Immune receptor repertoires in SC technologies

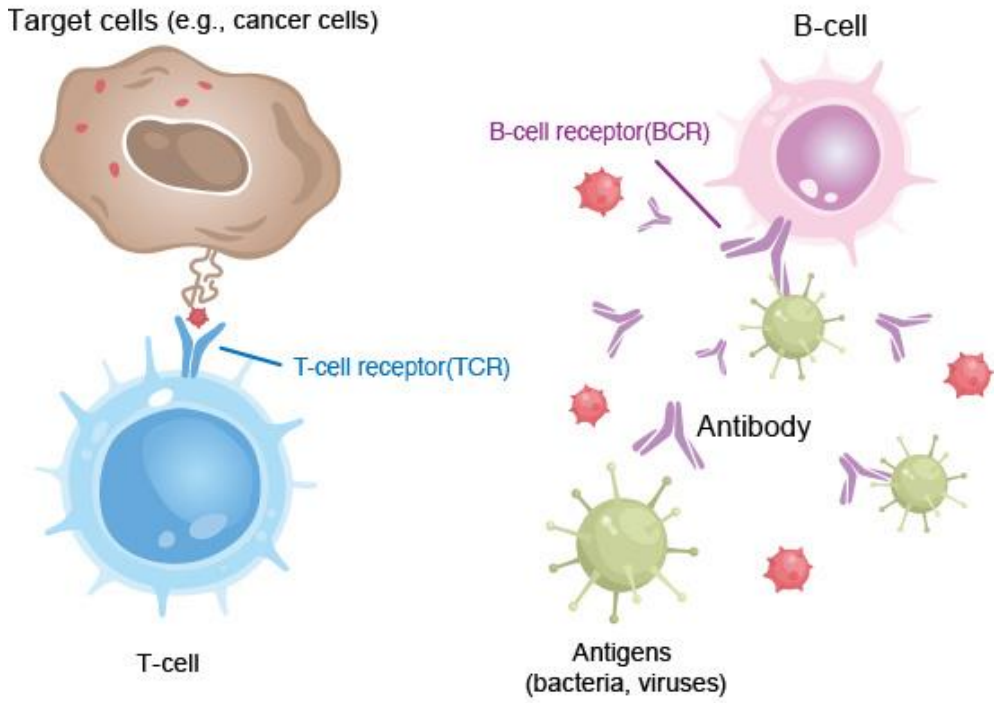
Anna Lorenc (Ania)
al16@sanger.ac.uk

TRYNKA GROUP

wellcome sanger institute

TCRs and BCRs (AIRR): keystones of adaptive immunity

TCR



BCR

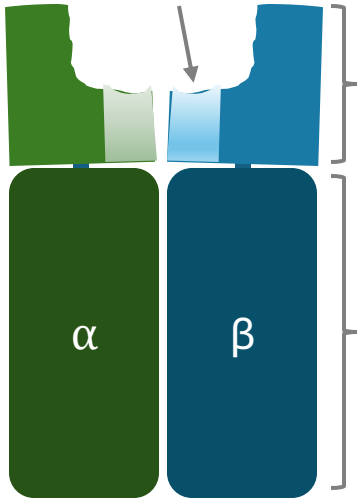
- Present on all T cells
- Reacts to a molecule (usually a short peptide) in HLA (MHC) context
- Results in the cell activation
- Depending on the type of the T cell: ↑ or ↓ immune response

- Present on all B cells
- React to a molecule
- When secreted – antibody
- Isotypes
- Many immune cells react to the external ‘handle’ of the antibody

