Andreas Mayer
Division of Infection and Immunity
Institute for the Physics of Livings Systems

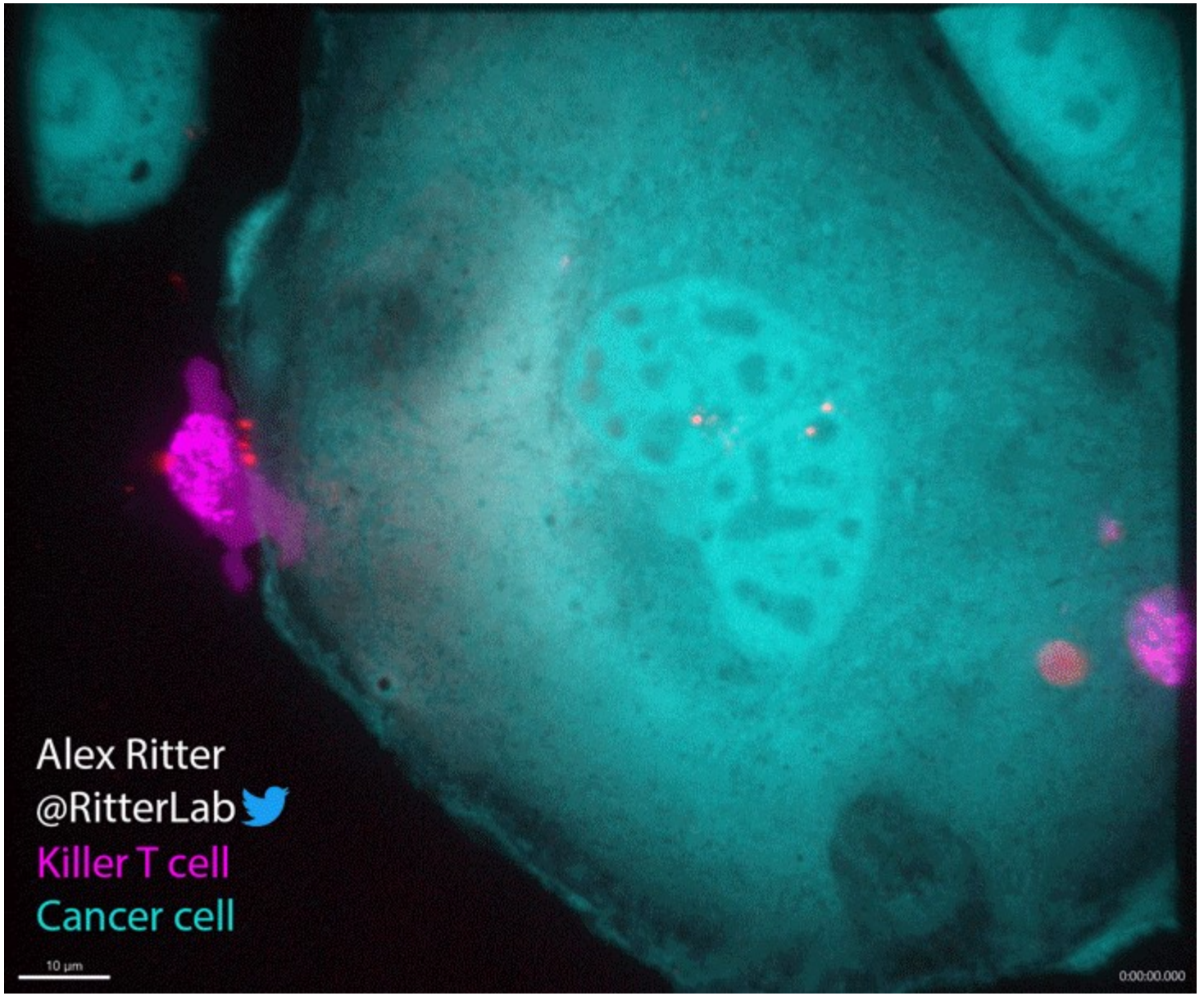
EPFL BEvAS

Imprints of antigen-driven selection in immune repertoires

LDN QIMMUNO

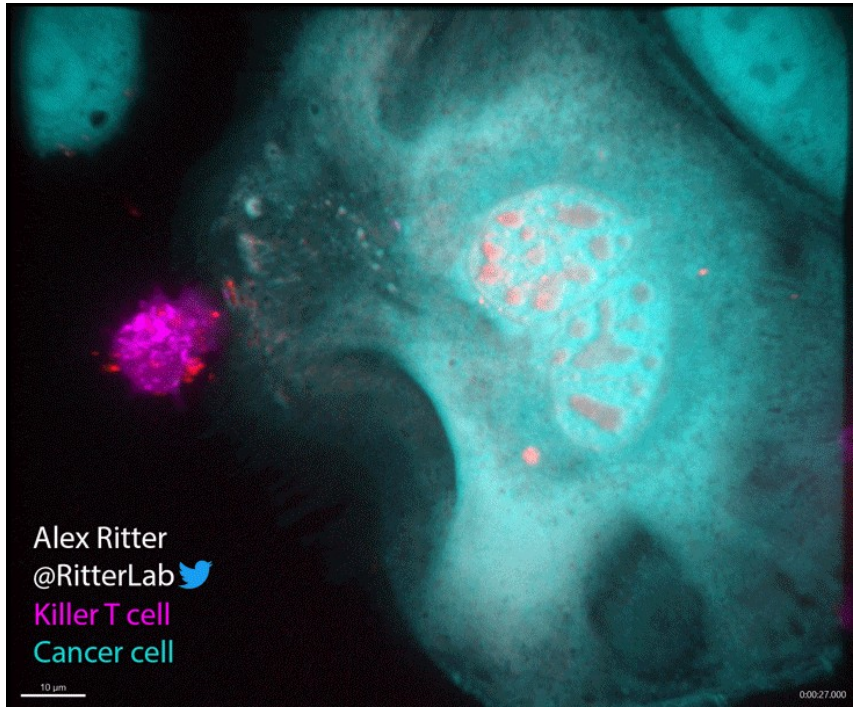
