CS-E5875 High-Throughput Bioinformatics
Immune cell receptor sequencing

Harri Lähdesmäki
(most slides by Emmi Jokinen)

Department of Computer Science
Aalto University

November 27, 2020

Outline

- Immune system, T cells and T cell receptors
- Motivation and objectives
- TCR sequencing data
- Kernel methods
- Gaussian processes
- Results

Outline

- **Immune system, T cells and T cell receptors**
- Motivation and objectives
- TCR sequencing data
- Kernel methods
- Gaussian processes
- Results

Human immune system

- Humans are exposed to millions of potential pathogens daily, through contact, ingestion, and inhalation.

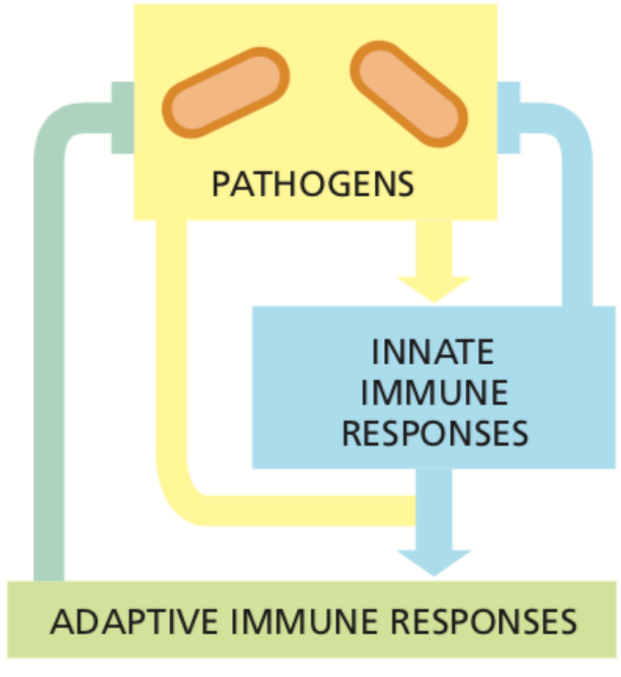


Figure: [1]

Innate Immune responses

- General defence reactions
- Three lines of defences:
  - Physical and chemical barriers
  - Cell-intrinsic responses
    - An individual cell recognizes that it has been infected and takes measures to kill or cripple the invader
  - A specialized set of proteins and phagocytic cells that recognize conserved features of pathogens and become quickly activated to help destroy invaders

Adaptive immune responses

- Highly specific responses
- Slow to develop on first exposure to a new pathogen (can take a week or so)
- Provide long-term protection
- Activated by innate immune system
- Carried out by lymphocytes
  - Antibody responses (B cells)
  - **T-cell-mediated responses**

T cells

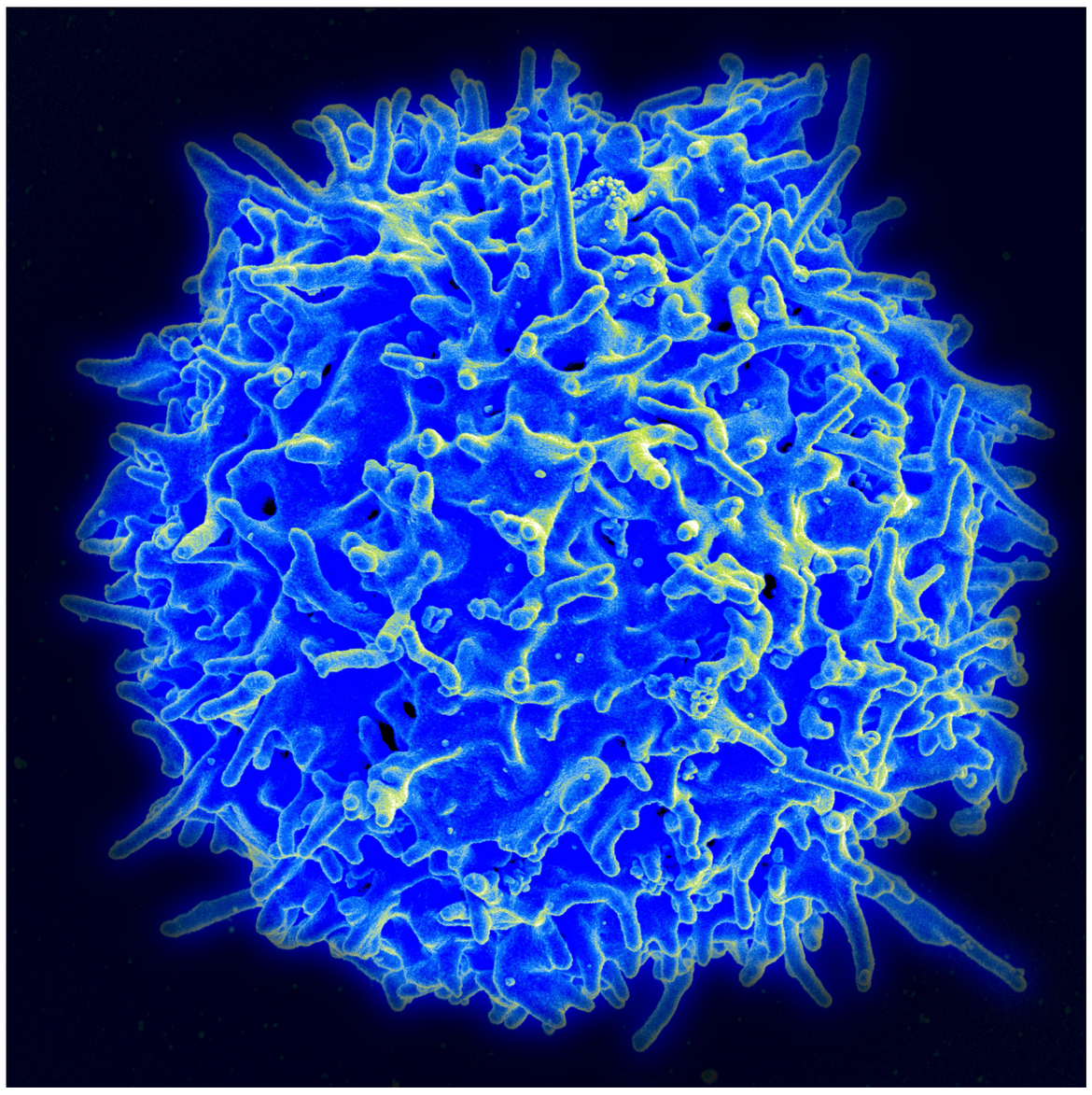


Figure: https://en.wikipedia.org/wiki/T_cell

- T cells are white blood cells (lymphocytes) that are distinguished from other lymphocytes by the presence of a T-cell receptor (TCR) on the cell surface
- T cells play a central role in the immune response

T cells and T cell receptors (TCRs)

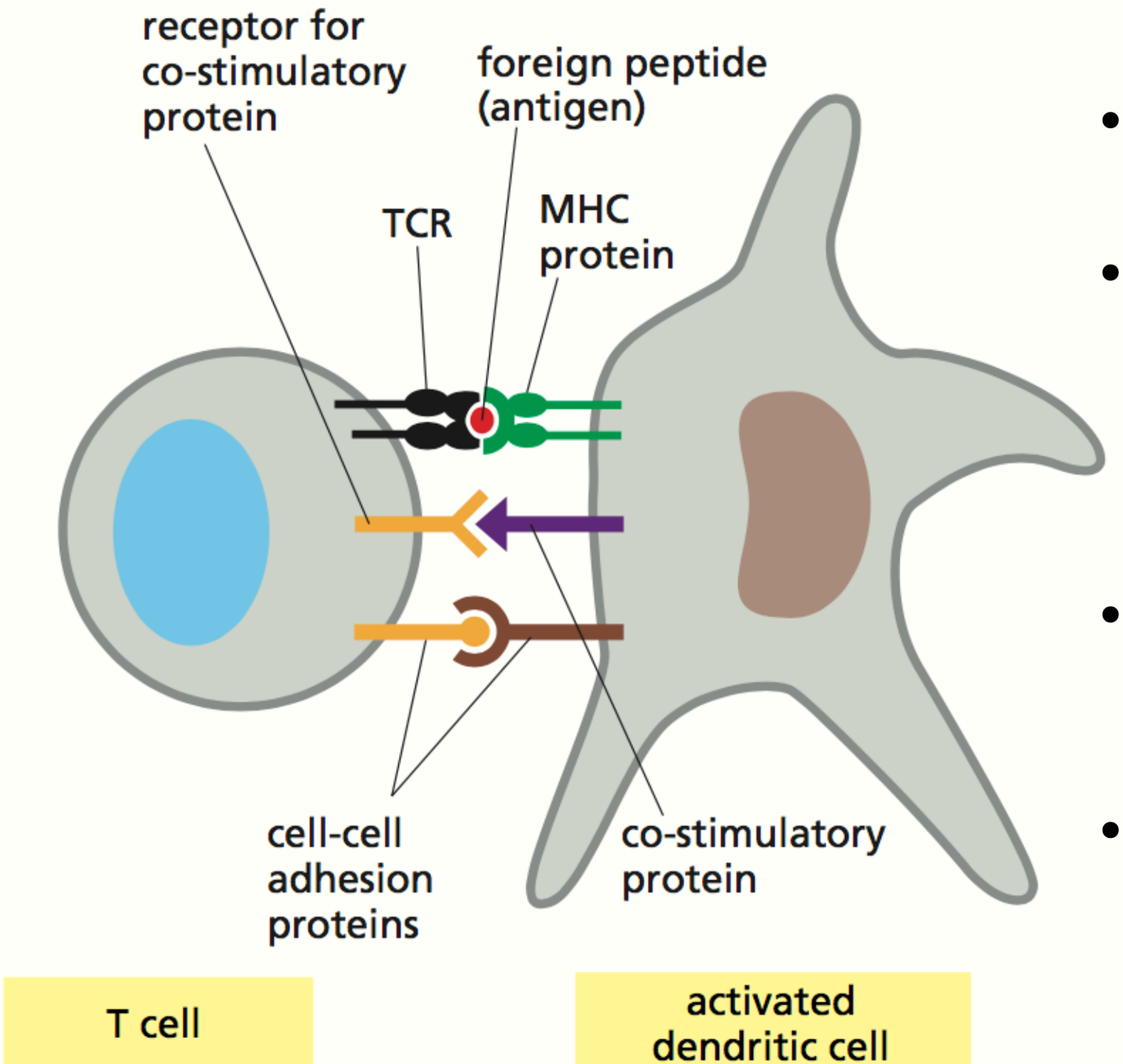


Figure: [1]

- T cells are activated by foreign antigens (peptides)
- Peptides are displayed by major histocompatibility complex (MHC) proteins located on the surface of antigen-presenting cells (usually dendritic cells)
- Peptide-MHC complex is recognised by T cell receptor (TCR)
- Upon T cell activation via TCR, T cells proliferate and differentiate into effector cells

T cell receptors (TCRs)

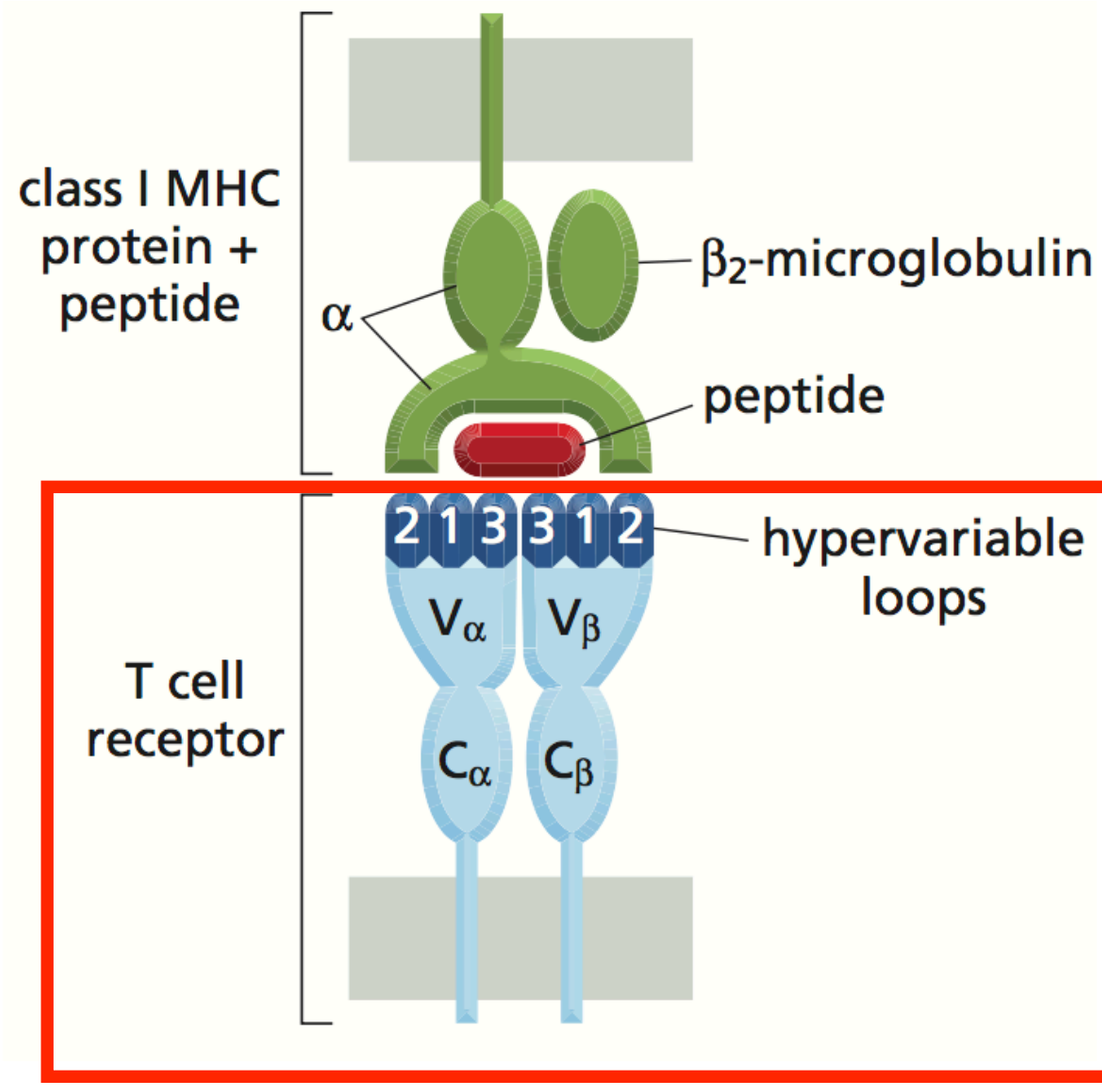


Figure: [1]

- The T-cell receptor (TCR) gene is expressed in T cells and found on the surface of T cells
- The TCR is a heterodimer composed of two different protein chains, alpha and beta
- Antigen (peptide) specificity is determined by hyper variable loops, so-called complementary determining regions (CDR) 1, 2 and 3

TCR diversity

- Each individual T cell can (in principle, but not in practice) have a unique TCR gene in DNA: different TCRs recognise different peptides
- **V(D)J recombination:** TCRs are manufactured from variable (V), diversity (D), joining (J) and constant (C) gene fragments through a process of somatic gene rearrangement