picture: www.repertoire.co.jp

TCRs and BCRs (AIRR): keystones of adaptive immunity

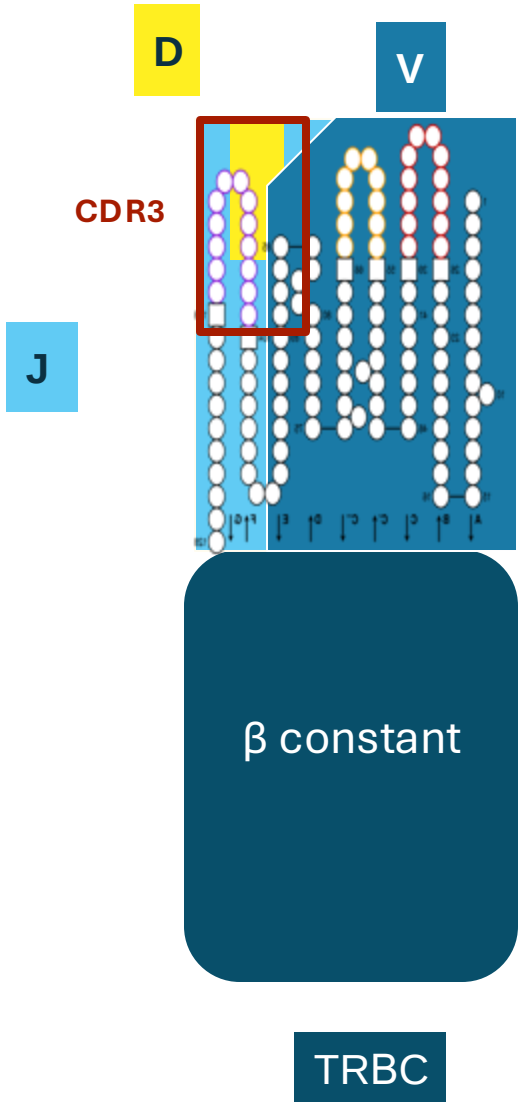
Complementarity
Determining
Region
3



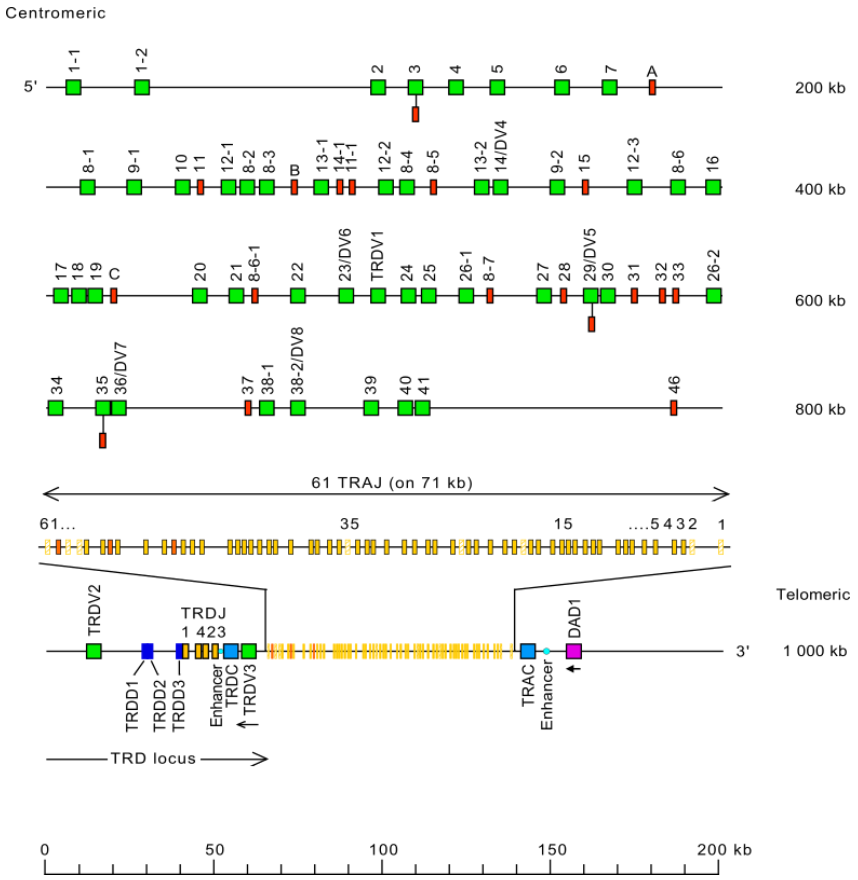
variable region

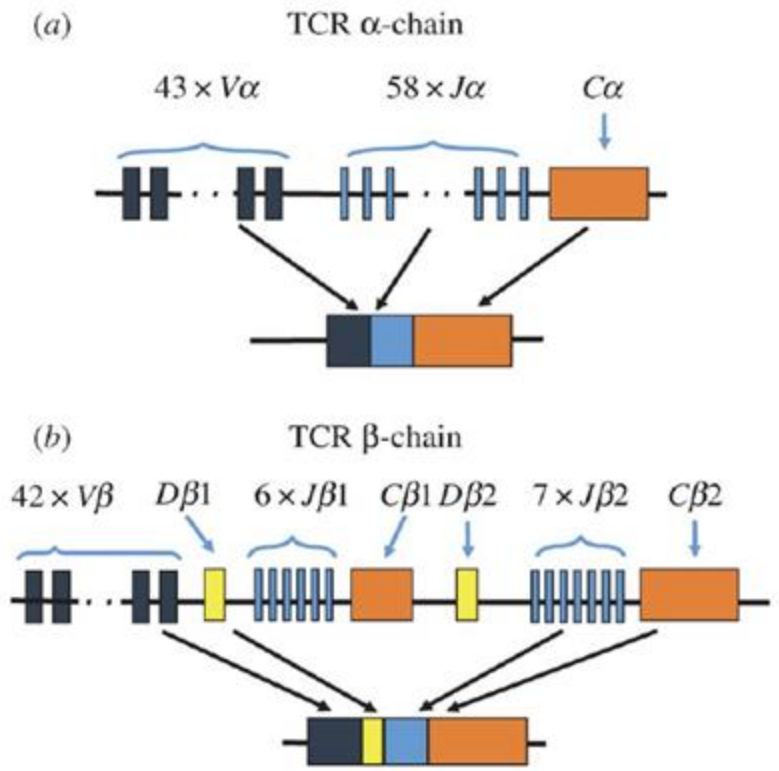
constant region

coded by...



TCRs and BCRs (AIRR): keystones of adaptive immunity





Each cell

T cell

Each cell

T cell

Somatic recombination

Pictures: IMGT.org, Janeway

TCRs and BCRs (AIRR): keystones of adaptive immunity

Somatic recombination: stochastic combinations of V, (D) and J gene rearrangements and imprecise joining contribute to the diversity of the TCR repertoire.

	V	J	D	V del	J del	Ins	combinations
α	43	58	-	15	12	15	4.3×10^{14}
β	42	13	3	10	12	18	5.1×10^{15}

$\alpha\beta = 10^{30}$

At any given time: 5×10^{11} T cells, 100g*; 10^{8} receptors
All possible 10^{21}g $= 10^{18}$ kg

\+ genetic diversity: alleles of every V, D, J gene → under-researched, most software rely on “known” germline variants

What the data tell us

What can we get from the single cell data?

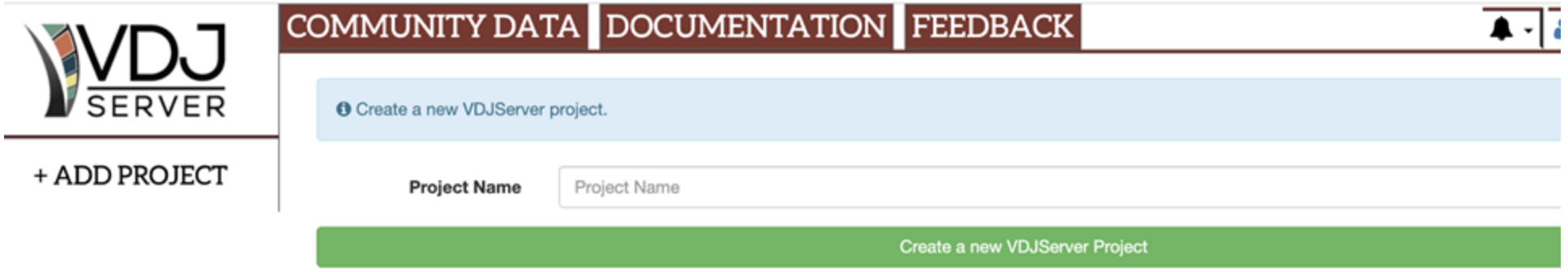
DNA TCR/BCR sequence unique to one cell*
If a cell multiplies (eg infection), more cells with identical receptor