@andimscience

www.qimmuno.com



Ritter lab, Altos San Francisco

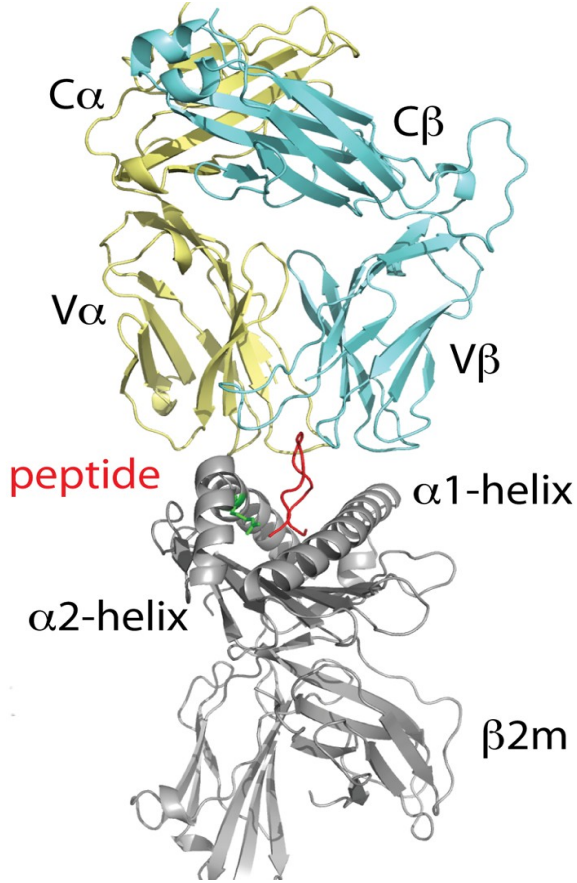
How do killer T cells know what to attack?



Ritter lab, Altos San Francisco

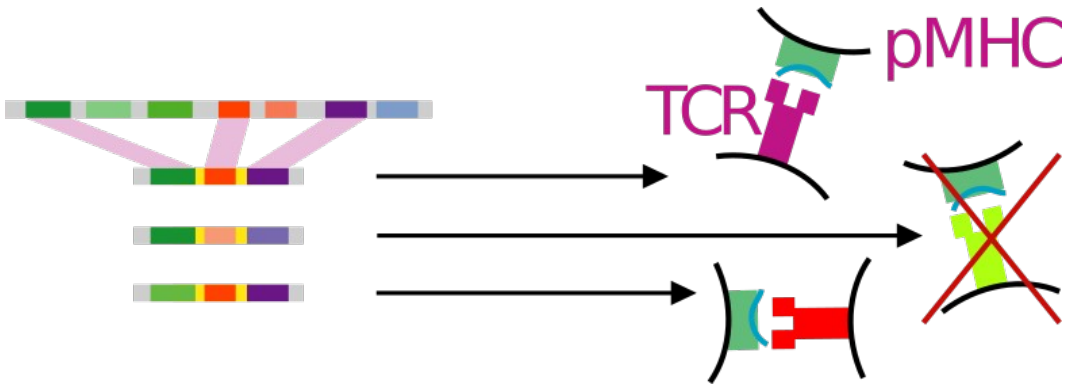
T cell receptor (TCR)