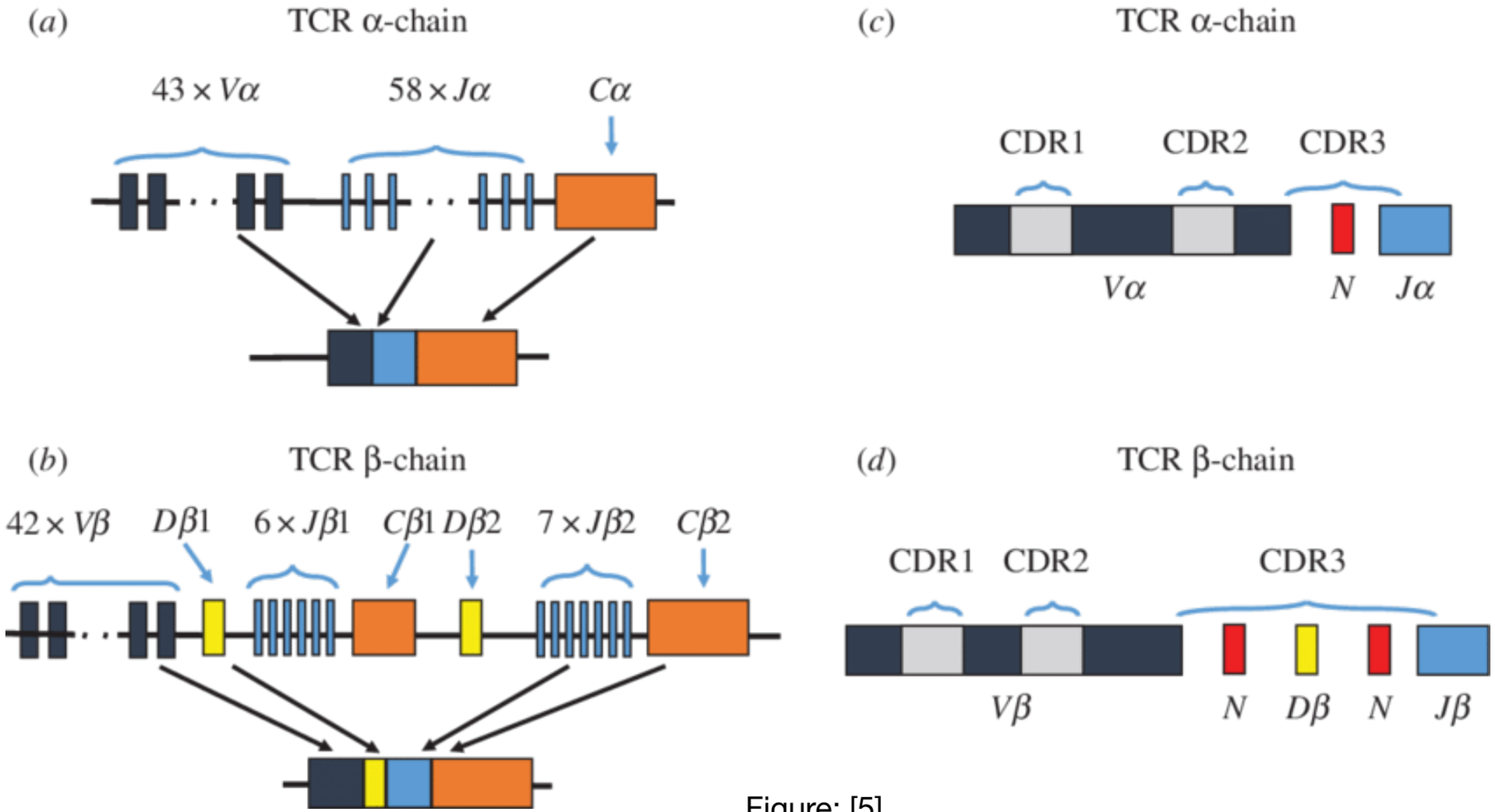
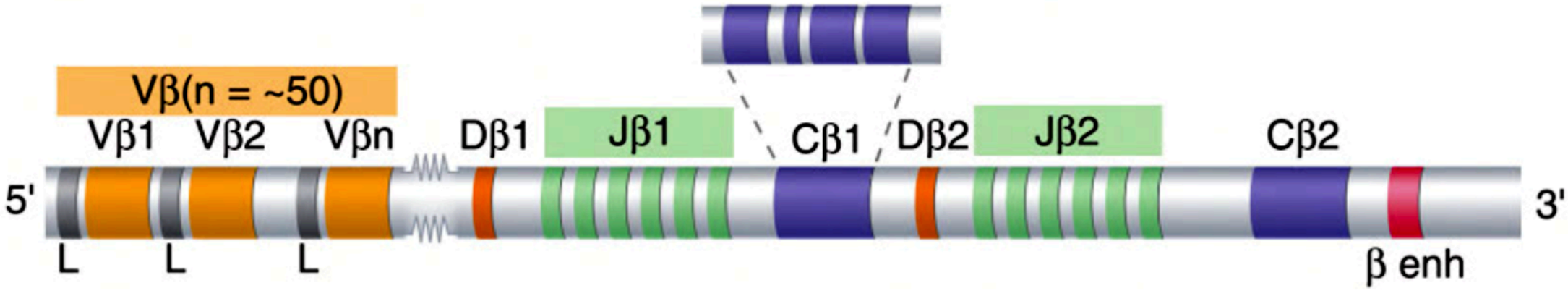


Figure: [5]

TCR diversity

- TCRα chain locus: 45 V-gene and 50 J-gene segments
- TCRβ chain locus: ~50 V-gene, 2 D-gene and 12 J-gene segments
- **Junctional diversification:** During the joining of these gene segments nucleotides can be lost from the ends of the segments, and one or more can also be inserted



Human TCRβ chain locus

Figure: Cellular and Molecular Immunology. Abul K. Abbas, Andrew H. H. Lichtman, Shiv

Antigen-binding site

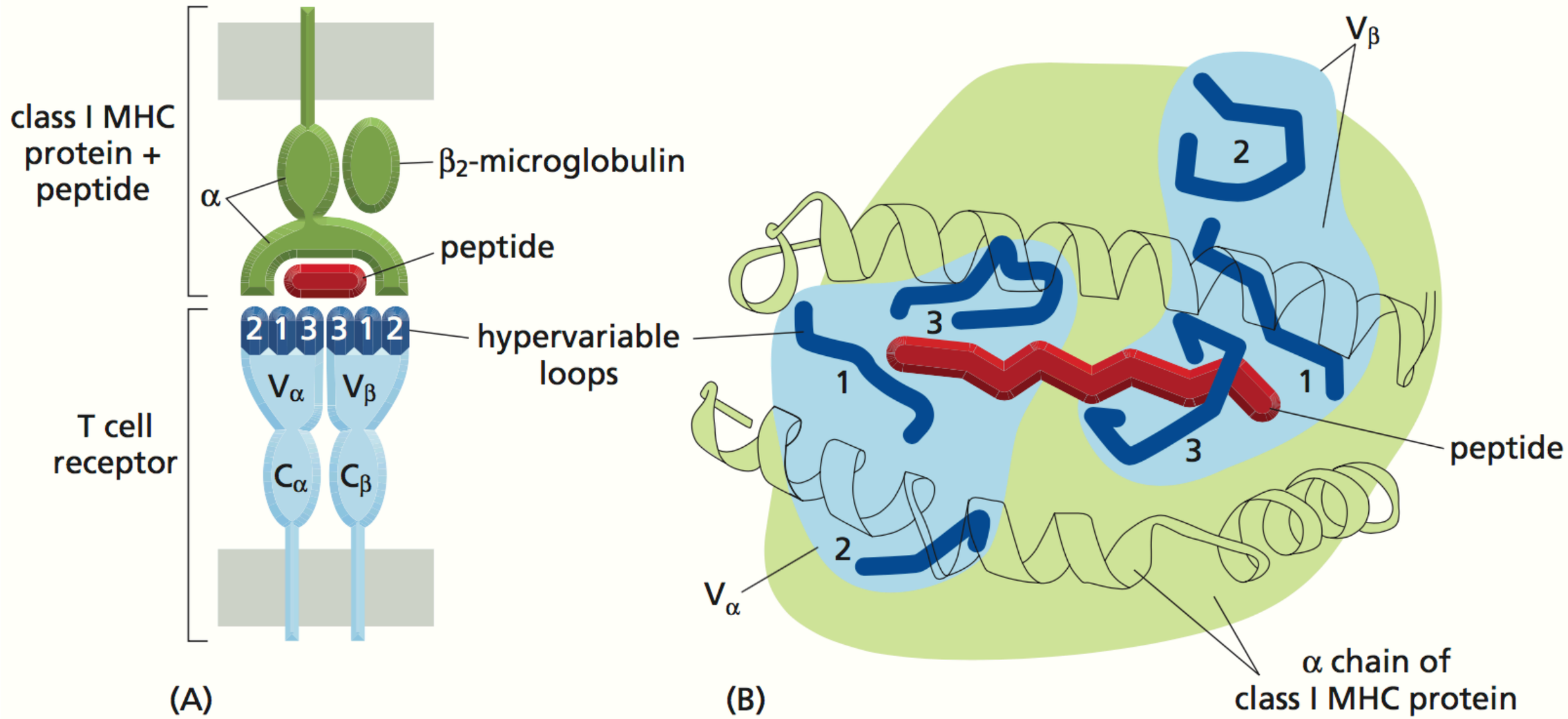
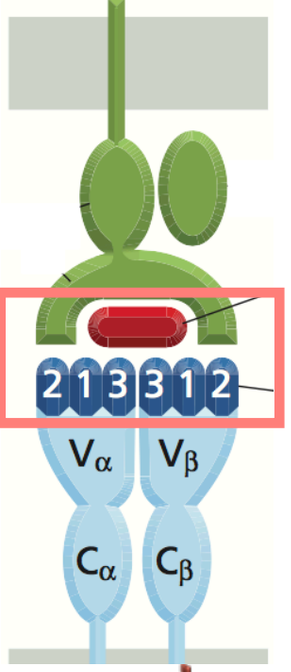


Figure: [1]

- CDR3 primarily interacts with the peptide and is most variable
- CDR1 and CDR2 (and CDR2.5) mainly bind to the walls of the peptide-binding groove, but have sometimes been observed to be in contact with the peptide

TCR + pMHC



CDR3α
CDR2.5α
CDR2α
CDR1α
CDR3β
CDR2.5β
CDR2β
CDR1β

peptide: NLVPMVATV (pp65)

Peptide binding groove of MHC

TCR repertoire

- Each T cell has potentially an unique TCR
- TCRs of an individual are called a TCR repertoire
- After a T cell has recognized an epitope, it starts to proliferate
    - The resulting set of T cells with identical TCRs is called a clone
        - T cells from large clones are more likely to be sampled
- TCR repertoire contains an immunological memory of all immunological stimuli an individual has had during lifetime
    - Viruses, microbes, other environmental exposures
    - Vaccines

Complexity of TCR repertoires

- ~10^{18} possible TCRs
  - ~10^{12} T cells in a human
    - ~10^{8} distinct TCRs in a human (young adult)
    - If a sample contains e.g. around 50 000 T cells
      - It’s about 0.000005 % of all T cells
- On average, each T cell recognises at least 1 million individual peptides
- A peptide can be recognised by several TCRs.

Outline

- Immune system, T cells and T cell receptors
- **Motivation and objectives**
- TCR sequencing data
- Kernel methods
- Gaussian processes
- Results

How to utilize TCRs?

Improved diagnostics

- Better understanding of an individual's immune status in different diseases

Personalized medicine

- Which patients would respond to different medications?

Repertoire level studies: utilize TCR repertoires of different subjects

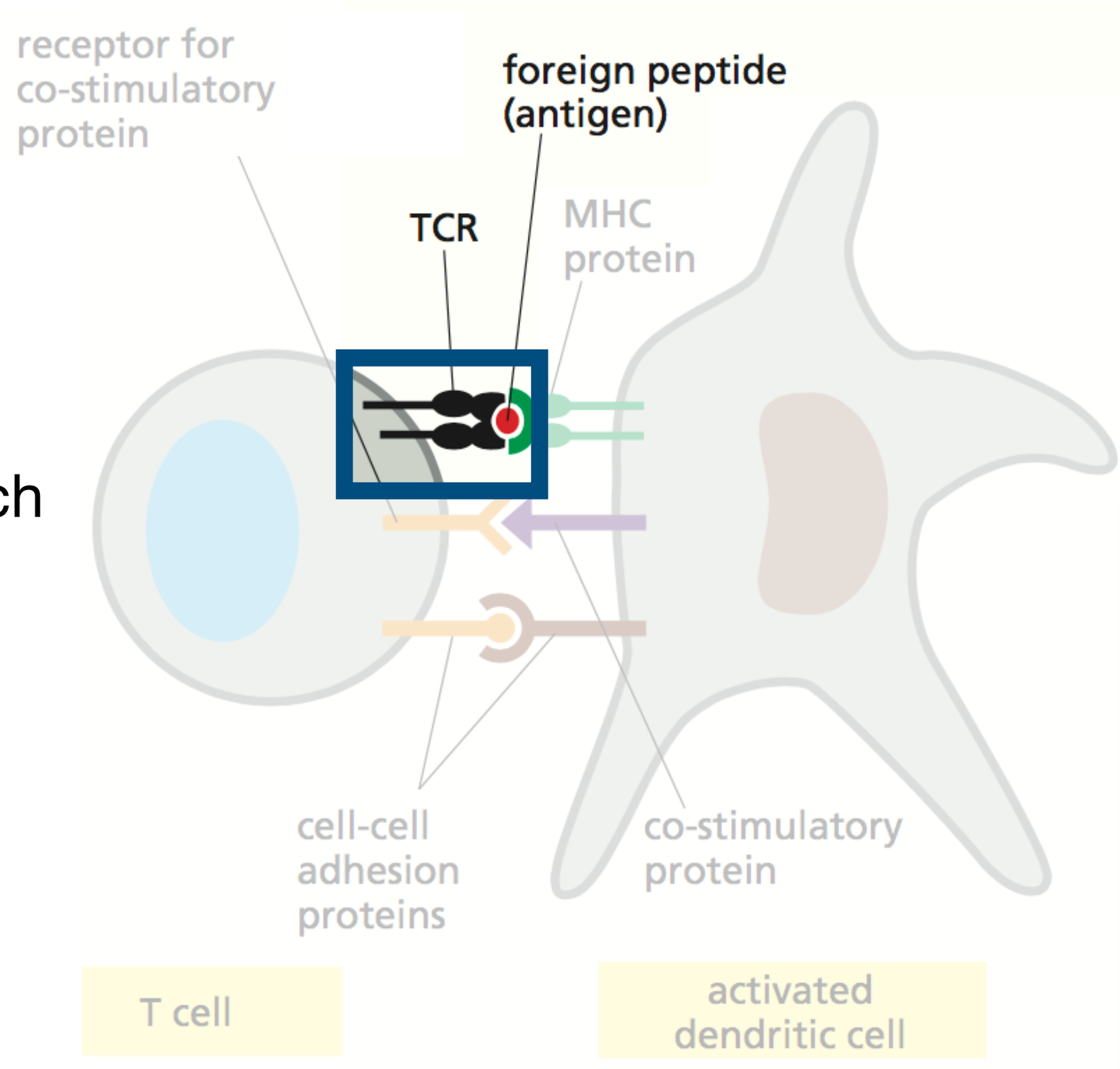
- E.g. find TCRs associated with some condition