→ We have a barcode for cells of the same origin
→ We can guess that a receptor is involved in something when many cells have the same receptor
→ We might detect an ongoing infection/inflammation
→ We might already know what receptor is against

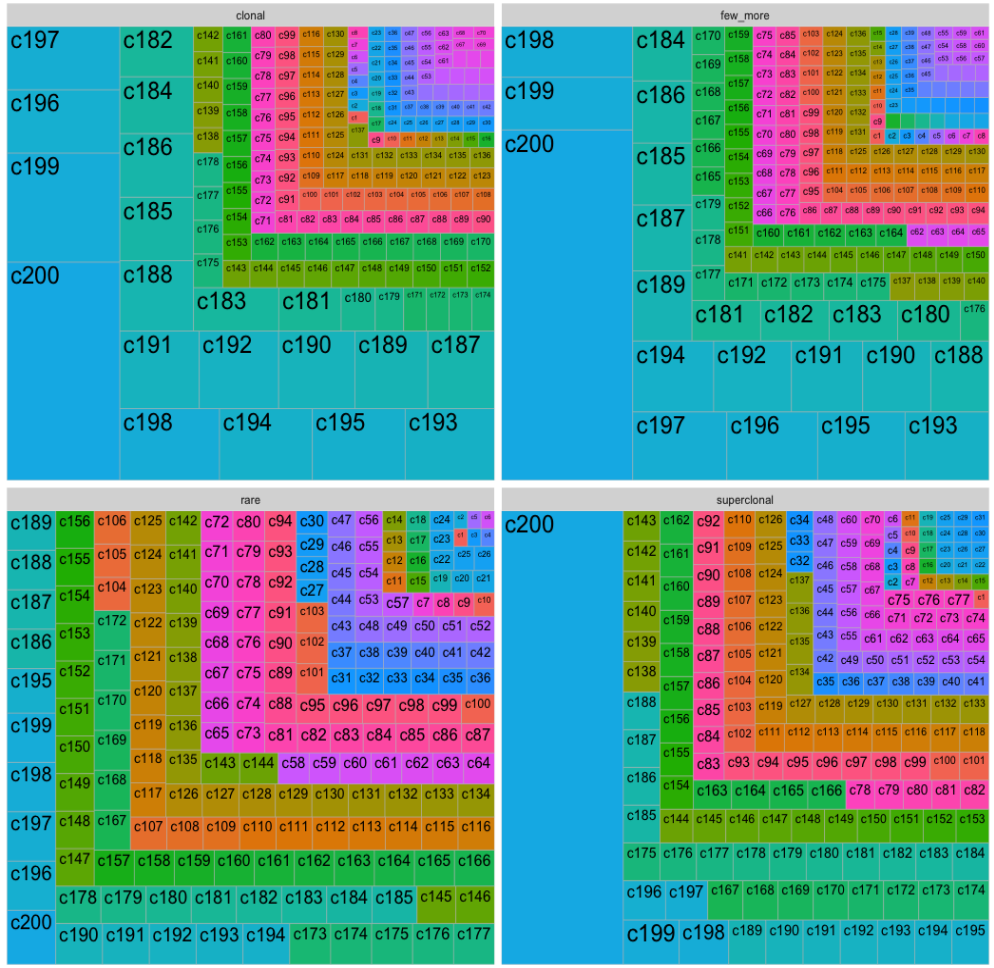
Constraints
→ Limited knowledge about TCR-antigen-(HLA)
→ *Not really unique
→ BCRs undergo the additional process of the somatic hypermutation
→ Very few TCRs/BCRs captured

What the data tell us

VDJdb



What the data tell us



- R = richness – number of unique clonotypes (each might be present in more than 1 cell)
- p_i = relative abundance of clonotype i (0-1)
- q = order of diversity
- N = total number of receptors (cells)

Hill diversity

$$ {}^qD = (\sum_{i=1}^{R} p_i^q)^{1/(1-q)} = exp(\frac{1}{1-q} ln(\sum_{i=1}^{R} p_i^q)) = exp({}^qH) $$