Peptide presented on target cell

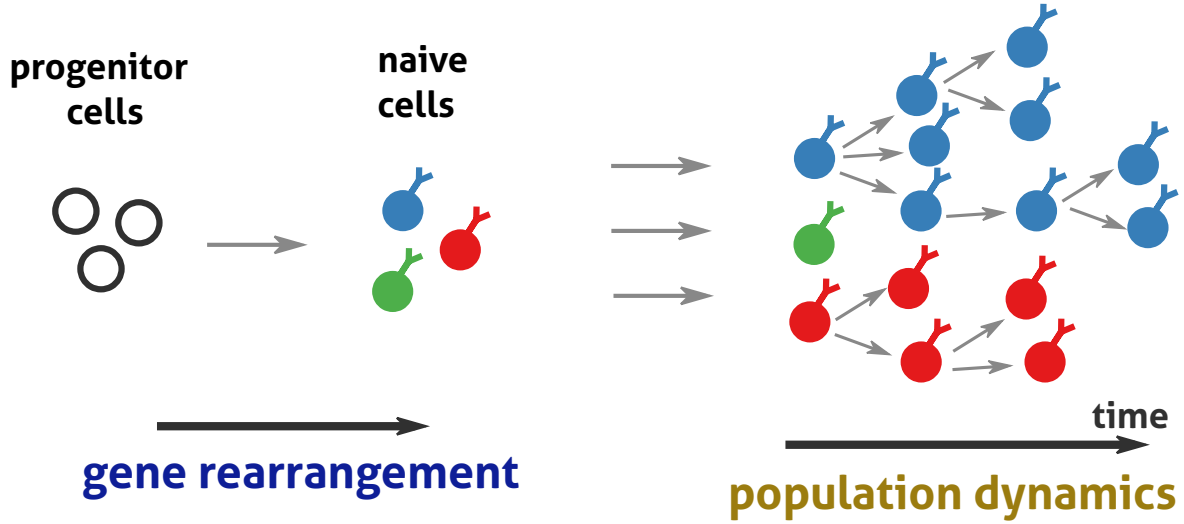


Cc-by Reboul et al. Plos CB 2012

Defending against the unknown



Cc-by Sethna et al. Plos CB 2020



1. Generate cells with diverse receptors through genetic recombination

2. Amplify clones with useful receptor specificity dynamically

Diverse pathogens finding their match

- pMHC diversity
  # peptides = 20 # amino acids

  k=9 → 10^{12} up to k=25 → 10^{32}

  >10^4 HLA alleles

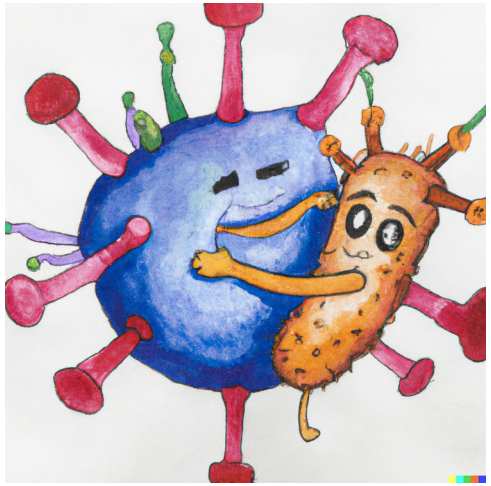
- TCR diversity

  # insertion profiles = 4 # insertions

  - Beta : 25 VDins + 25 DJins ~ 10^{30}
  - Alpha : 25 VJins ~ 10^{15}

} > 10^{30}