Sequence level studies

- E.g. determine epitope specificity of individual TCRs

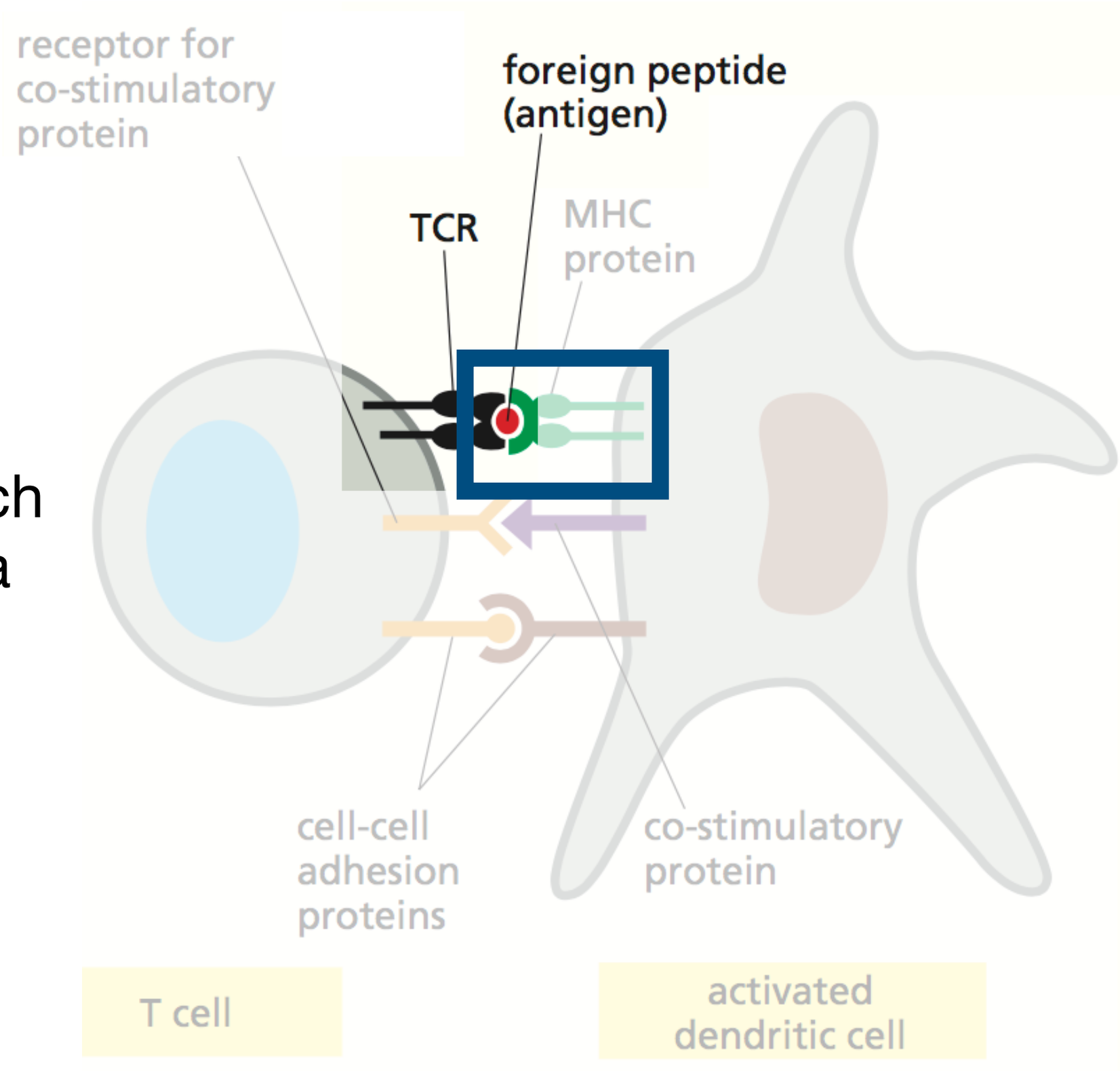
Goal

- Determine which peptides TCRs recognize



Another Goal

- Determine which peptides bind a given MHC

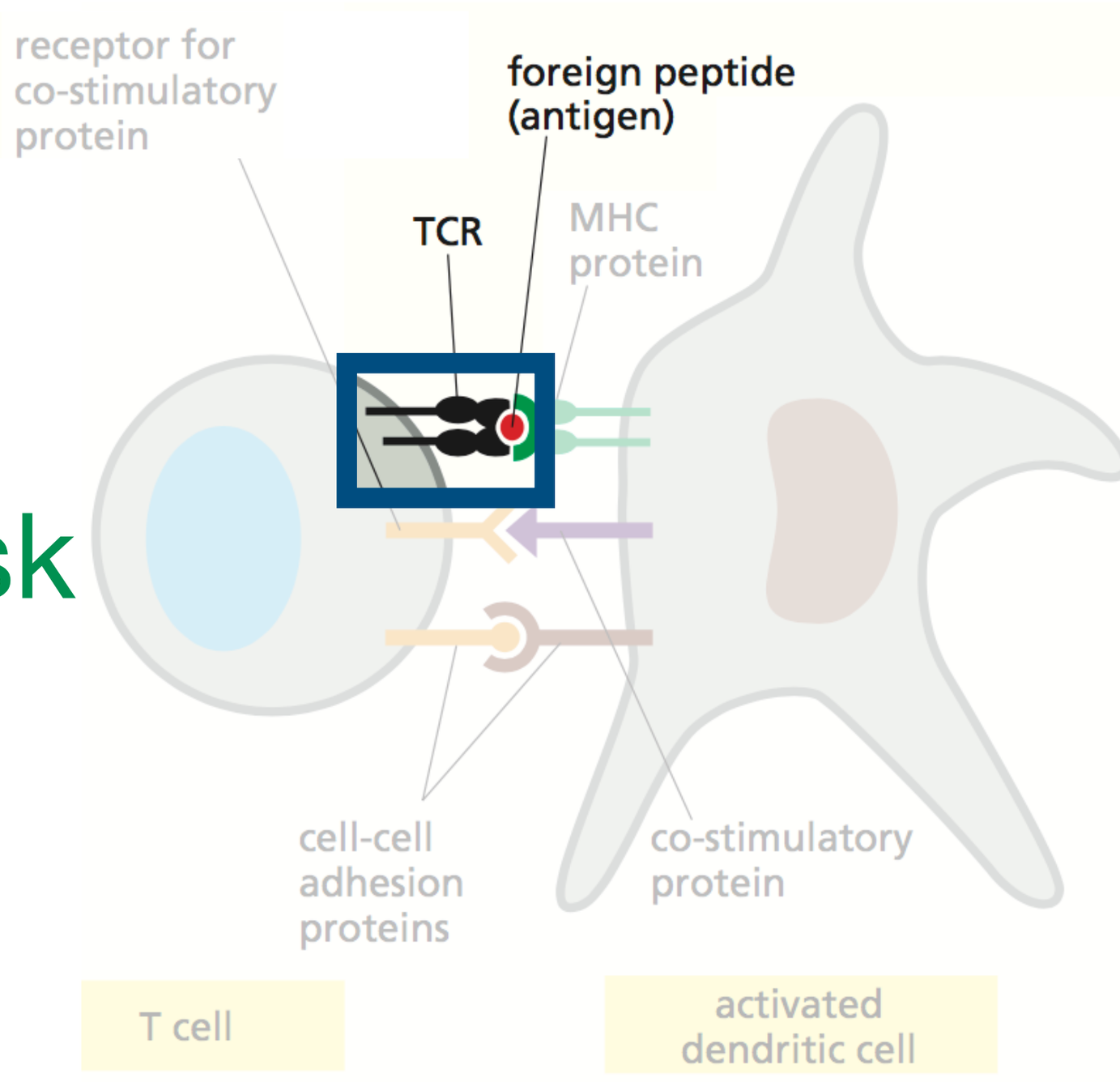


Why machine learning?

- **“Perfect” solution**:
  - Test experimentally which peptides all possible TCRs (~10^{18}) recognize
  - Impossible
- **Machine learning solution**:
  - Assume that similar TCRs behave similarly
  - Based on known specificities of some TCRs, predict specificities for new TCRs (supervised learning)

Supervised learning

- A learning process which looks at annotated data to then automatically annotate similar un-annotated data



Classification task

- Binary classification:
  - Predict whether a TCR recognizes and binds to a certain peptide or not

Outline

- Immune system, T cells and T cell receptors
- Motivation and objectives
- **TCR sequencing data**
- Kernel methods
- Gaussian processes
- Results

TCR sequencing

- TCRs can be quantified by sequencing
  - Targeted sequencing for TCR locus in DNA using C-gene selective primer (TCR-seq)
  - RNA-seq
- Additionally, one can first select T cells that recognize a specific peptide, and sequence the TCR gene from only those cells
  - Epitope-specific, tetramer-sorted TCR-seq

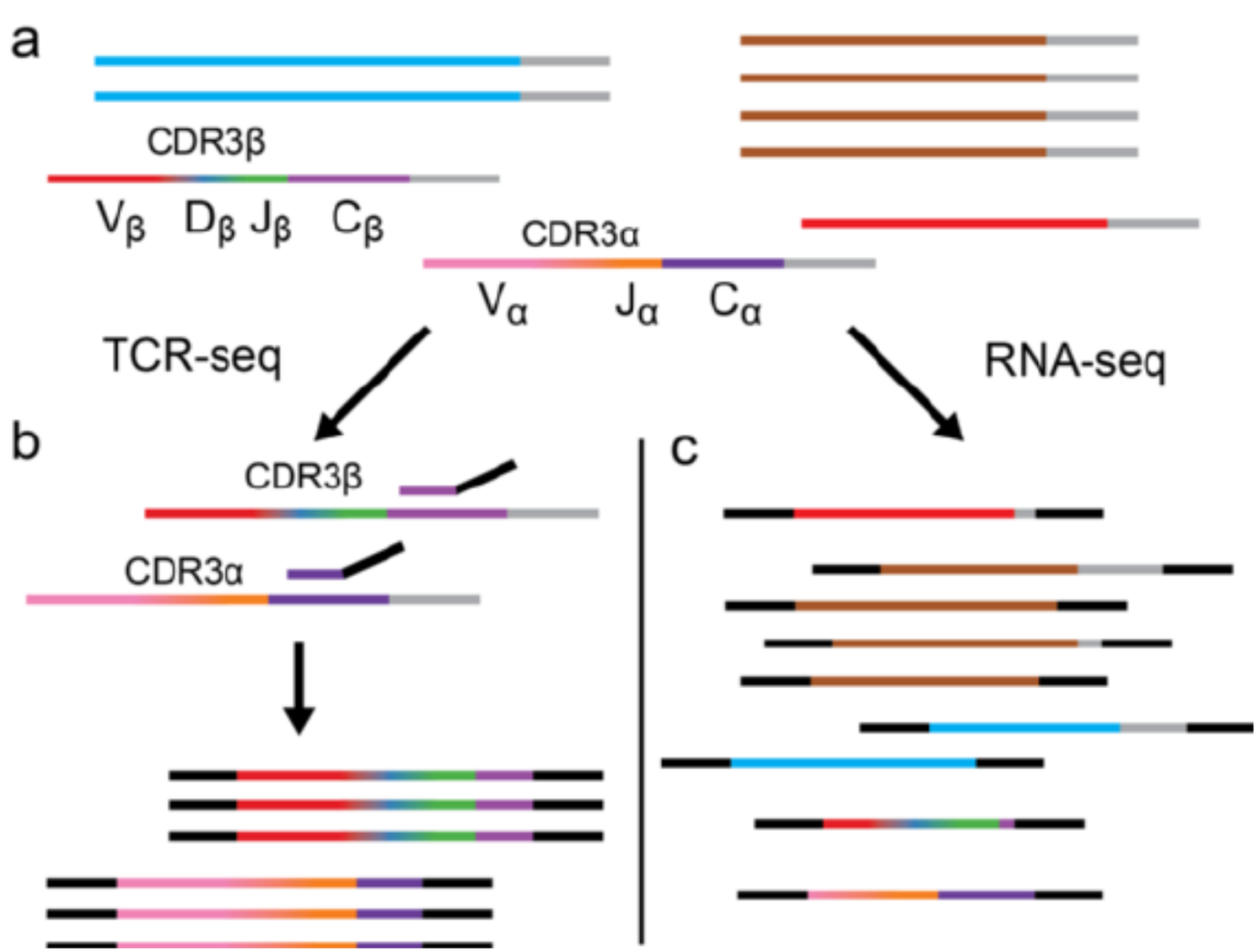


Figure: [6]

Quantification of TCRs from TCR-seq

- Align TCR-seq sequencing reads against V, D and J genes
- Similar to RNA-seq read alignment but with lots of mismatches and indels
- Several tools: e.g. MiXCR

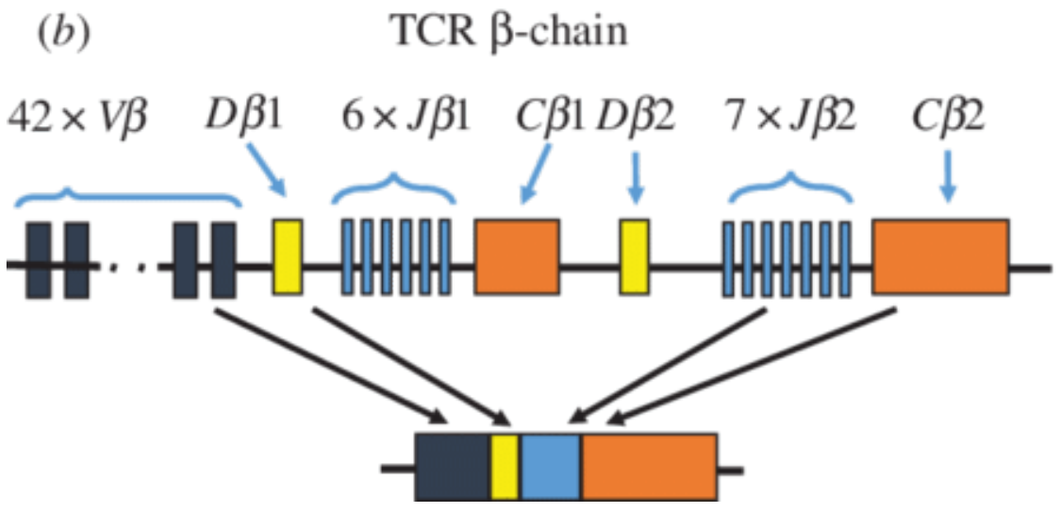


Figure: [5]

Quantification of TCRs from TCR-seq

- Align TCR-seq sequencing reads against V, D and J genes
- Similar to RNA-seq read alignment but with lots of mismatches and indels
- Several tools: e.g. MiXCR

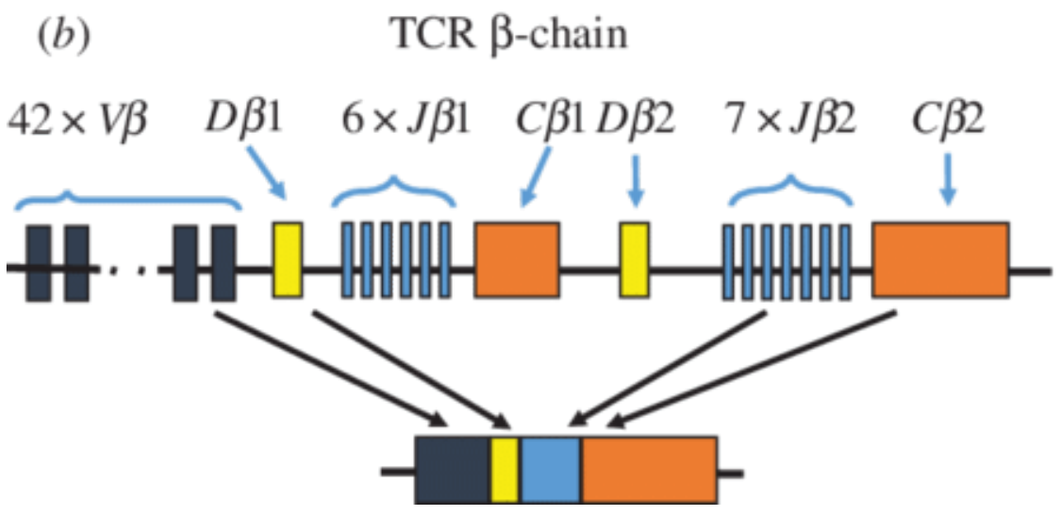


Figure: [5]