Renyi entropy

$$ H = \frac{1}{1-q} ln(\sum_{i=1}^{R} p_i^q) = ln(D^q) $$

Clonality

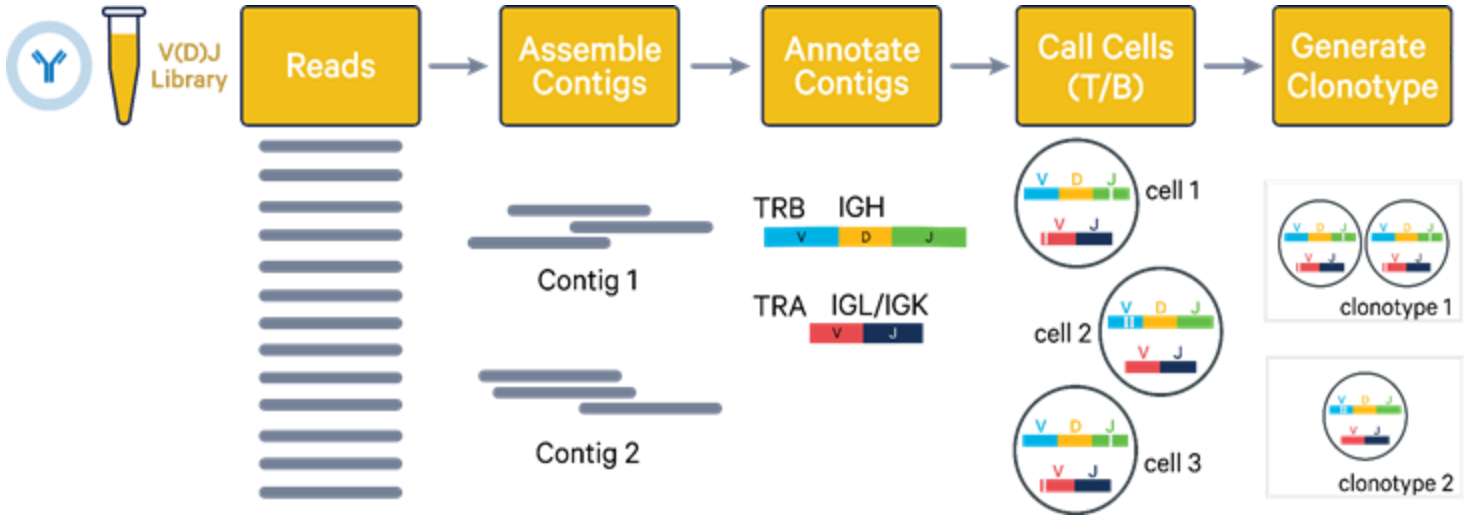
$$ C = 1 - \frac{H}{lnR} $$

Shannon entropy

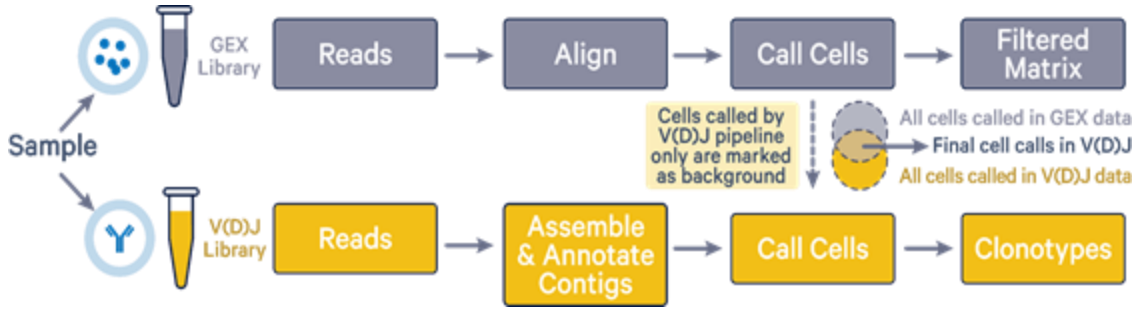
$$ H = -\sum_{i=1}^{R} p_i \ln p_i = ln(D^1) $$

How the data is produced

Within one barcode:



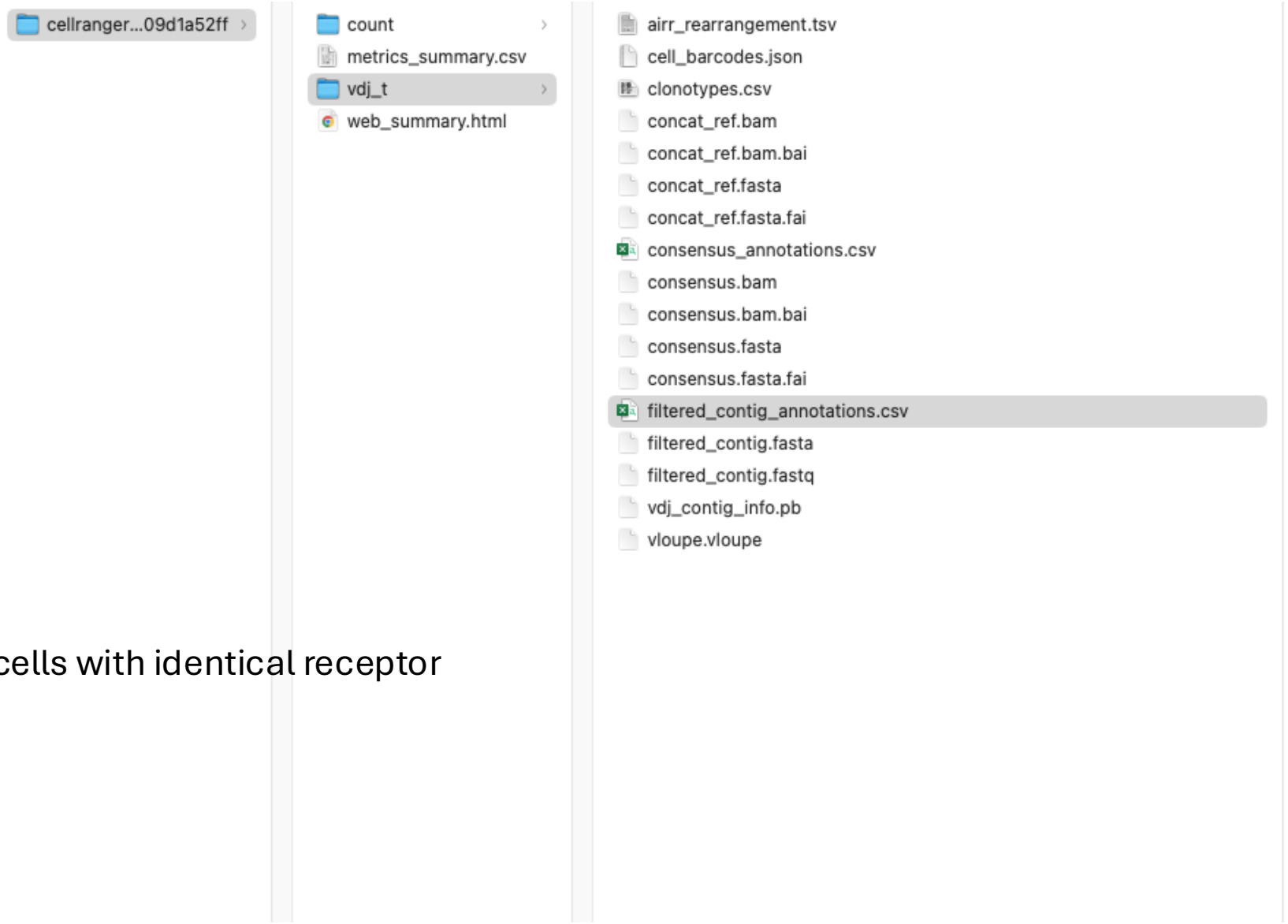
Cell assignment is from GEX



10x GENOMICS™ **CellRanger**

Other technologies similar approach

How the data is produced



Per identified chain info
Alignments to the reference
Per clonotype info
Per cell info

Clonotype = all cells with identical receptor

questions

NCBI > GEO > **Accession Display** Not logged in | Lo

GEO help: Mouse over screen elements for information.