} > 10^{45}

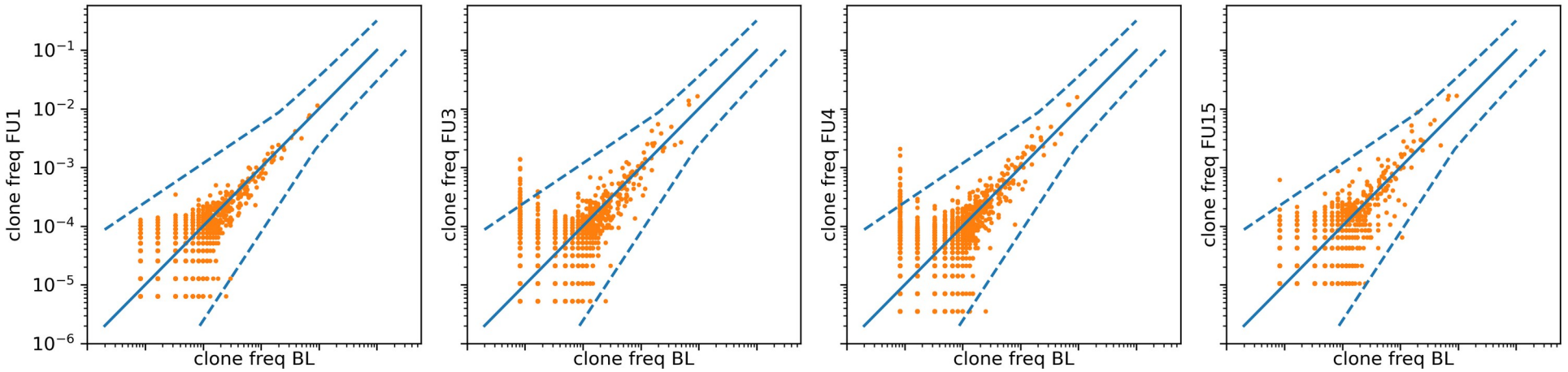


DallE's impression of T cells matching pathogens

Tracking immune responses by TCRseq

Large diversity of receptors
→ built-in “barcodes” for clones

Week 2: PCR+ for SARS-CoV-2



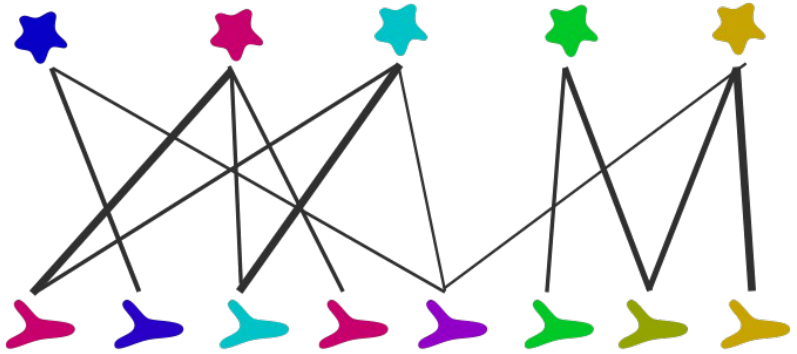
Dashed lines – null variation in statistical model of sampling process

Milighetti et al bioRxiv 2022

Bottleneck: Decoding receptor specificity?

Match dynamic clones to TCRs of **known specificity:**

- **sort specific T cells** for different epitopes
- **sequence** their TCRs



> 10^{30} pMHC

> 10^{45} TCR

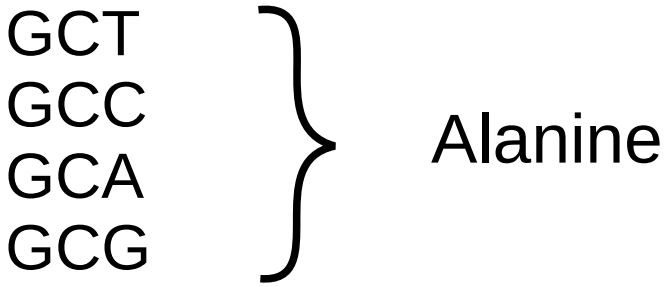
But:

Individuals respond with distinct TCRs to same epitope

Degeneracy of the TCR sequence code

Different TCRs bind the same pMHC target

Degeneracy of the genetic sequence code



Different codons translate to the same amino acid

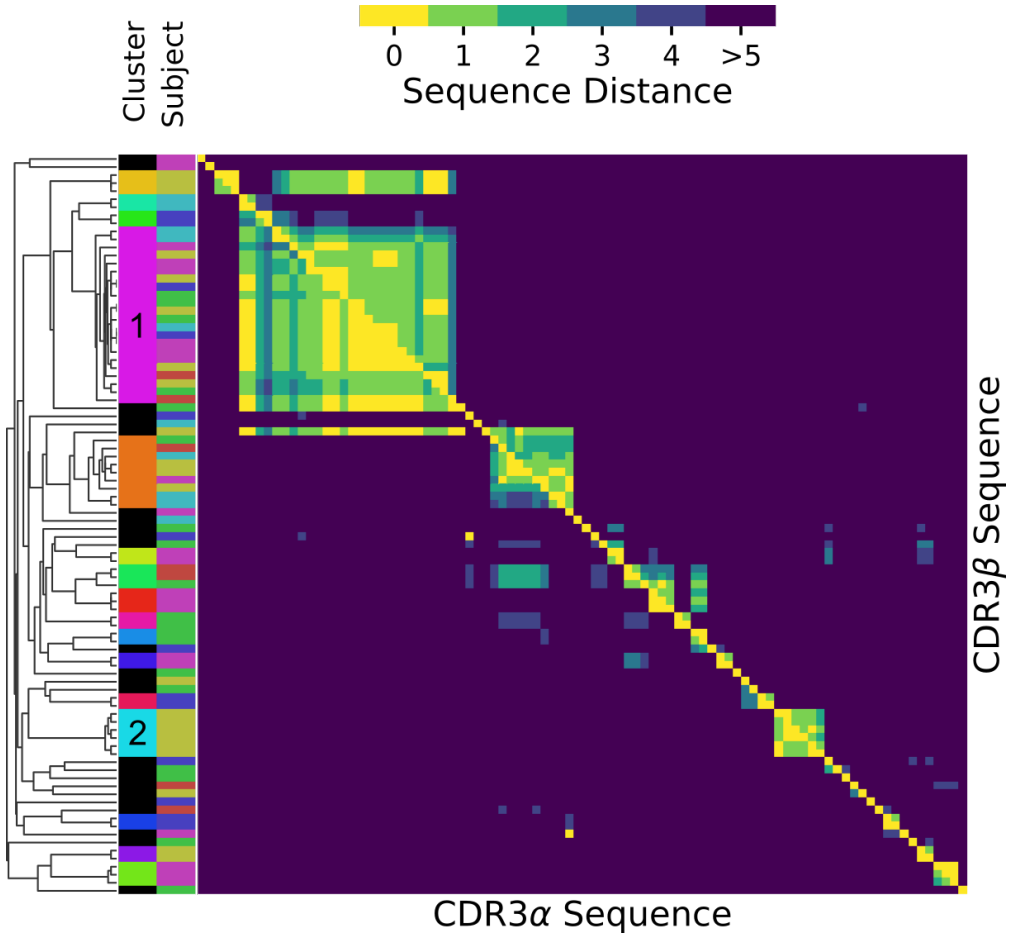
Learning the rules of the TCR binding code

Given sequence of two TCRs
Prob(TCR_1 and TCR_2 bind same ligand)?

Measuring correlation functions of selection factors on sequence space
Mayer, Callan PNAS 2023

Learning the right sequence space direction
Pyo, Henderson, Wingreen, **Mayer**, in preparation

pMHC-specific TCRs are enriched for similar sequences

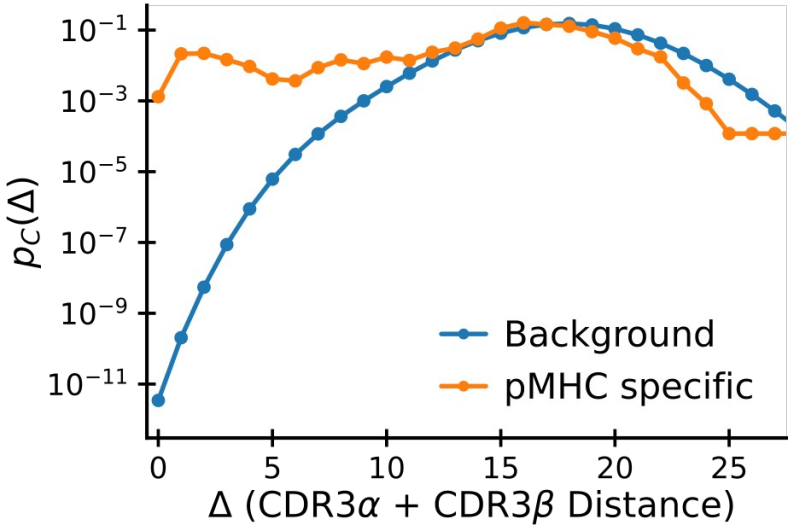


data: Dash … Thomas Nature 2017

Distribution of off-diagonal entries:
Probabilities of near-coincidences

$$p_C(\Delta) = \sum_{\{\boldsymbol{\sigma},\boldsymbol{\sigma}'\}} P(\boldsymbol{\sigma})P(\boldsymbol{\sigma}')\boldsymbol{I}_{d(\boldsymbol{\sigma},\boldsymbol{\sigma}')=\Delta}$$