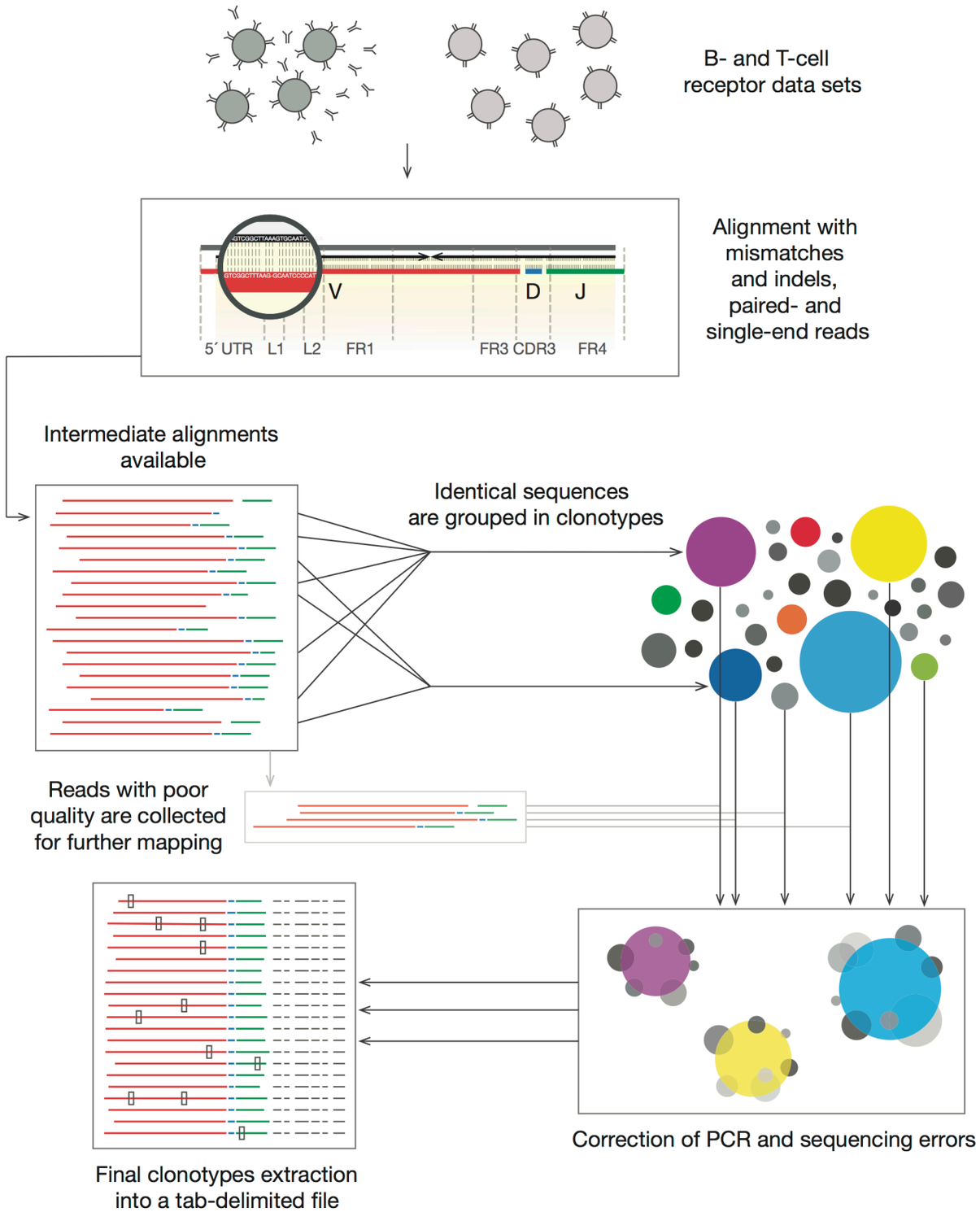


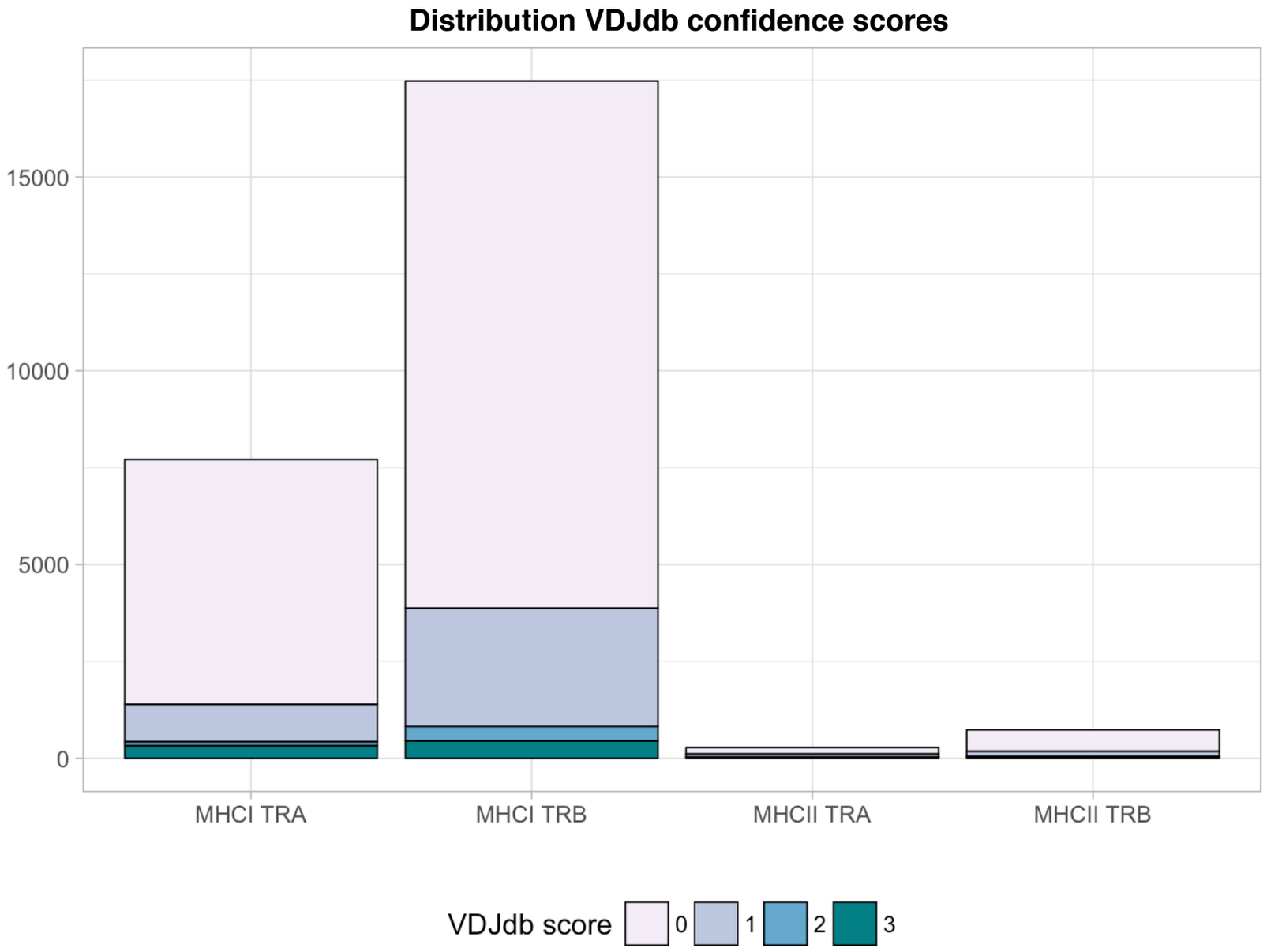
Figure: https://mixcr.readthedocs.io/en/master/

Epitope-specific TCRs

- Epitope-specific TCRs are stored e.g. in VDJdb

  https://vdjdb.cdr3.net

- TCRs recognizing epitopes from e.g.
  - Influenza A
  - Cytomegalovirus
  - HIV
  - Epstain Barr Virus
  - Sars-Cov-2



0 - critical information missing, 1 - medium confidence,
2 - high confidence, 3 - very high confidence.

Control sequences

- Negative controls may also be needed (e.g. for supervised analysis)

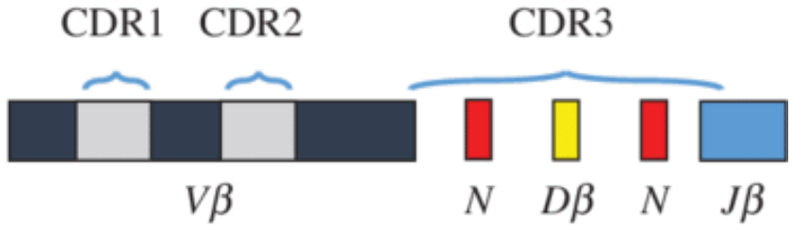
- Generally TCRs that recognize an epitope are sequenced, not TCRs that do not recognize that epitope

- We can take TCRs that appear only once (singletons) in a subject's TCR reportoire

- We can assume that these TCRs are unlikely to recognize a certain epitope

TCR amino acid sequences

- Usually a TCR is presented by its CDR3 amino acid sequence and V- and J-genes
- CDR1, CDR2 and CDR2.5 are completely determined by V-gene and allele



  - We can construct a table of CDR1, CDR2 and CDR2.5 sequences corresponding to all possible V-genes and alleles

- Examples of TCRβ sequences:

CDR3	CDR1	CDR2	CDR2.5
CASSIQALLTF	SGHDY	FNNNVP	PNASF
CASSVVGGNEQFF	SGDLS	YYNGEE	FPDLH
CASSVAQLAGGTDTQYF	SGDLS	YYNGEE	FPDLH
CSARDPSGLAGGLAETQYF	DFQATT	SNEGSKA	ASLTL

How to utilize sequences?

No alignment
CASSIQALLTF
CASSVVGGNEQFF
CASSVAQLAGGTDTQYF
CSARDPSGLAGGLAETQYF

With alignment
CASSIQ--------ALLTF
CASSVVG------GNEQFF
CASSVAQLA--GGTDTQYF
CSARDPSGLAGGLAETQYF

- Alignment free methods
  - \+ Sequences can have arbitrary lenghts
  - \- Cannot consider position specific information
- Methods that use aligned sequences
  - \+ Can utilize position specific information
  - \+ Can utilize amino acid features (more easily)
  - \- Good alignment can be difficult to get
  - \- New sequences need to be added to the alignment
  - \- New sequences cannot be longer than those in the original alignment

Alignment-free comparisons

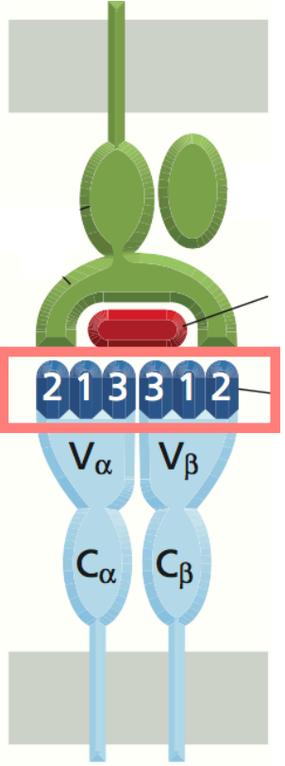
- Edit distance: Levenshtein distance
  - Minimum number of single amino acid changes (insertions, deletions, substitutions) between two sequences:
  - CASSLYF → C**A**AS**S**LY**F** → CAASLYW: distance is 3

insert delete substitute

- k-mer or motif frequencies
  - Define a set of k-mers,

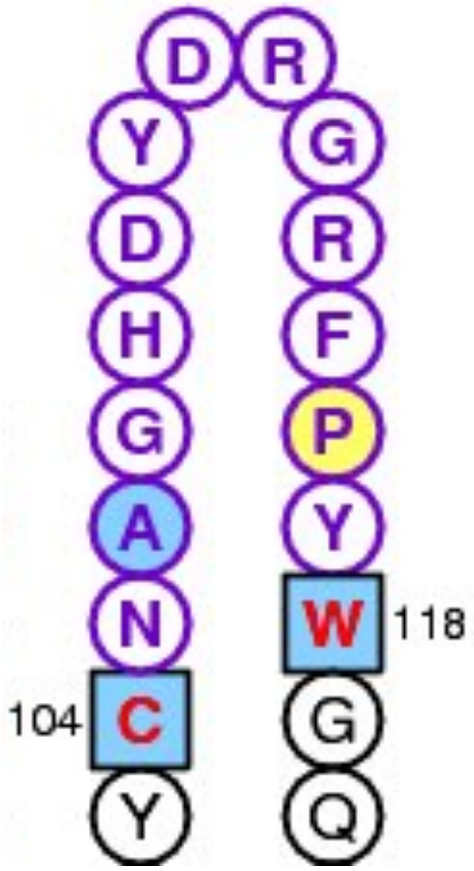
    all possible or some smaller set
- Can be used to define “similar” TCRs
- Do not consider similarity between amino acids

	CAS	ASS	SSL	SLY	⋯
CASSLYFF	1	1	1	1	⋯
CASSIQALLTF	1	1	0	0	⋯
CASSVVGGNEQFF	1	1	0	0	⋯
CAVGDRGYEQYF	0	0	0	0	⋯
⋮	⋮	⋮	⋮	⋮	⋱



Aligning TCR sequences

- There is a limited number of CDR1, CDR2 and CDR2.5 sequences, and we know what they are
  - They can all be aligned according to IMGT definitions
- We assume that CDR3 sequences form simple loops
  - We add gap at the top of the loop for shorter sequences (according to IMGT numbering)
  - Easy to add new sequences to the alignment
- Examples of aligned TCRβ sequences



CDR3	CDR1	CDR2	CDR2.5
CASSIQ--------ALLTF	SGH--------DY	FNN----NVP	P-NASF
CASSVVG------GNEQFF	SGD--------LS	YYN----GEE	F-PDLH
CASSVAQLA--GGTDTQYF	SGD--------LS	YYN----GEE	F-PDLH
CSARDPSGLAGGLAETQYF	DFQ------ATT	SNEG---SKA	A-SLTL

One-hot encoding

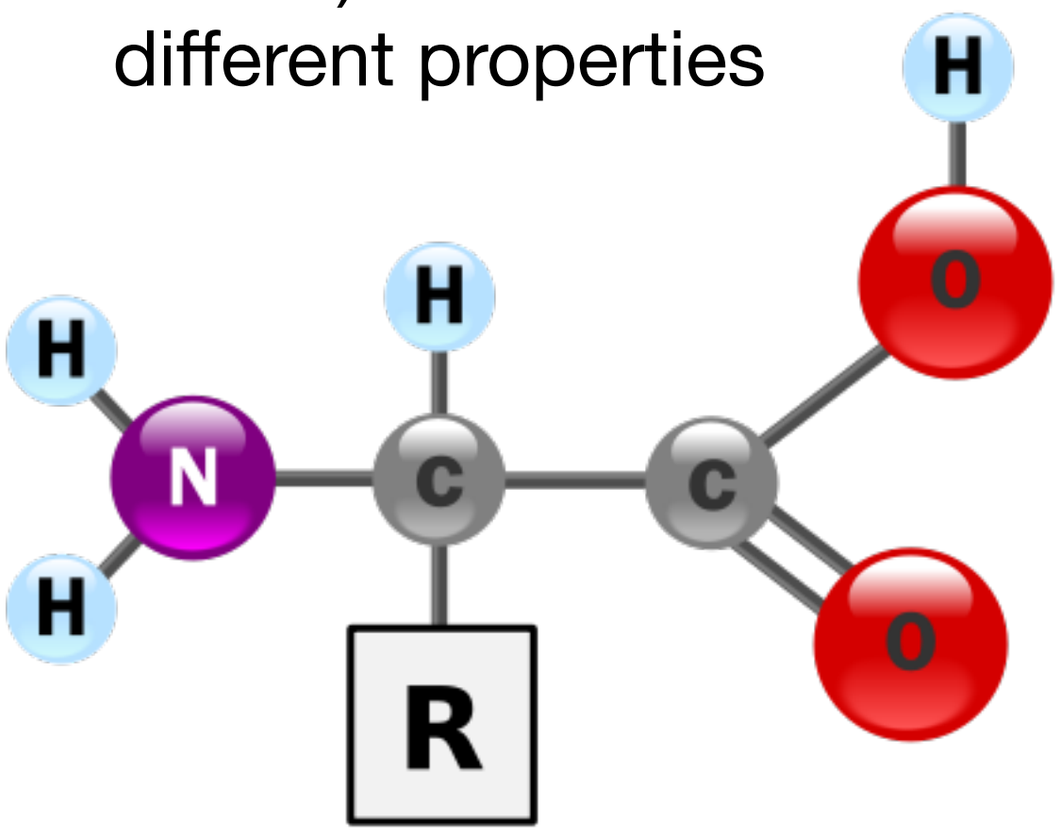
- Most simple numeric presentation
  - Sequences as vectors with constant length
  - Does not consider similarity between amino acids

