequence_alignment_aa</b></li><li>• <b>d_sequence_alignment</b></li><li>• <b>d_sequence_alignment_aa</b></li><li>• <b>j_sequence_alignment</b></li><li>• <b>j_sequence_alignment_aa</b></li><li>• <b>c_sequence_alignment</b></li><li>• <b>c_sequence_alignment_aa</b></li><li>• <b>v_germline_alignment</b></li><li>• <b>v_germline_alignment_aa</b></li><li>• <b>d_germline_alignment</b></li><li>• <b>d_germline_alignment_aa</b></li><li>• <b>j_germline_alignment</b></li><li>• <b>j_germline_alignment_aa</b></li><li>• <b>c_germline_alignment</b></li><li>• <b>c_germline_alignment_aa</b></li></ul>	<ul style="list-style-type: none"><li>• <b>v_sequence_start</b></li><li>• <b>v_sequence_end</b></li><li>• <b>v_germline_start</b></li><li>• <b>v_germline_end</b></li><li>• <b>v_alignment_start</b></li><li>• <b>v_alignment_end</b></li><li>• <b>d_sequence_start</b></li><li>• <b>d_sequence_end</b></li><li>• <b>d_germline_start</b></li><li>• <b>d_germline_end</b></li><li>• <b>d_alignment_start</b></li><li>• <b>d_alignment_end</b></li><li>• <b>j_sequence_start</b></li><li>• <b>j_sequence_end</b></li><li>• <b>j_germline_start</b></li><li>• <b>j_germline_end</b></li><li>• <b>j_alignment_start</b></li><li>• <b>j_alignment_end</b></li></ul>	<ul style="list-style-type: none"><li>• <b>fwr1</b></li><li>• <b>fwr1_aa</b></li><li>• <b>cdr1</b></li><li>• <b>cdr1_aa</b></li><li>• <b>fwr2</b></li><li>• <b>fwr2_aa</b></li><li>• <b>cdr2</b></li><li>• <b>cdr2_aa</b></li><li>• <b>fwr3</b></li><li>• <b>fwr3_aa</b></li><li>• <b>cdr3</b></li><li>• <b>cdr3_aa</b></li><li>• <b>fwr4</b></li><li>• <b>fwr4_aa</b></li><li>• <b>np1</b></li><li>• <b>np1_aa</b></li><li>• <b>np2</b></li><li>• <b>np2_aa</b></li></ul>
<h3>Identifiers</h3> <ul style="list-style-type: none"><li>• <b>sequence_id</b></li><li>• <b>rearrangement_id</b></li><li>• <b>rearrangement_set_id</b></li><li>• <b>cell_id</b></li><li>• <b>clone_id</b></li><li>• <b>germline_database</b></li></ul>	<h3>Primary Annotations</h3> <ul style="list-style-type: none"><li>• <b>locus</b></li><li>• <b>v_call</b></li><li>• <b>d_call</b></li><li>• <b>j_call</b></li><li>• <b>c_call</b></li><li>• <b>rev_comp</b></li><li>• <b>productive</b></li><li>• <b>vj_in_frame</b></li><li>• <b>stop_codon</b></li><li>• <b>junction</b></li><li>• <b>junction_aa</b></li><li>• <b>duplicate_count</b></li><li>• <b>consensus_count</b></li></ul>	<h3>Junction Lengths</h3> <ul style="list-style-type: none"><li>• <b>junction_length</b></li><li>• <b>np1_length</b></li><li>• <b>np2_length</b></li><li>• <b>n1_length</b></li><li>• <b>n2_length</b></li><li>• <b>p3v_length</b></li><li>• <b>p5d_length</b></li><li>• <b>p3d_length</b></li><li>• <b>p5j_length</b></li></ul>	<h3>Region Positions</h3> <ul style="list-style-type: none"><li>• <b>fwr1_start</b></li><li>• <b>fwr1_end</b></li><li>• <b>cdr1_start</b></li><li>• <b>cdr1_end</b></li><li>• <b>fwr2_start</b></li><li>• <b>fwr2_end</b></li><li>• <b>cdr2_start</b></li><li>• <b>cdr2_end</b></li><li>• <b>fwr3_start</b></li><li>• <b>cdr3_start</b></li><li>• <b>cdr3_end</b></li><li>• <b>fwr4_start</b></li><li>• <b>fwr4_end</b></li></ul>