Scope: Self Format: HTML Amount: Quick GEO accession: GSE145926 GO

Series GSE145926 Query DataSets for GSE145926

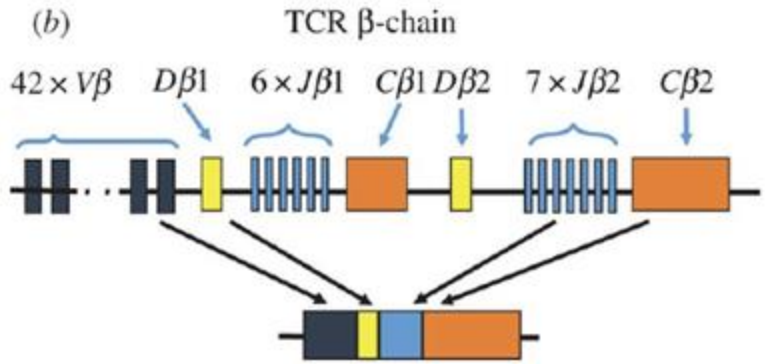
Status	Public on Apr 22, 2020
Title	Single-cell landscape of bronchoalveolar immune cells in COVID-19 patients
Organism	Homo sapiens
Experiment type	Expression profiling by high throughput sequencing Other
Summary	Immune characteristics associated with Coronavirus Disease-2019 (COVID-19) severity are currently unclear. We characterized bronchoalveolar lavage fluid (BALF) immune cells from patients with varying severity of COVID-19 disease and from healthy subjects using single-cell RNA-sequencing. Proinflammatory monocyte-derived macrophages were abundant in the BALF from severe COVID-9 patients. Moderate cases were characterized by the presence of highly clonally expanded tissue-resident CD8+ T cells. This atlas of the bronchoalveolar immune-microenvironment suggests potential mechanisms underlying pathogenesis and recovery in COVID-19.
Overall design	Using 10x genomics to measure single-cell RNA sequence (scRNA-seq)/TCR-seq to comprehensively characterize the lung immune microenvironment in the bronchoalveolar lavage fluid (BALF) from 6 severe and 3 moderate COVID-19 patients and 3 healthy control.
Contributor(s)	Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, Cheng L, Li J, Wang X, Wang F, Liu L, Amit I, Zhang S, Zhang Z
Citation(s)	Liao M, Liu Y, Yuan J, Wen Y et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med* 2020 Jun;26(6):842-844. PMID: 32398875 Zhang Z, Zhang L, Wang K, Xie T et al. Single-cell landscape of bronchoalveolar immune cells in patients with immune checkpoint inhibitor-related pneumonitis. *NPJ Precis Oncol* 2024 Oct 5;8(1):226. PMID: 39369126 medRxiv: https://doi.org/10.1101/2020.02.23.20026690
Submission date	Feb 25, 2020
Last update date	Nov 05, 2024
Contact name	Zheng Zhang
Organization name	Shenzhen 3rd People's Hospital
Street address	No. 29, Bulan Road
City	Shenzhen
State/province	Guangdong
ZIP/Postal code	454171
Country	China
Platforms (1)	GPL23227 BGISEQ-500 (Homo sapiens)
Samples (21) More...	GSM4339769 BALF, C141 (scRNA-seq) GSM4339770 BALF, C142 (scRNA-seq)

GSM4339769 BALF, C141 (scRNA-seq)
GSM4339770 BALF, C142 (scRNA-seq)
GSM4339771 BALF, C143 (scRNA-seq)
GSM4339772 BALF, C144 (scRNA-seq)
GSM4339773 BALF, C145 (scRNA-seq)
GSM4339774 BALF, C146 (scRNA-seq)
GSM4385990 BALF, C141 (TCR-seq)
GSM4385991 BALF, C142 (TCR-seq)
GSM4385992 BALF, C143 (TCR-seq)
GSM4385993 BALF, C144 (TCR-seq)
GSM4385994 BALF, C145 (TCR-seq)
GSM4385995 BALF, C146 (TCR-seq)
GSM4475048 C51 (scRNA-seq)
GSM4475049 C52 (scRNA-seq)
GSM4475050 C100 (scRNA-seq)
GSM4475051 C148 (scRNA-seq)
GSM4475052 C149 (scRNA-seq)
GSM4475053 C152 (scRNA-seq)
GSM4475054 C148 (TCR-seq)
GSM4475055 C149 (TCR-seq)
GSM4475056 C152 (TCR-seq)