A	R	N	D	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	-
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

C	A	S	S	-	-	-	L	Y	F	F
0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	1
0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	1	0	0	0	0

Amino acid properties

- There are 20 naturally occurring amino acids
- R-groups (or side chains) determine their different properties

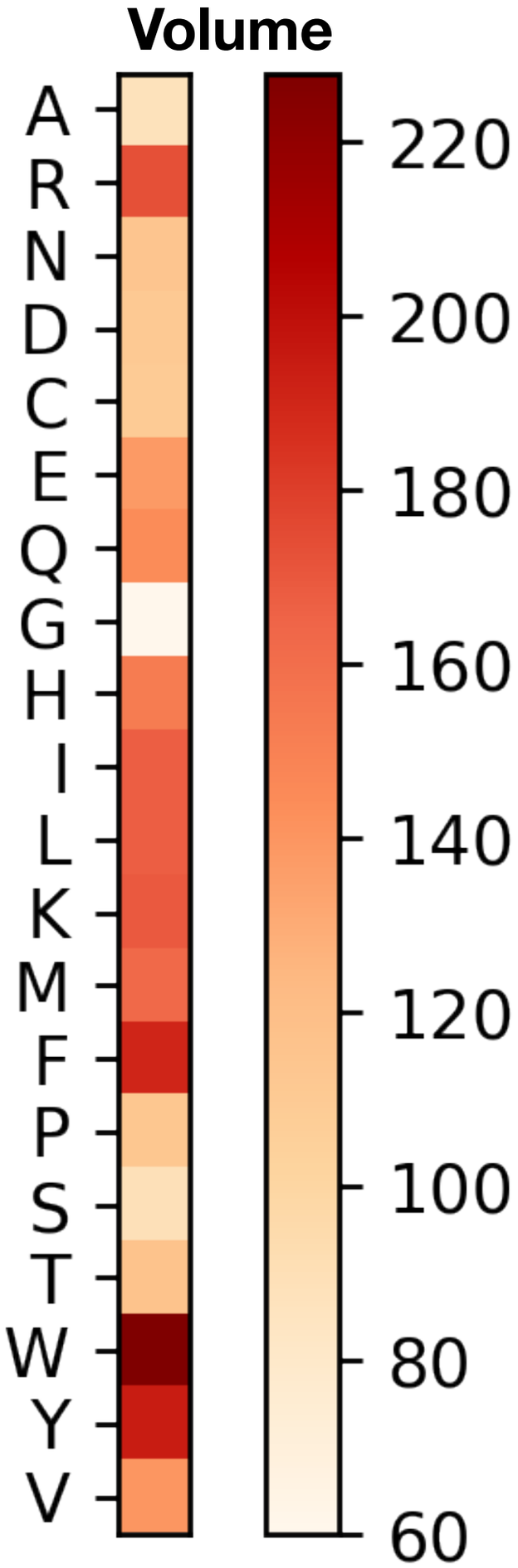


Amino acid	Abbreviation		Chemical	Volume	Hydropathy
Alanine	Ala	A	alipahatic	87	hydrophobic
Arginine	Arg	R	basic	173	hydrophilic
Aspargine	Asn	N	amide	114	hydrophilic
Aspartic acid	Asp	D	acid	111	hydrophilic
Cysteine	Cys	C	sulfur	109	hydrophobic
Glutamic acid	Glu	E	acid	138	hydrophilic
Glutamine	Gln	Q	amide	144	hydrophilic
Glysine	Gly	G	aliphatic	60	neutral
Histidine	His	H	basic	153	neutral
Isoleucine	Ile	I	alipahatic	167	hydrophobic
Leucine	Leu	L	alipahatic	167	hydrophobic
Lycine	Lys	K	basic	169	hydrophilic
Methionine	Met	M	sulfur	163	hydrophobic
Phenyalanine	Phe	F	aromatic	190	hydrophobic
Proline	Pro	P	Cyclic	113	neutral
Serine	Ser	S	hydroxyl	89	neutral
Threonine	Thr	T	hydroxyl	116	neutral
Tryptohophan	Trp	W	aromatic	228	hydrophobic
Tyrosine	Tyr	Y	aromatic	194	neutral
Valine	Val	V	alipahatic	140	hydrophobic

Feature presentation

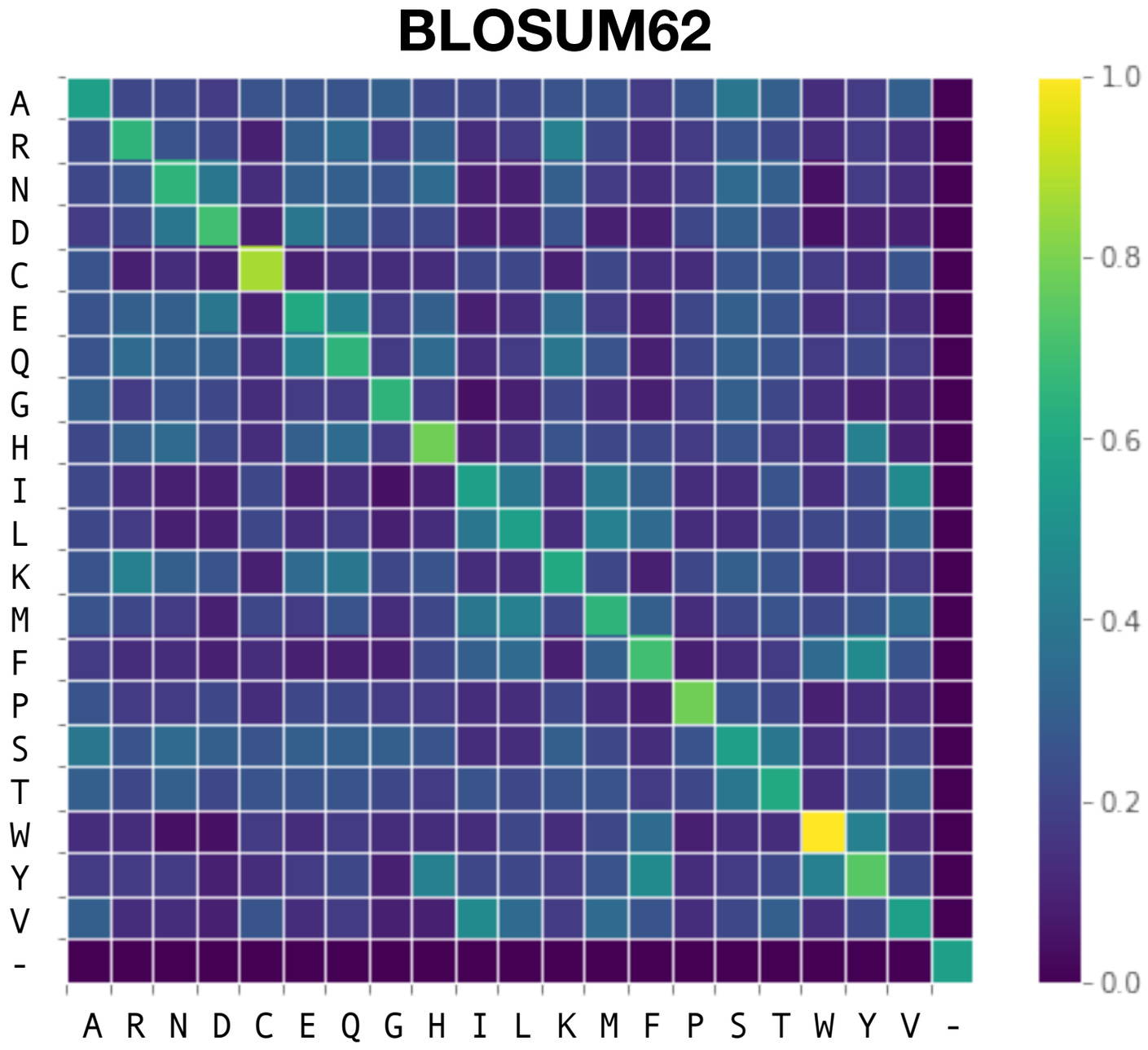
- Use the different amino acid properties as features
- Concatenate them to make feature vectors for each amino acid, e.g.

$$\begin{bmatrix} \text{volume} \\ \text{charge} \\ \text{hydrophobicity} \\ \text{polarity} \end{bmatrix}$$



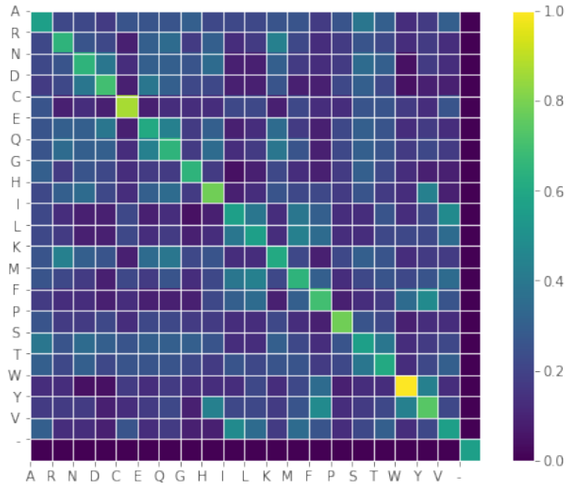
Substitution matrices

- Describe how easily an amino acid can be substitued with another
- Can be based e.g. on:
  - Sequence comparison
  - Sequence comparison by protein blocks
  - Chemical similarity
  - Structural or physical similarity



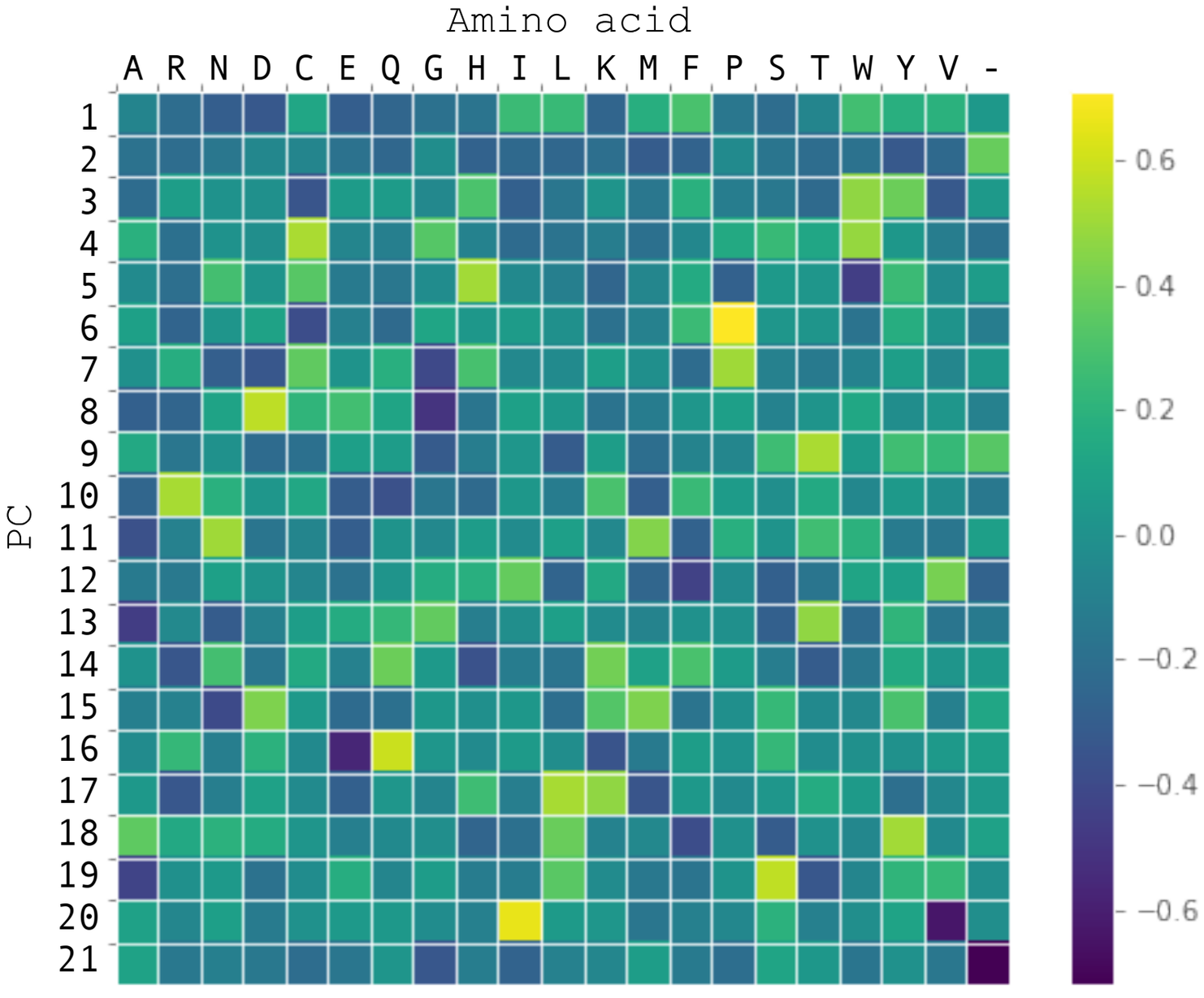
With added gap (-) and scaled into range [0,1]