# Adaptive immune receptor repertoire (AIRR)

## Repertoire:

All B or T cells with their antigen receptor present in an individual at a given time

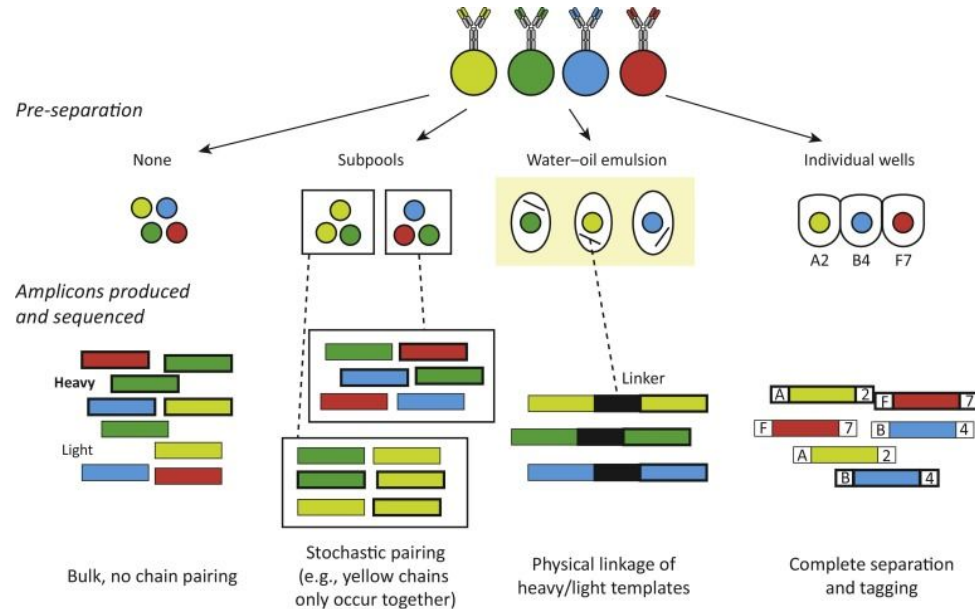
- $\sim 10^{12}$  possible combinations
- “Public” receptors are extremely rare



# Single-cell sequencing including AIRR (VDJ-seq)

First single-cell receptors sequenced long before the single-cell era!

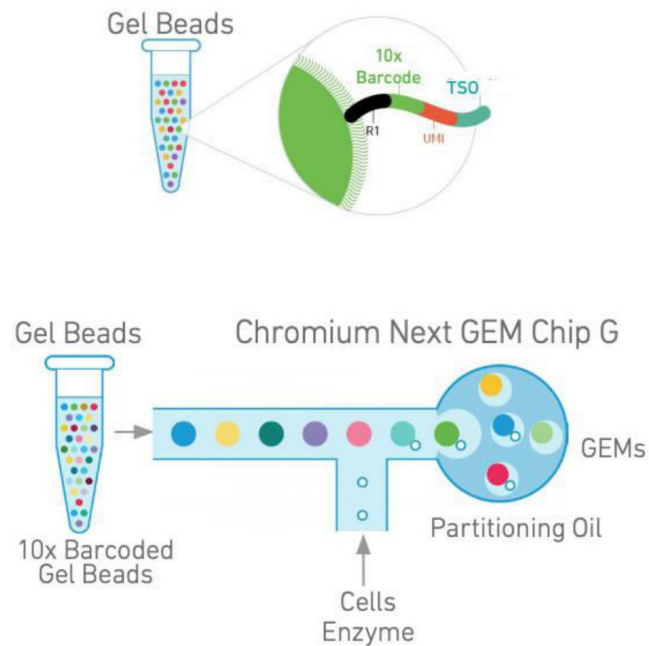
**Challenge: maintain heavy/light chain association.**