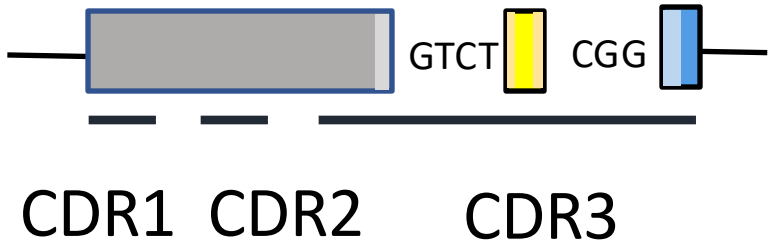
barcode	is_cell	contig_id	high_confide	length	chain	v_gene	d_gene	j_gene	c_gene	full_length	productive	cdr3	cdr3_nt	reads	umis	raw_clonotype_id	raw_
AAACCTGCATGGTCAT-1	TRUE	AAACCTGCATGGT	TRUE	492	TRA	TRAV13-2	None	TRAJ10	TRAC	TRUE	TRUE	CAEKSSGGG	TGTGCAGAG	37974	11	clonotype129	clon
AAACCTGCATGGTCAT-1	TRUE	AAACCTGCATGGT	TRUE	501	TRB	TRBV6-6	None	TRBJ2-5	TRBC2	TRUE	TRUE	CASSYGTTGI	TGTGCCAGC	12806	7	clonotype129	clon
AAACCTGCATGGTCAT-1	TRUE	AAACCTGCATGGT	TRUE	509	TRA	TRAV6	None	TRAJ40	TRAC	TRUE	TRUE	CALRSGTYKY	TGTGCTCTAC	17114	5	clonotype129	clon
AAACCTGCATGGTCAT-1	TRUE	AAACCTGCATGGT	TRUE	318	TRA	None	None	TRAJ5	TRAC	FALSE	FALSE	None	None	6288	2	clonotype129	None
AAACCTGGTTTAGCTG-1	TRUE	AAACCTGGTTTAG	TRUE	494	TRB	TRBV7-6	None	TRBJ2-3	TRBC2	TRUE	TRUE	CASRSIEADT	TGTGCCAGC	57140	6	clonotype130	clon
AAACCTGGTTTAGCTG-1	TRUE	AAACCTGGTTTAG	TRUE	497	TRA	TRAV17	None	TRAJ21	TRAC	TRUE	TRUE	CATDGDNFN	TGTGCTACG	8332	4	clonotype130	clon
AAACCTGTCAATCACG-1	TRUE	AAACCTGTCAATCA	TRUE	471	TRB	TRBV29-1	None	TRBJ2-7	TRBC2	TRUE	TRUE	CSVEGTATYE	TGCAGCGTT	39068	10	clonotype131	clon
AAACCTGTCAATCACG-1	TRUE	AAACCTGTCAATCA	TRUE	576	TRA	TRAV8-4	None	TRAJ32	TRAC	TRUE	TRUE	CAVSDGFGG	TGTGCTGTGA	30316	11	clonotype131	clon
AAACGGGAGAACTCGG-1	TRUE	AAACGGGAGAACT	TRUE	634	TRA	TRAV8-2	None	TRAJ41	TRAC	TRUE	TRUE	CVGNSGYALI	TGTGTTGGGA	6840	4	clonotype25	clon
AAACGGGAGAACTCGG-1	TRUE	AAACGGGAGAACT	TRUE	471	TRB	TRBV10-2	None	TRBJ2-3	TRBC2	TRUE	TRUE	CASNLAGPTI	TGCGCCAGC	38840	17	clonotype25	clon
AAACGGGAGAACTCGG-1	TRUE	AAACGGGAGAACT	TRUE	387	TRB	None	None	TRBJ1-5	TRBC1	FALSE	FALSE	None	None	16520	5	clonotype25	None
AAACGGGAGAACTCGG-1	TRUE	AAACGGGAGAACT	TRUE	607	TRB	None	None	TRBJ1-5	TRBC1	FALSE	FALSE	None	None	11888	3	clonotype25	None
AAACGGGAGAACTCGG-1	TRUE	AAACGGGAGAACT	TRUE	510	TRA	TRAV1-2	None	TRAJ6	TRAC	TRUE	FALSE	CAVPHQEEA	TGTGCTGTC	3442	3	clonotype25	None

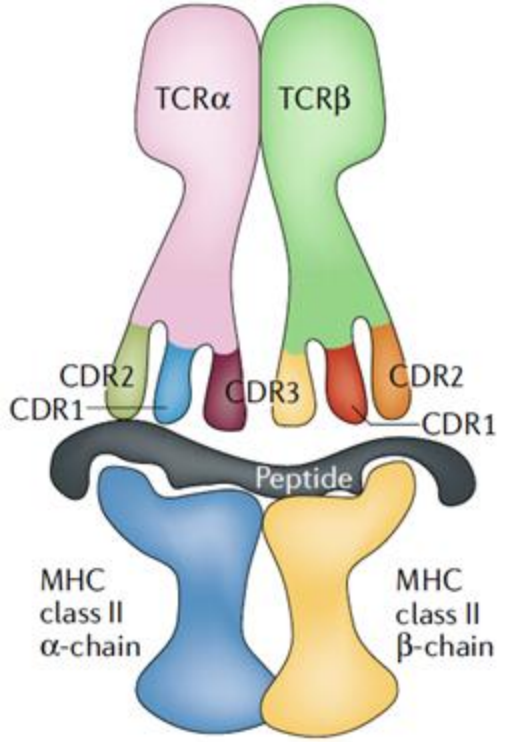
Additional slides

Most variable parts - **C**omplementarity **D**etermining **R**egions – interact with pMHC complex