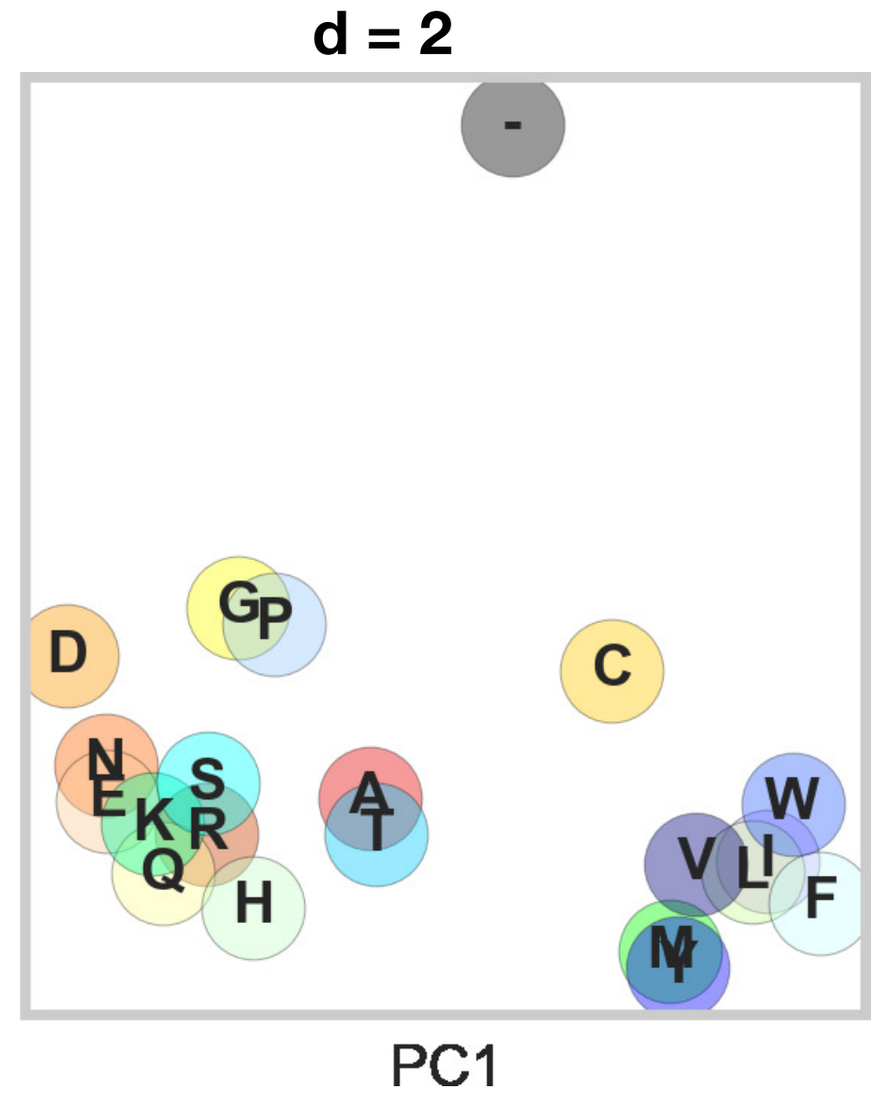
Amino acid features with BLOSUM62

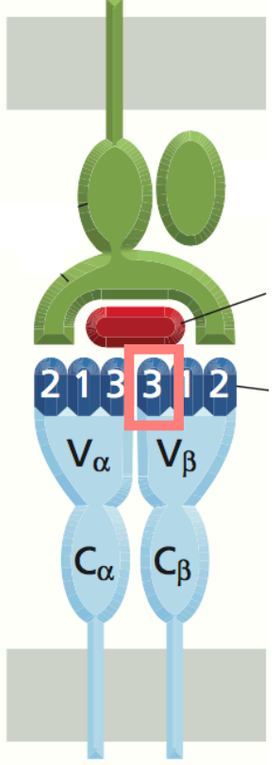


PCA of BLOSUM62

→ feature vectors (size: *d*×1*)* for each amino acid

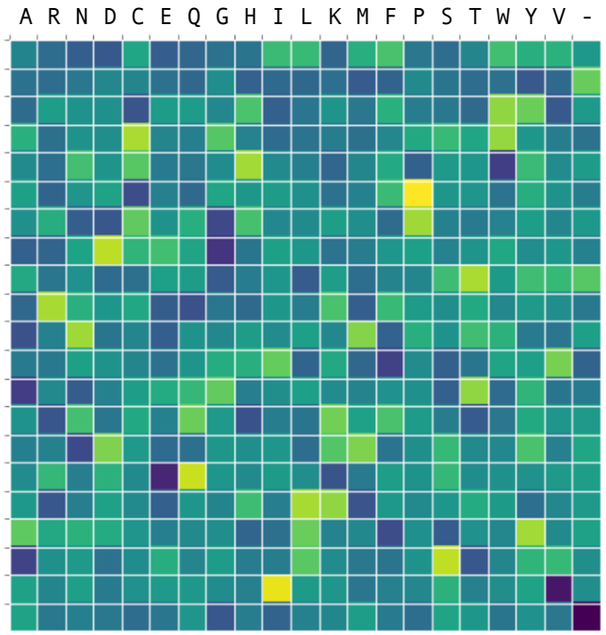




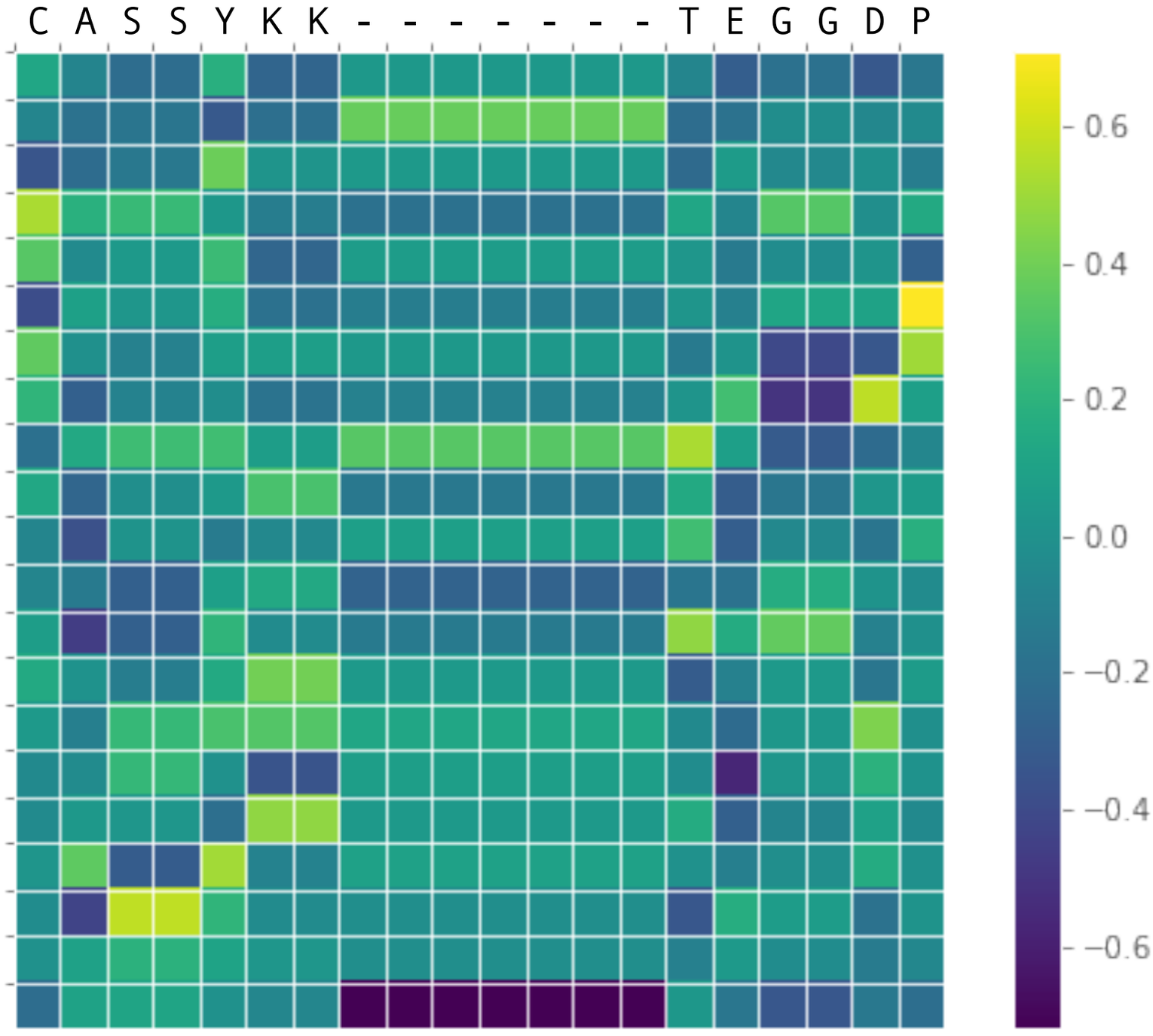


CDR3 presentation with BLOSUM62

PCA of BLOSUM62
→ feature vectors (size: *d*×1)
for each amino acid



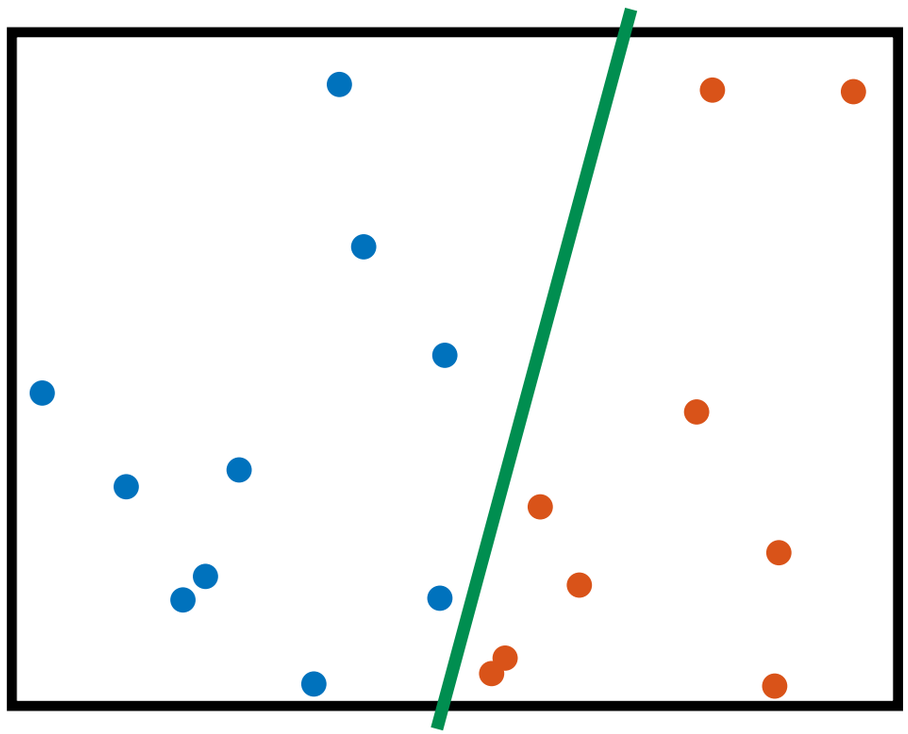
Sequence presentation (size: *l*×*d* or (*l*·*d*)×1)



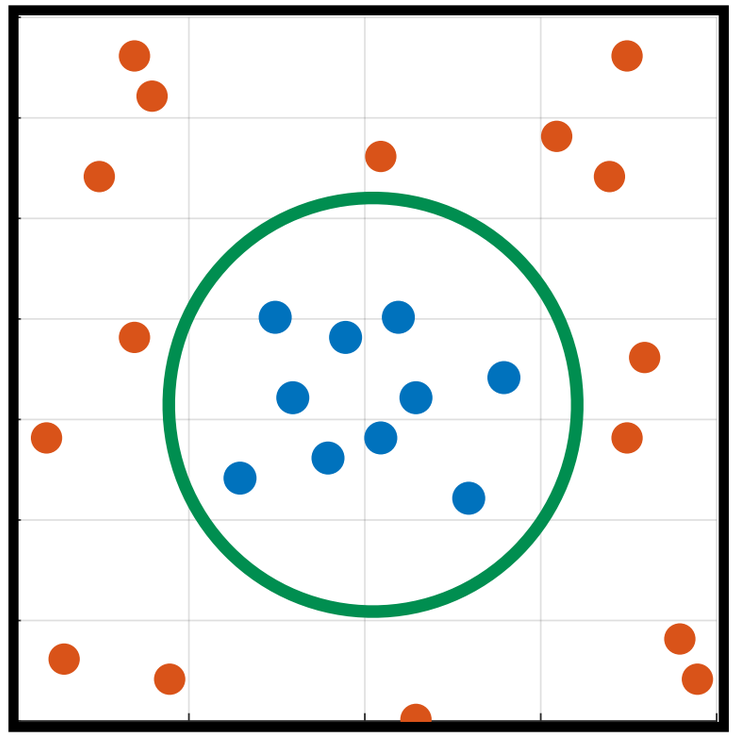
Outline

- Immune system, T cells and T cell receptors
- Motivation and objectives
- TCR sequencing data
- **Kernel methods**
- Gaussian processes
- Results

Classification



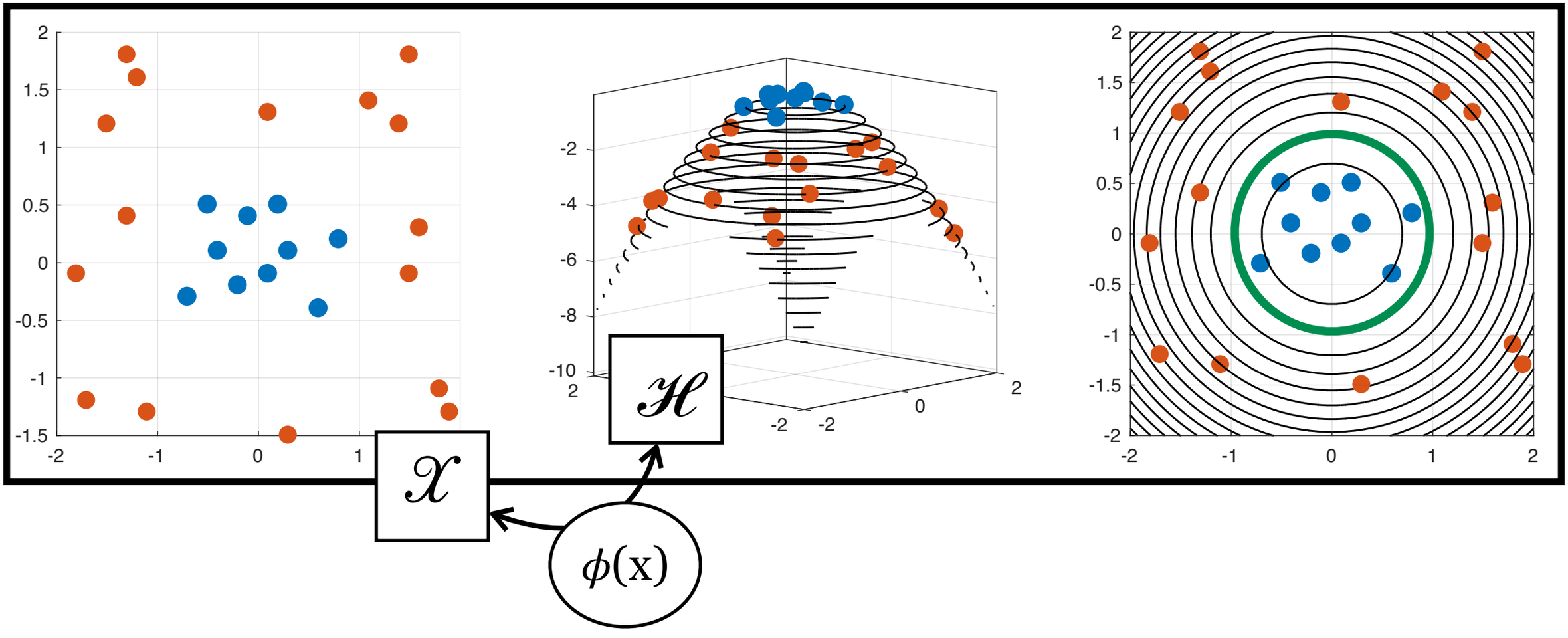
- Linear classification
  - Fairly simple



- Non-linear classification
  - More difficult
  - Can be implemented with kernels

Kernels (1/3)

- Kernel functions allow us to encode the similarity of TCRs
- Kernels can map data x ∈ $\mathcal{X}$ to a higher dimensional space $\mathcal{H}$, where it is linearly separable



Kernels (2/3)

- Definition:

For a non-empty set $\mathcal{X}$, a function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a kernel if there exists a Hilbert space $\mathcal{H}$ and a function $\phi : \mathcal{X} \to \mathcal{H}$ such that $\forall x, x' \in \mathcal{X}, k(x, x') := \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$

- A commonly used kernel is Gaussian kernel (or radial basis function (RBF) or squared exponential (SE)):

$$k(\mathbf{x}, \mathbf{x}'|\theta) = \sigma^2 \exp\left(-\frac{(\mathbf{x} - \mathbf{x}')^T(\mathbf{x} - \mathbf{x}')}{2\ell^2}\right),$$

where ℓ is the length-scale parameter, σ^2 is the magnitude parameter and $\theta = (\ell, \sigma^2)$.