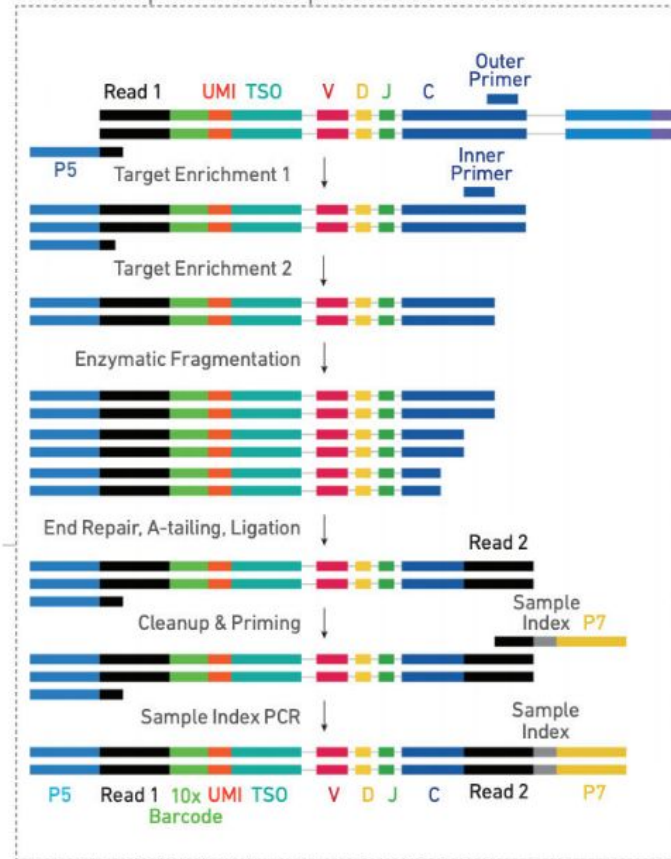
Wardemann, Busse (2017)