Janeway





La Gruta *Nat Rev Imm 2018*

Final CDR3 length: 7-40 aminoacids
Median 13

IMGT – source of references

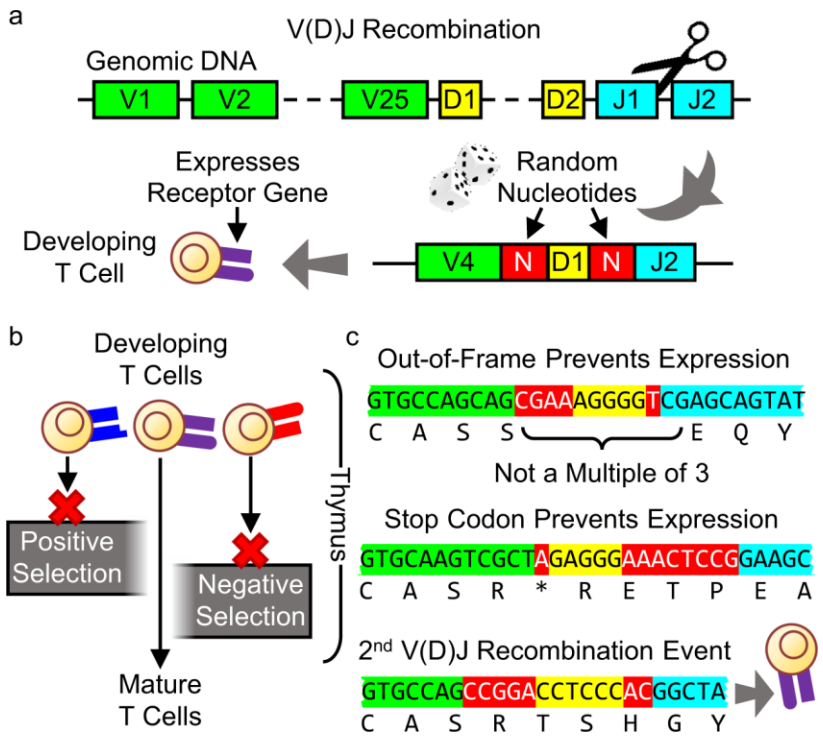
References
www.IMGT.org

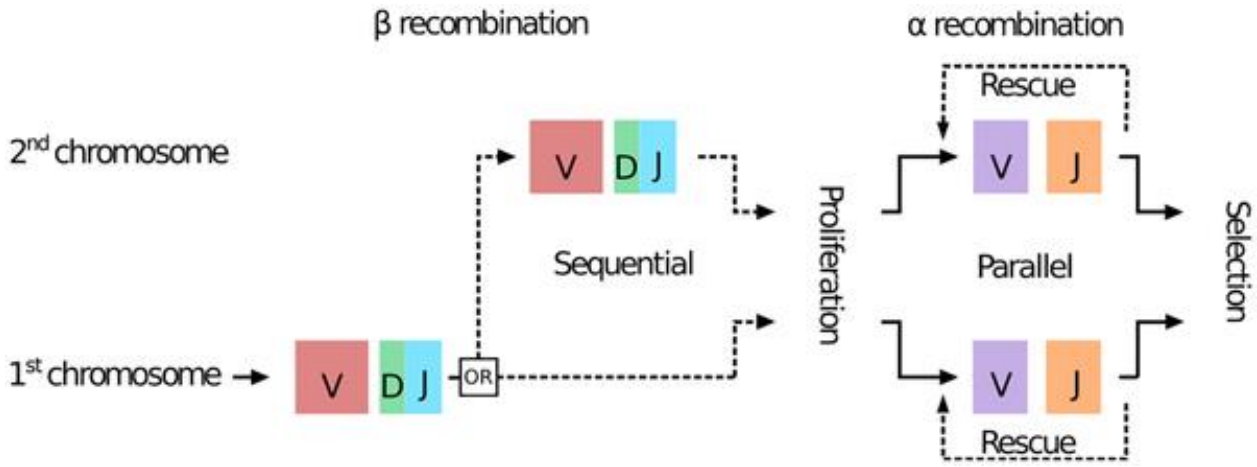
TRBV subgroup	TRBV gene name	Fct	TRBV allele name	Accession number	Confirmed by genetics and/or data	
2	2	F	V2*01	L36092/U66059	+	g65 ,R22 |a237 |
		(F)	V2*02	M62379		g65>a,R22>H|
		(F)	V2*03	M64351		|a237>g|
3	3-1	F	V3-1*01	U07977	+	t174 |c181 ,L61 |c225 |c256 ,c258 ,H86 |
		(F)	V3-1*02	L06889		t174>c|c181>a,L61>I|c225>a|c256>a,c258>a,H86>K|
4	4-1	F	V4-1*01	U07977	+	t93 |
		(F)	V4-1*02	M13855		t93>a|
	4-2	F	V4-2*01	U07975	+	t263 ,F88 |
		(F)	V4-2*02	X58811		t263>g,F88>C|
	4-3	F	V4-3*01	U07978	+	t84 |g183 |t263 ,F88 |
		(F)	V4-3*02	X58812		|t263>c,F88>S|
		(F)	V4-3*03	L06888		|g183>t|
		(F)	V4-3*04	X57616		t84>g|
5	5-1	F	V5-1*01	L36092/U66059	+	a2 ,K1 |a9 |t28 ,Y10 |a64 ,S22 |c137 ,P46 |c
		(F)	V5-1*02	M14271		a2>g,K1>R|a9>g|t28>c,Y10>H|a64>g,S22>G|c137>t,P46>L|c
	5-3	ORF	V5-3*01	X61439	+	g254 ,C85 |
		ORF	V5-3*02	AF009660	+	g254>a,C85>Y|
	5-4	F	V5-4*01	L36092/U66060	+	t60 |t212 ,71F |g257 ,86S |
		(F)	V5-4*02	X57615		|g257>a,86S>N|
		(F)	V5-4*03	S50547		t60>a|
		(F)	V5-4*04	X58804		|t212>c,71F>S|

Practicalities: not always possible/necessary, different references!
Underrepresentation of Non-Europeans

wellcome sanger institute

DNA level





Ostmeyer et al 21

Dupic et all, Plos Comp Biol 2019

- Nonproductive sequences

Bulk methods do not see it!

- Allelic exclusion is leaky: ~7% of cells with 2x TCRβ (1% both expr), 7-30% with 2xTCRα
- Final length: few to ~40 aas, majority 13aa
- Non-random frequency of specific recombination events → pgen

Some practical hints

The same VDJ – different CDR3
Different VDJ – the same CDR3

Convergence:

aa1==aa2
nuc1 != nuc2

Functional comparison: aminoacid level
“Tag” analysis: DNA level

HLA-dependence

Not always full TRB reconstruction possible

Sampling issues: →

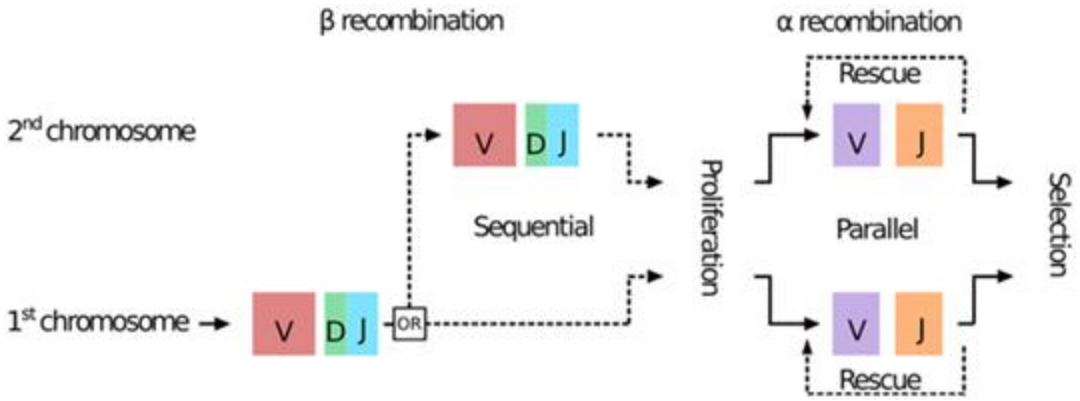
- Size of the repertoire
- Cell-type dependency
- Tisssue dependency
- Clonotypes of importance might be rare

The naive T-cell receptor repertoire has an extremely broad distribution of clone sizes
https://doi.org/10.7554/eLife.49900

Known TCR sequences: selection bias

Differences in recombination of TRA and TRB

- Both alleles recombine at the same time
- Rescue mechanism
- 1-2 recombined alleles (2 in 7-30% of cells)
- Positive/negative selection



Dupic et all, Plos Comp Biol 2019

- Has to interact with MHC in the thymus
- If recombination unsuccessful, another round
- Single product, leakiness <1%

Very very very very wasteful

- Recombination products Out of frame and with stop codons (non-productive sequences)
- Negative/positive selection

Both together: LOWER estimate of selection factor is 99.9%

We can estimate probability of a recombination-derived TCR

Select a V gene

↓ P1(V=V_i)

Select a J gene

↓ P2(J=J_i)

Select number of V deletions

↓ P3(m=1,2,3….n)

Select number of J deletions

↓ P4(m=1,2,3,…)

Select no. of nucleotide additions

→ P4(Nt=A,T,C,G)

p5,p6 ….