Kernels (3/3)

- Examples of kernel functions:

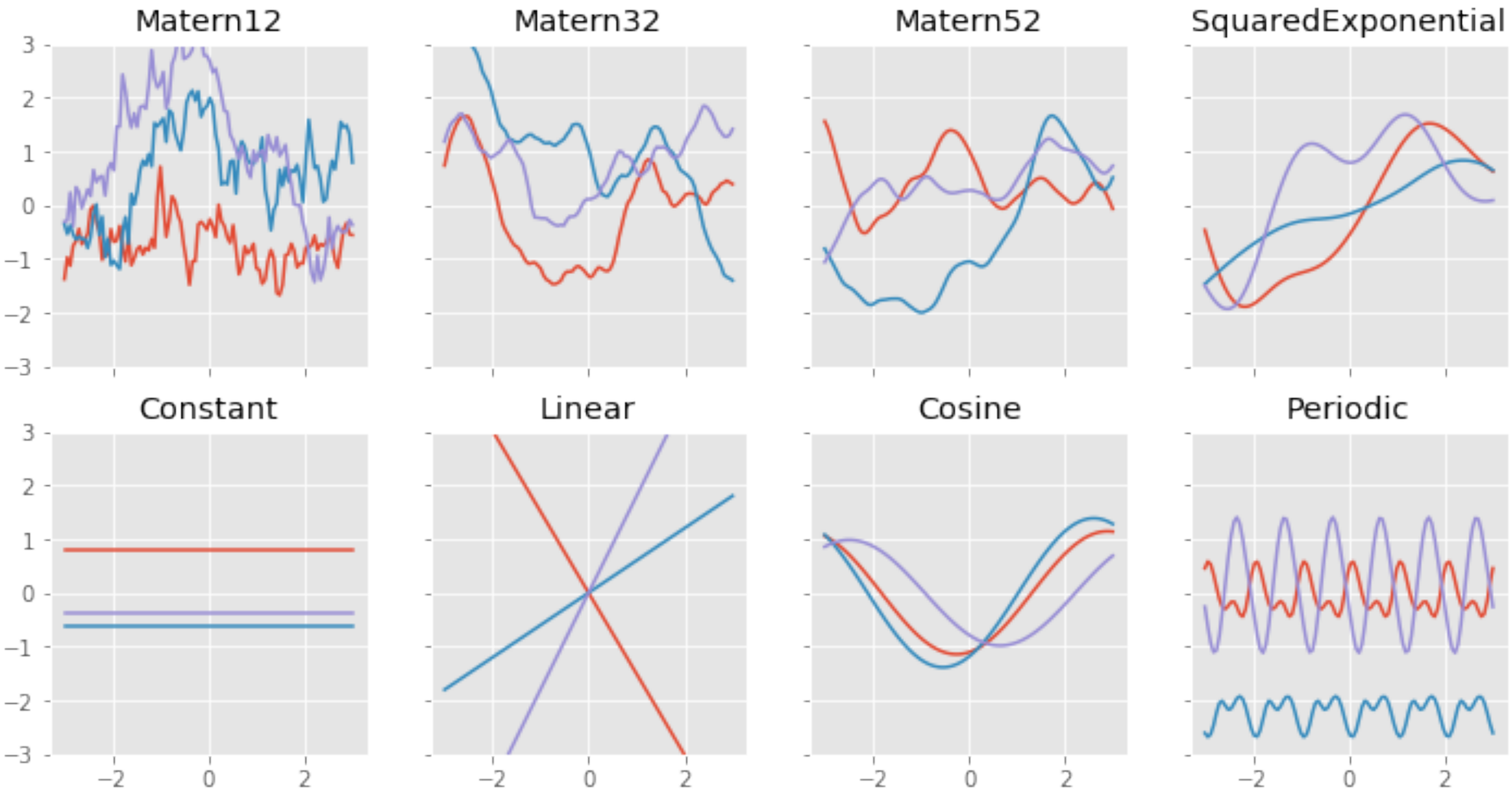


Figure: https://gpflow.readthedocs.io/en/develop/notebooks/kernels.html

Outline

- Immune system, T cells and T cell receptors
- Motivation and objectives
- TCR sequencing data
- Kernel methods
- **Gaussian processes**
- Results

GP classification

- A probabilistic classifier that uses kernels
- Can learn non-linear decision boundaries
- Learns suitable complexity of the boundary from data
- Models the confidence of the predictions

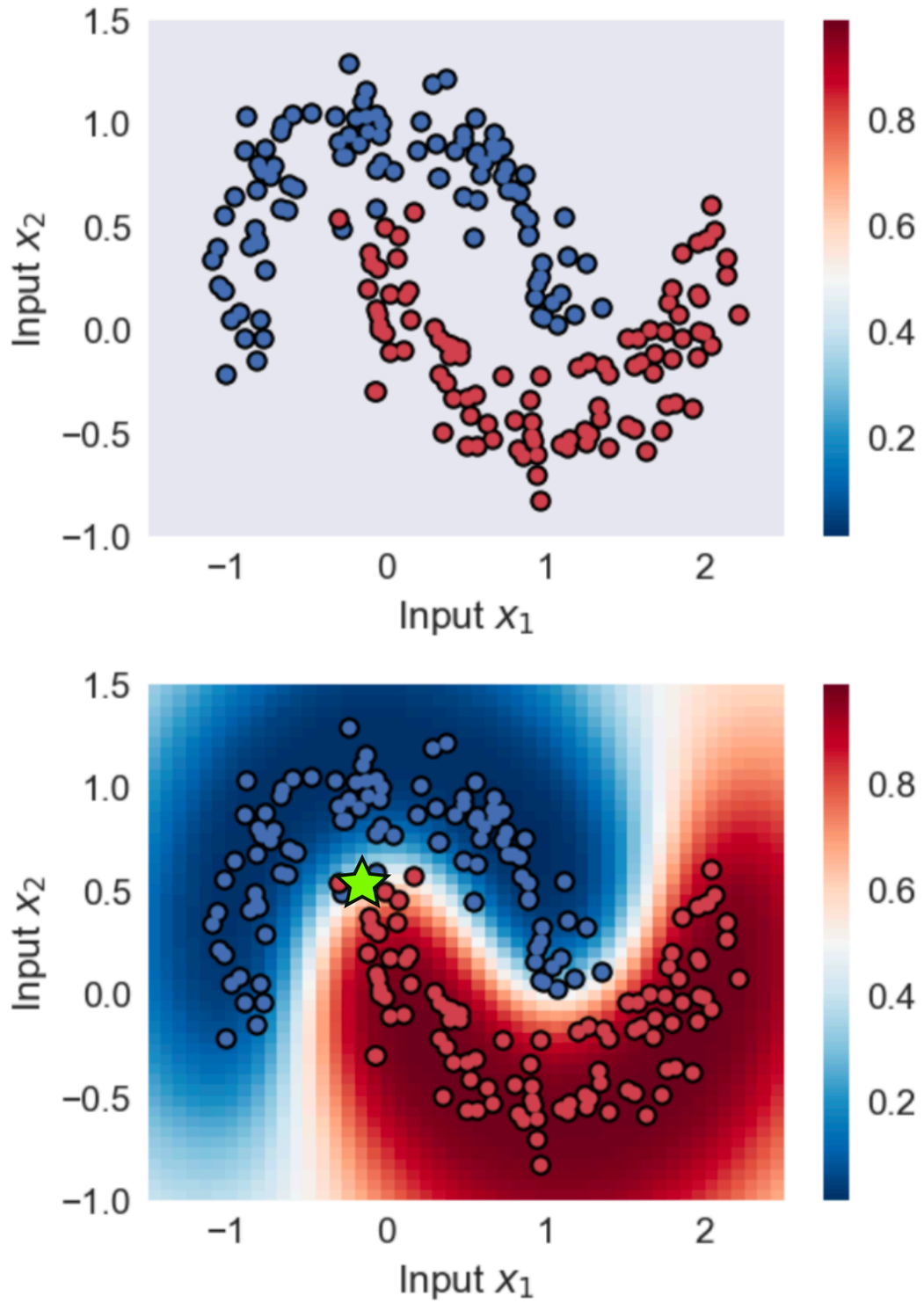
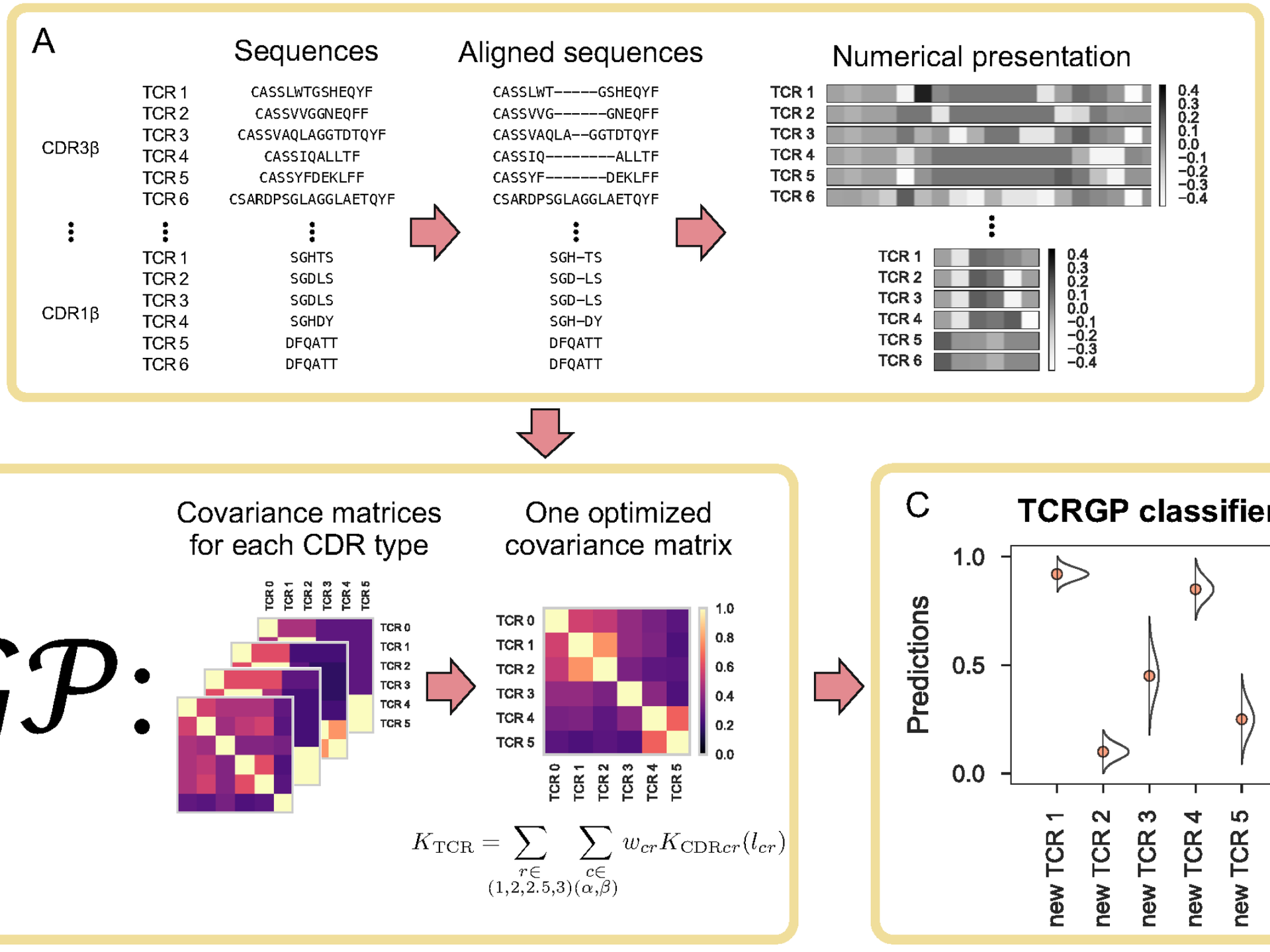


Figure: Michael Riis Andersen, Special course on Gaussian processes, Session 4, 2018