# 10x genomics VDJ sequencing: Linking heavy and light chain via the cellular barcode

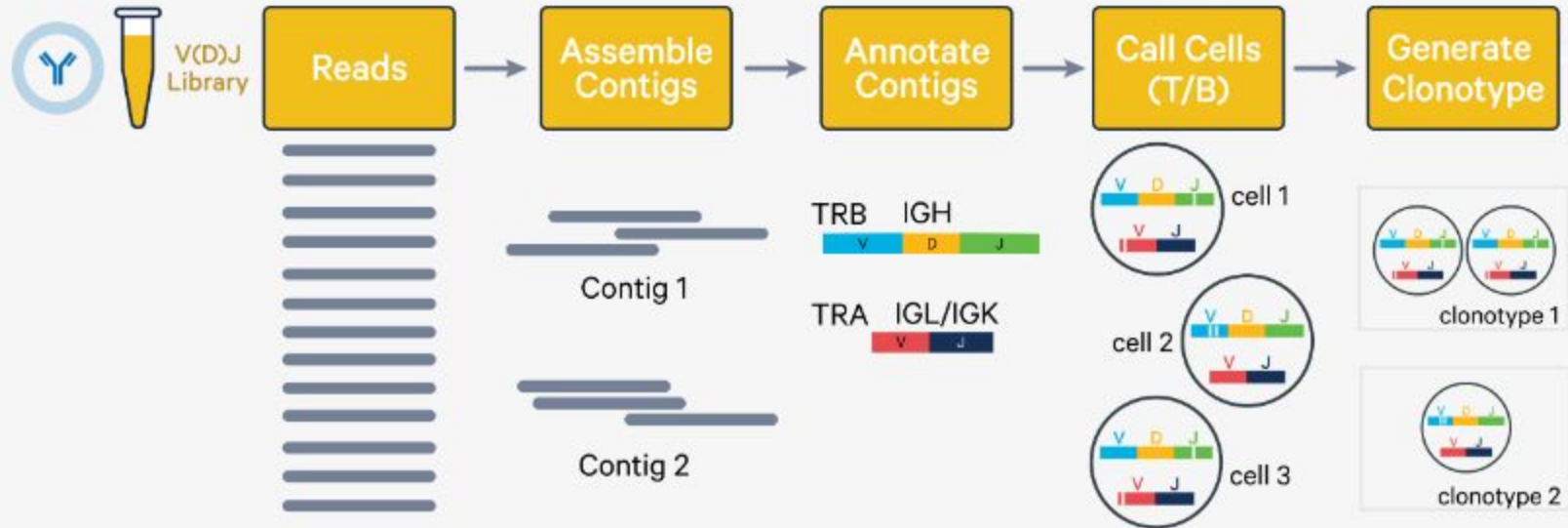


Pooled amplified cDNA processed in bulk

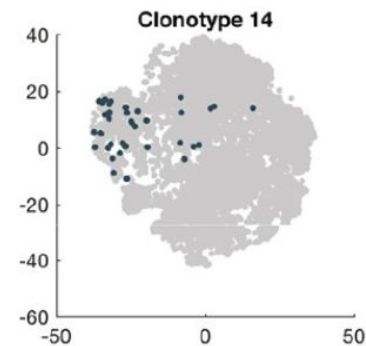
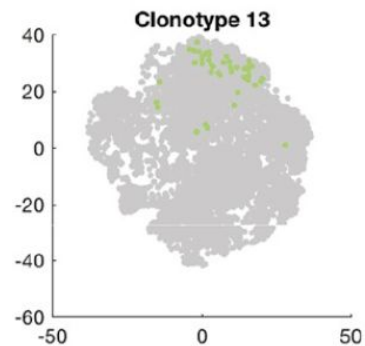
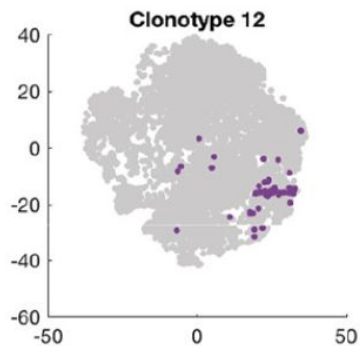
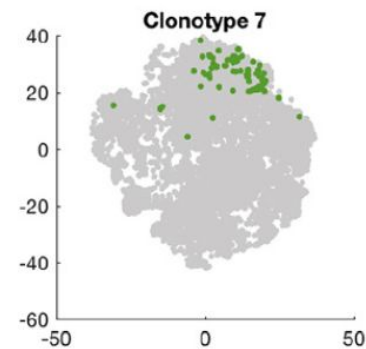
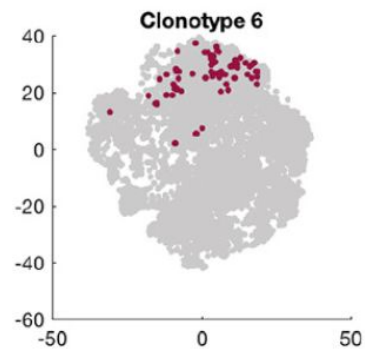
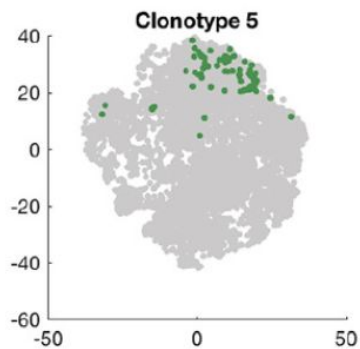
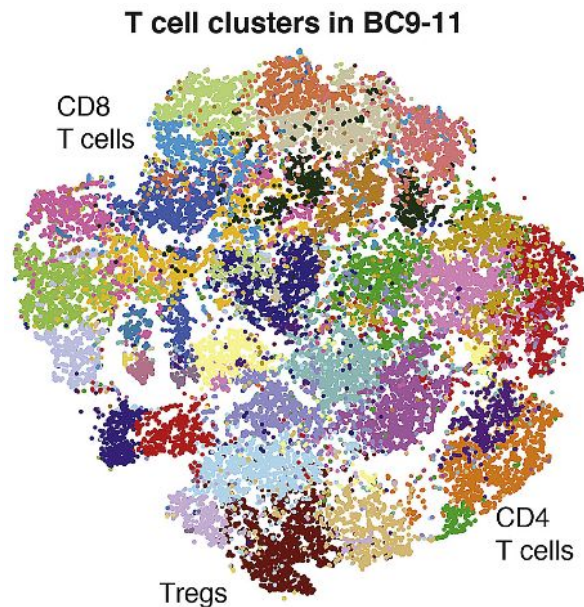