↺ Select nucleotide

Murugan et al. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. Proc Natl Acad Sci U S A. 2012 Oct 2;109(40):16161-6.

Sethna et al. OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. Bioinformatics. 2019 Sep 1;35(17):2974-2981

$$P_{\text{recomb}}(E) = P(V)P(D,J) \times P(\text{del}V|V)P(\text{del}J|J)P(\text{del}5'D, \text{del}3'D|D) \times P(\text{ins}VD)\prod_{i=1}^{\text{ins}VD} p_{VD}^{(2)}(x_i|x_{i-1})P(\text{ins}DJ)\prod_{i=1}^{\text{ins}DJ} p_{DJ}^{(2)}(y_i|y_{i-1}).$$

Probabilities to obtain from non-productive sequences

https://github.com/statbiophys/OLGA

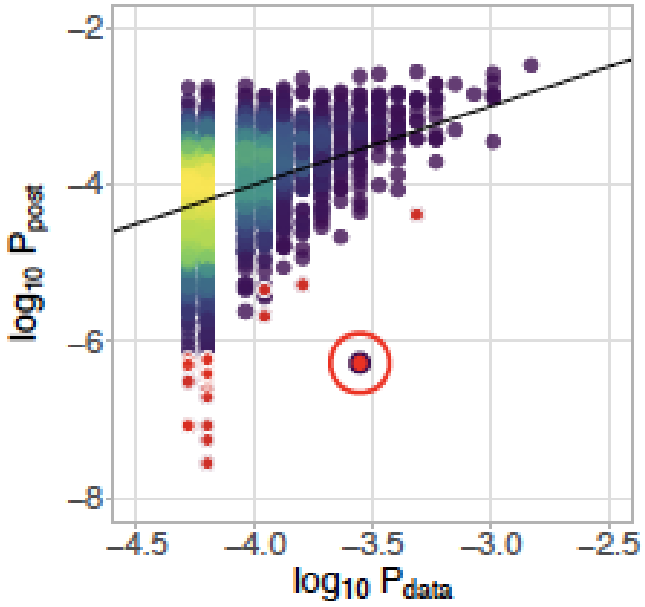
welcome sanger institute

We can estimate probability of a recombination-derived TCR

A nucleotide sequence with low probability of generation and shared between individuals

An aminoacid sequence with low probability of generation, coded by many nucleotides

- putatively selected sequences

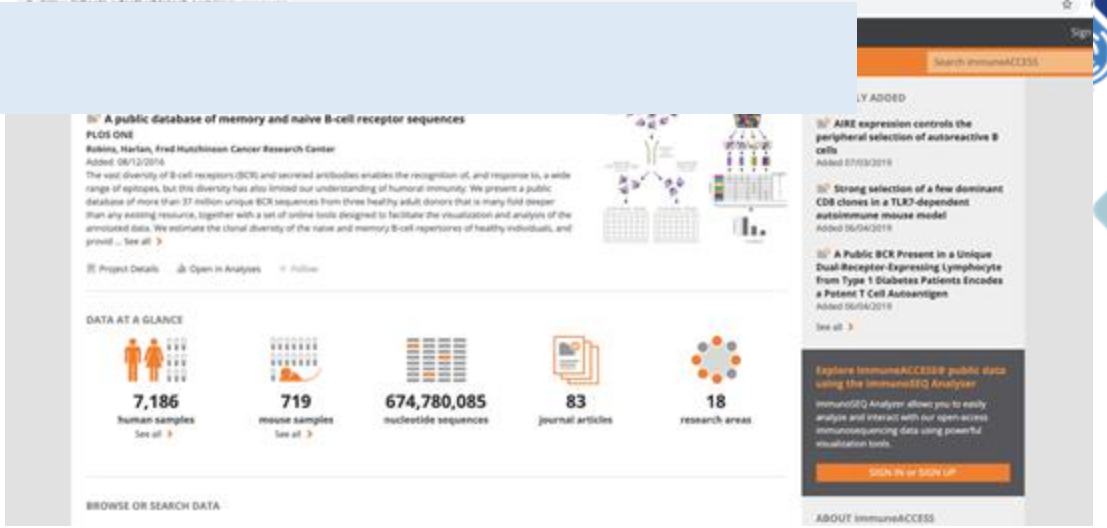


Pogorelyy, Elife 2018

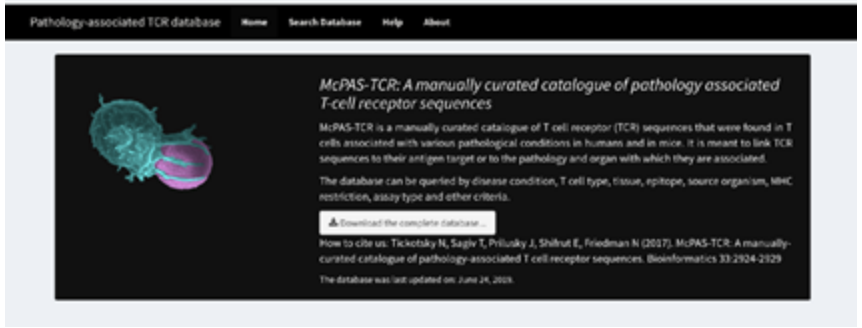
Databases and repositories

VDJdb

ImmuneACCESS



McPass-TCR



FASTQ: NCBI SRA