TCRGP pipeline



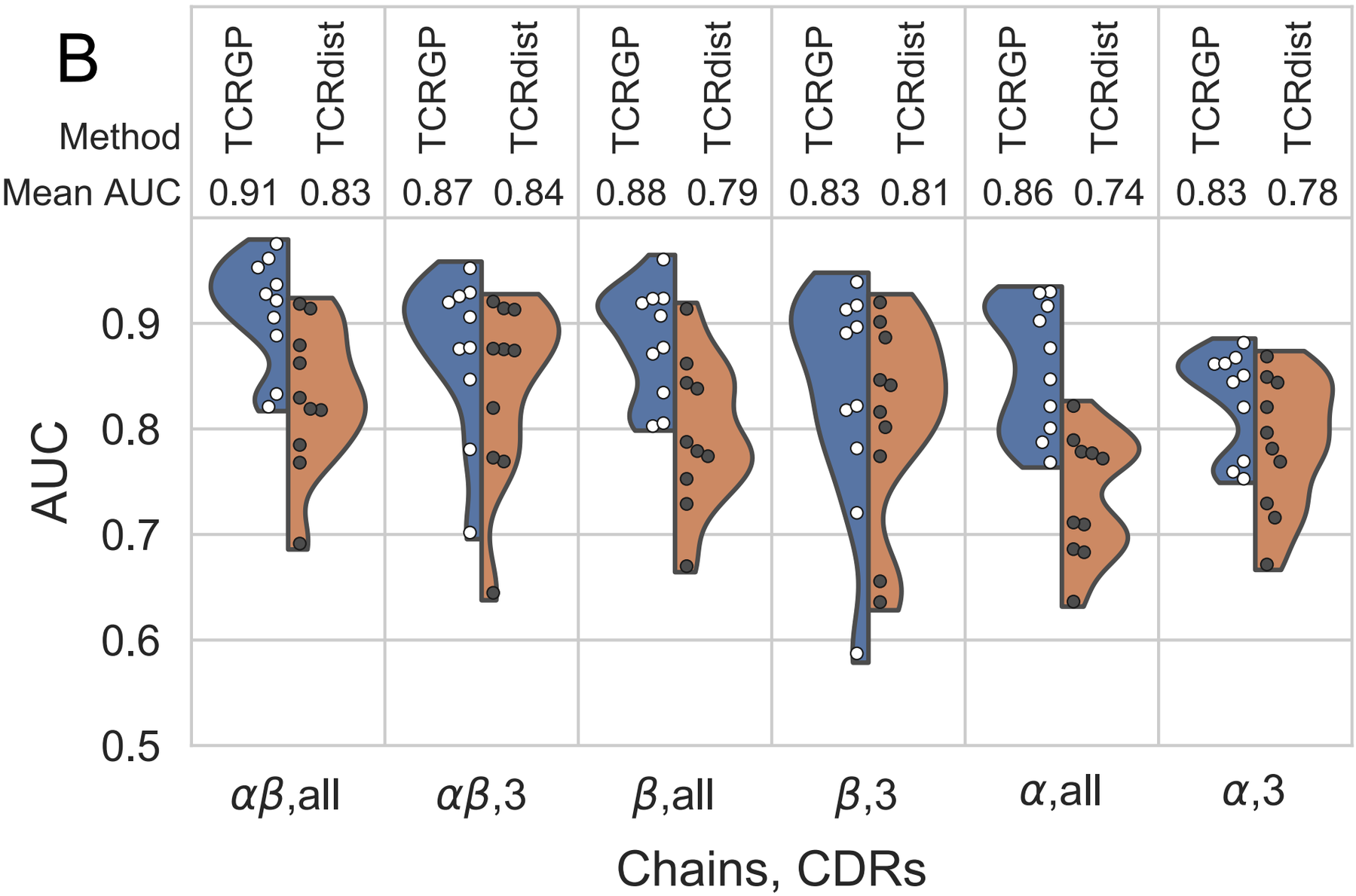
Epitope-specific TCR data

Species	Epitope species	Epitope gene	Epitope	MHC chain 1	MHC chain 2	Subjects	Samples	Unique TCR$\alpha\beta$s
Dash data								
Human	EBV	BMLF1$_{280\text{-}288}$	GLCTLVAML	HLA-A*02:01	-	6	76	69
	CMV	pp65$_{495\text{-}503}$	NLVPMVATV	HLA-A*02:01	-	10	61	60
	IAV	M1$_{58\text{-}66}$	GILGFVFTL	HLA-A*02:01	-	15	275	237
Mouse	IAV	PB1-F2$_{62\text{-}70}$	LSLRNPILV	D^b	-	9	117	117
	IAV	NP$_{366\text{-}374}$	ASNENMETM	D^b	-	24	305	263
	IAV	PA$_{224\text{-}233}$	SSLENFRAYV	D^b	-	15	324	293
	IAV	PB1$_{703\text{-}711}$	SSYRRPVGI	K^b	-	34	642	584
	mCMV	m139$_{419\text{-}426}$	TVYGFCLL	K^b	-	8	87	87
	mCMV	M38$_{316\text{-}323}$	SSPPMFRV	K^b	-	14	158	143
	mCMV	M45$_{985\text{-}993}$	HGIRNASFI	D^b	-	13	291	271
VDJdb data								
Human	CMV	pp65$_{123\text{-}131}$	IPSINVHHY	HLA-B*35	B2M	17	65	58
	CMV	pp65$_{417\text{-}426}$	TPRVTGGGAM	HLA-B*07	B2M	29	184	122
	CMV	pp65$_{495\text{-}503}$	NLVPMVATV	HLA-A*02	B2M	103	413	242
	EBV	BMLF1$_{280\text{-}288}$	GLCTLVAML	HLA-A*02	B2M	54	299	152
	EBV	BZLF1$_{190\text{-}197}$	RAKFKQLL	HLA-B*08	B2M	17	225	149
	EBV	BRLF1$_{109\text{-}117}$	YVLDHLIVV	HLA-A*02	B2M	6	66	51
	IAV	M1$_{58\text{-}66}$	GILGFVFTL	HLA-A*02	B2M	50	239	138
	IAV	HA$_{306\text{-}318}$	PKYVKQNTLKLAT	HLA-DRA*01	HLA-DRB1*01,04	11	56	50
	HCV	NS3$_{1073\text{-}1081}$	CINGVCWTV	HLA-A*02	B2M	7	76	39
	HCV	NS3$_{1406\text{-}1415}$	KLVALGINAV	HLA-A*02	B2M	4	65	65
	HCV	NS3$_{1436\text{-}1445}$	ATDALMTGY	HLA-A*01	B2M	7	152	139
	HSV-2	VP22$_{49\text{-}57}$	RPRGEVRFL	HLA-B*07	B2M	5	68	29
	YFV	NS4B$_{214\text{-}222}$	LLWNGPMAV	HLA-A*02	B2M	5	223	198
	DENV1	NS3$_{133\text{-}142}$	GTSGSPIVNR	HLA-A*11	B2M	11	65	59
	DENV3-4	NS3$_{133\text{-}142}$	GTSGSPIINR	HLA-A*11	B2M	8	51	46
	HIV-1	p24$_{30\text{-}40}$	KAFSPEVIPMF	HLA-B*57	B2M	44	134	104
	HIV-1	p24$_{48\text{-}56}$	TPQDLNTML	HLA-B*42,81	B2M	21	52	40
	HIV-1	p24$_{128\text{-}135}$	EIYKRWII	HLA-B*08	B2M	12	81	60
	HIV-1	p24$_{131\text{-}140}$	KRWIILGLNK	HLA-B*27	B2M	27	212	141
	HIV-1	p24$_{161\text{-}180}$	FRDYVDRFYKTLRAEQASQE	HLA-DRA*01	HLA-DRB1*01,07,11,15, HLA-DRB5*01	17	141	95
	HIV-1	p24$_{223\text{-}231}$	GPGHKARVL	HLA-B*07	B2M	1	62	53
	HIV-1	Nef$_{90\text{-}97}$	FLKEKGGL	HLA-B*08	B2M	21	104	78

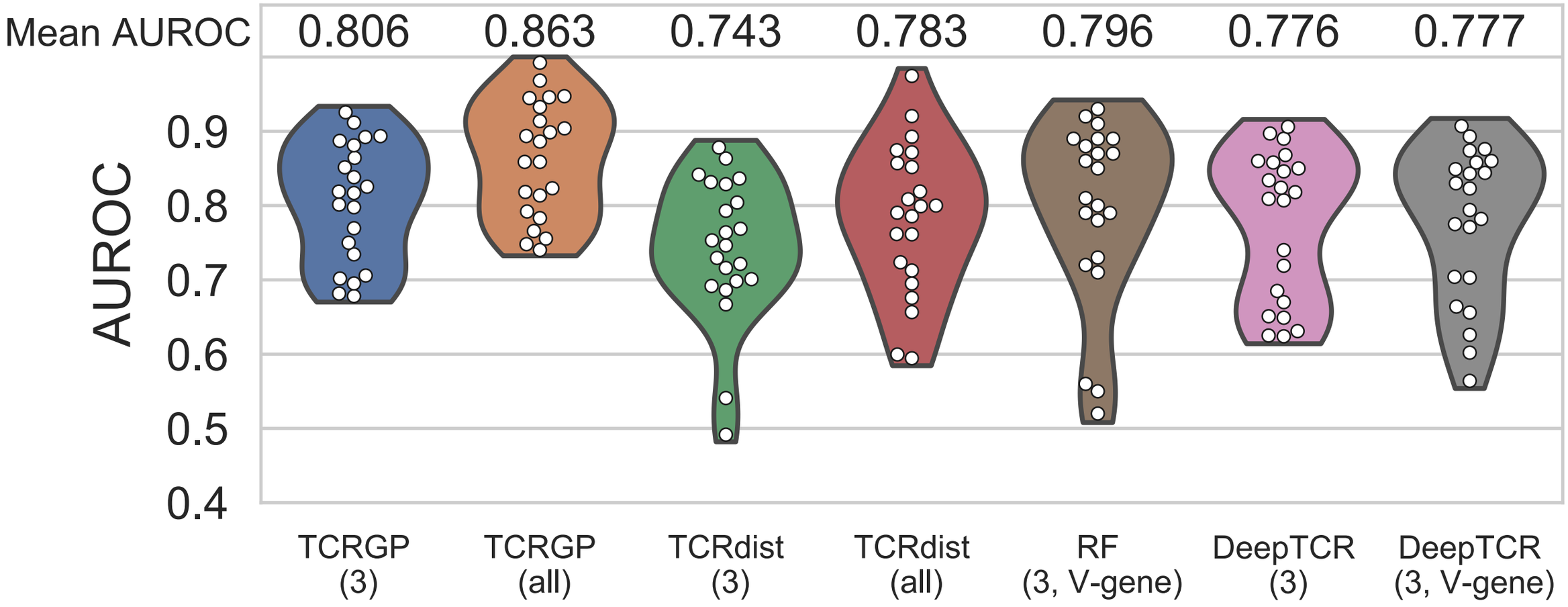
AUCs for 10 epitopes: Comparing TCRGP and TCRdist using leave-one-subject-out crossvalidation

B

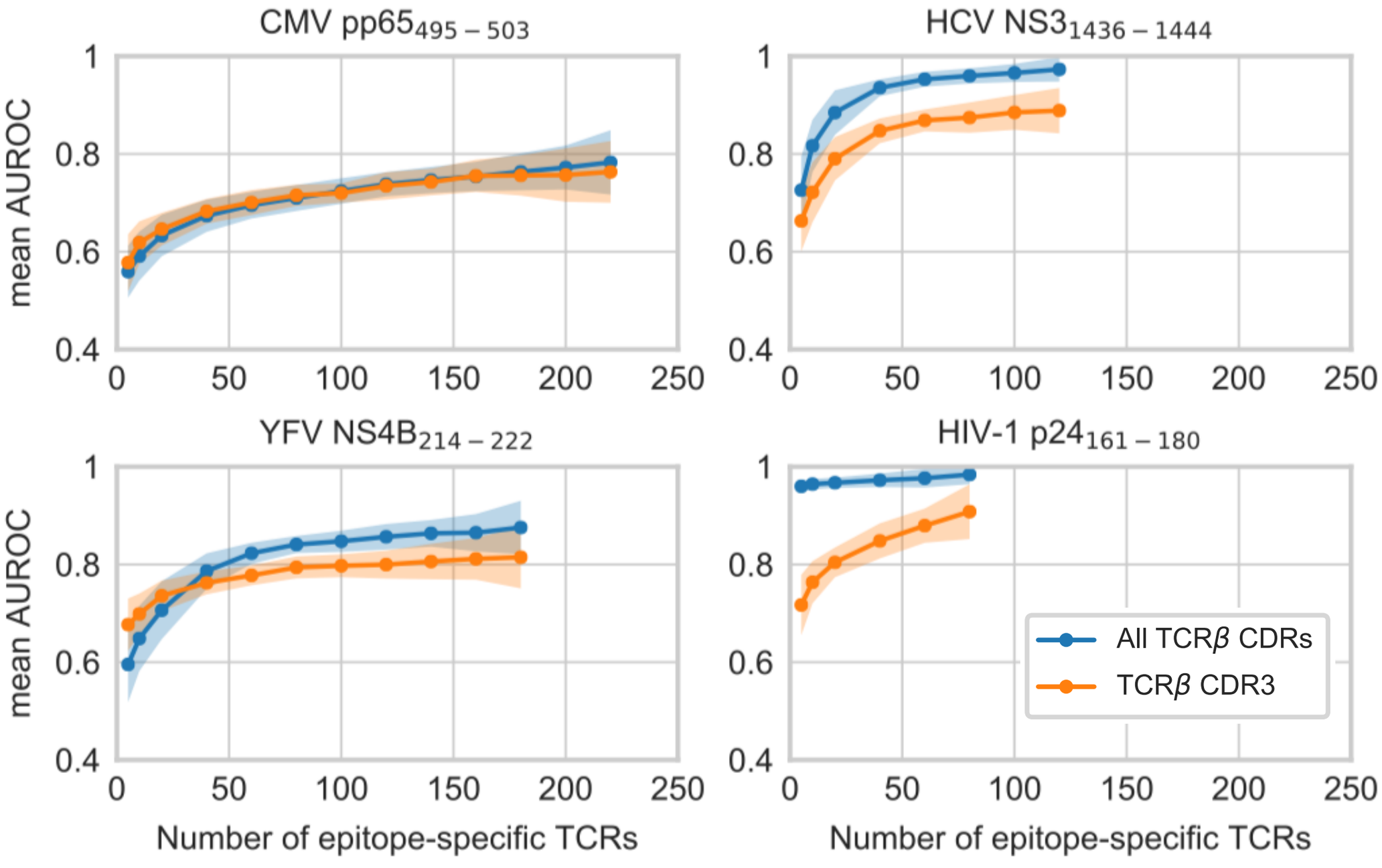
Method	TCRGP	TCRdist	TCRGP	TCRdist	TCRGP	TCRdist	TCRGP	TCRdist	TCRGP	TCRdist	TCRGP	TCRdist
Mean AUC	0.91	0.83	0.87	0.84	0.88	0.79	0.83	0.81	0.86	0.74	0.83	0.78



AUCs for 22 epitopes for VDJdb data: Comparing several methods

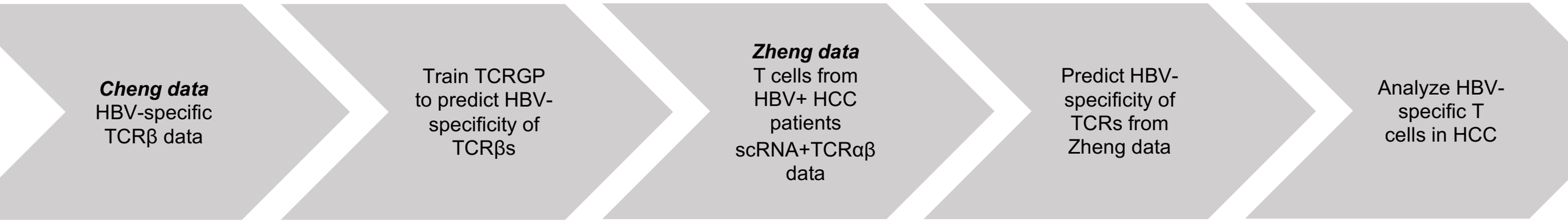


How many epitope-selected TCRs are needed to build a reliable/robust prediction model?



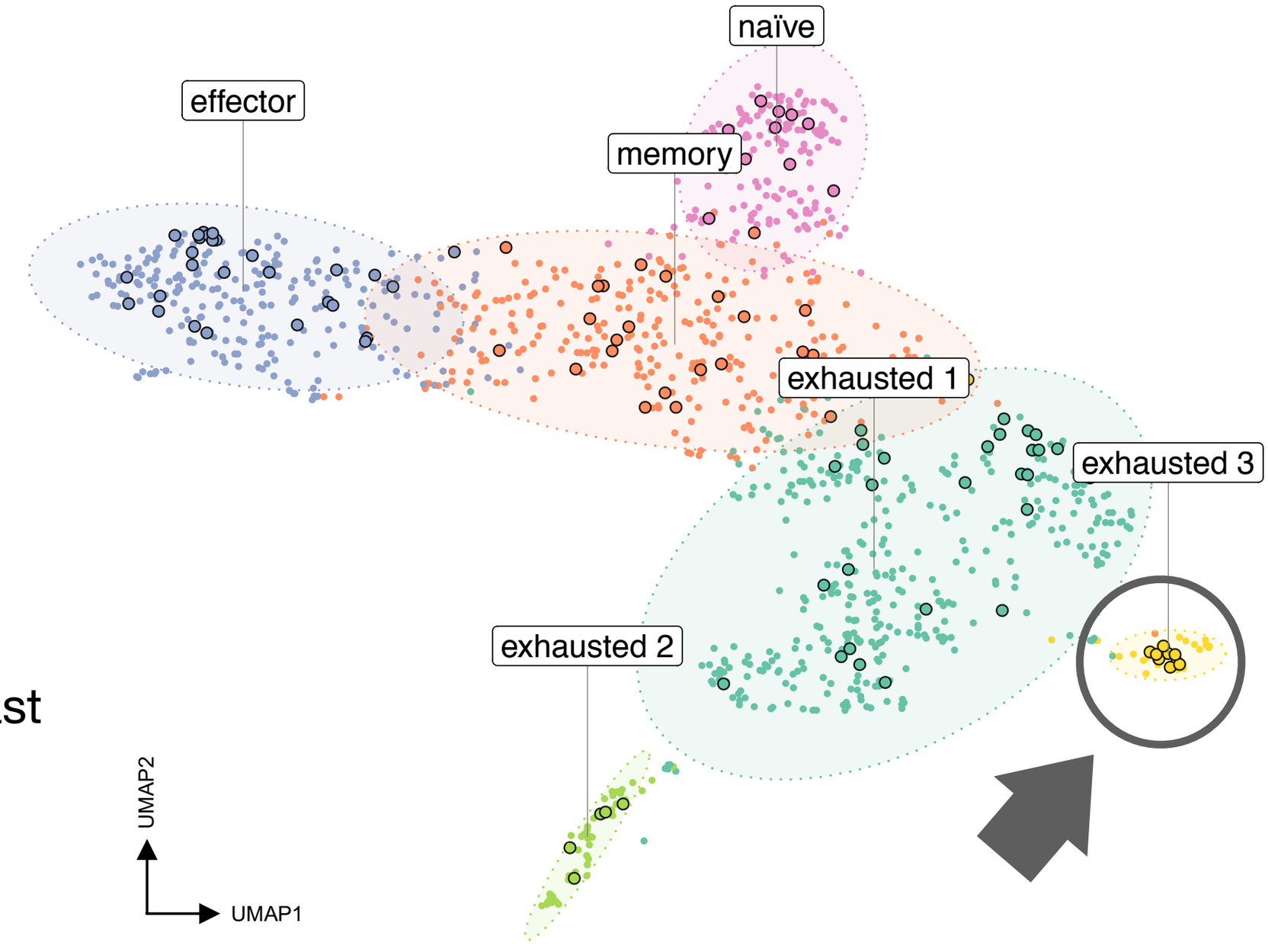
Combining TCR-peptide recognition prediction with scRNA-seq analysis

- Can we gain more insight into diseases using combined TCR-seq+scRNA-seq?
- An example of HBV virus in hepatocellular carcinoma (HCC)



Analysis of TCR-seq+scRNAseq from HBV+ hepatocellular carsinoma patients

- Can identify which phenotypes HBV-recognizing T cells are enriched to
- Most exhausted and least functional



Other references

[1] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2007). *Molecular biology of the cell*. Garland Science, 5 edition.

[2] Robins, H. S. *et al*. (2009). Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. *Blood*, 114(19), 4099–4107.

[3] Lefranc, M. (1999). The IMGT unique numbering for immunoglobulins, T-cell receptors, and Ig-like domains. *Immunologist*, 7(4), 132–136.

[4] Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. The MIT Press.

[5] Daniel Joseph Laydon, Charles R M Bangham, Becca Asquith (year). Estimating T-cell repertoire diversity: Limitations of classical estimators and a new approach.

[6] Scott Brown, Lisa A. Raeburn, Robert A. Holt (2015) Profiling tissue-resident T cell repertoires by RNA sequencing