# 10x genomics VDJ - read annotation

## Algorithm overview



# TCR Clonotypes and single cell transcriptomic data



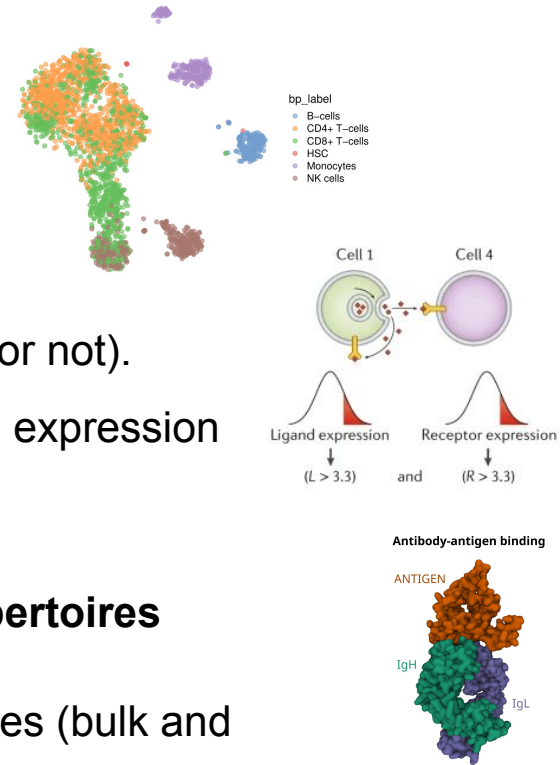
# Summary

## Immune cell types and interactions

- How to automatically annotation immune cell types (or not).
- How to score receptor ligand-receptor interactions in expression datasets.

## Antigen specificity and adaptive immune receptor repertoires

- How to annotate adaptive immune receptor repertoires (bulk and single-cell).



# Further reading

OSCAR: <https://bioconductor.org/books/release/OSCA/>

SingleR book: <http://bioconductor.org/books/release/SingleRBook/>

Review cellular interactions:

Armingol, E., Officer, A., Harismendy, O. *et al.* Deciphering cell–cell interactions and communication from gene expression. *Nat Rev Genet* 22, 71–88 (2021). <https://doi.org/10.1038/s41576-020-00292-x>

Review B/T cell repertoires:

Philip Bradley and Paul G. Thomas. Using T Cell Receptor Repertoires to Understand the Principles of Adaptive Immune Recognition. *Annual Reviews Immunology* (2019). <https://doi.org/10.1146/annurev-immunol-042718-041757>

Katharina Imkeller, Hedda Wardemann. Assessing human B cell repertoire diversity and convergence. *Immunological Reviews* (2018). <https://doi.org/10.1111/imr.12670>