Sequences at distance Δ



Mayer, Callan PNAS 2023

Detecting selection in sequence repertoires

$P(\boldsymbol{\sigma})$

Blood sample

CASSWNGPTYEQYF
CSAPLGGGEQFF
CADPFRDRGSNQPQHF
CASSPGQGSYEQYF
CAISGQGTGEKYQPQHF
CSARDGTGNGYTF
CASSTIEGQGGRHTQYF
....

$Q(\boldsymbol{\sigma})$

Selection

$Q(\boldsymbol{\sigma})P(\boldsymbol{\sigma})$

Sorted for specificity

CSARDRVGNGYTF
CASSPGQGSYEQYF
CSARDGTGNGYTF
CSARDGTGNGYTF
...

How can we tell that the specific sequences are selected?

as opposed to just being a random subsample

Mayer, Callan PNAS 2023

Why this statistics?

Population genetics approach to detecting selection?
dN/dS – compare rates of non-synonymous and synonymous changes

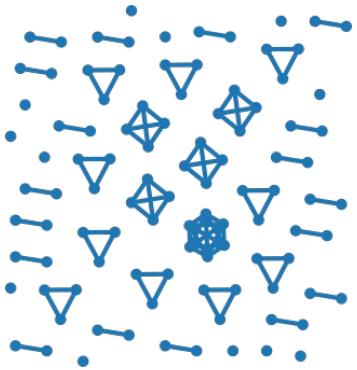
Here: selection on standing variation

→ coincidence counting

p_C = Prob(AA(NT_1) = AA(NT_2))

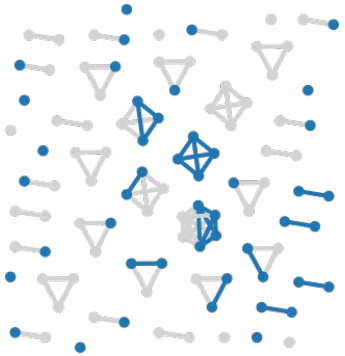
NT Sequence
Identical AA Sequence

$$p_C = \frac{2|E|}{|V|(|V|-1)}$$



p_C = 0.02

Select random cells



p_C = 0.02

Mayer, Callan PNAS 2023

Why this statistics?

Population genetics approach to detecting selection?
dN/dS – compare rates of non-synonymous and synonymous changes

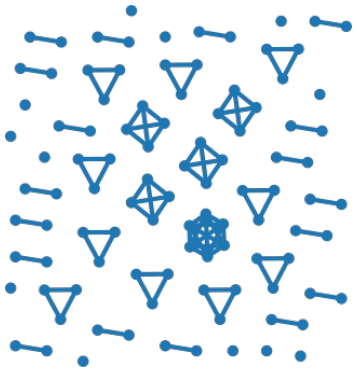
Here: selection on standing variation

→ coincidence counting

p_C = Prob(AA(NT_1) = AA(NT_2))

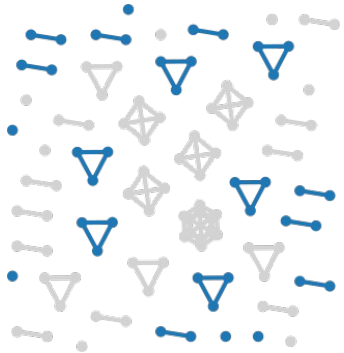
NT Sequence
Identical AA Sequence

$$p_C = \frac{2|E|}{|V|(|V|-1)}$$



p_C = 0.02

Select random clonotypes



p_C = 0.04

Mayer, Callan PNAS 2023

Theory of near coincidence counting

$$\frac{p_C[QP](\Delta)}{p_C[P](\Delta)} = \langle Q(\boldsymbol{\sigma})Q(\boldsymbol{\sigma}')\rangle_{\boldsymbol{\sigma}\overset{\Delta}{\approx}\boldsymbol{\sigma}'}$$

Pairs at d=Δ

Near-coincidence ratios = how selection co-varies with sequence

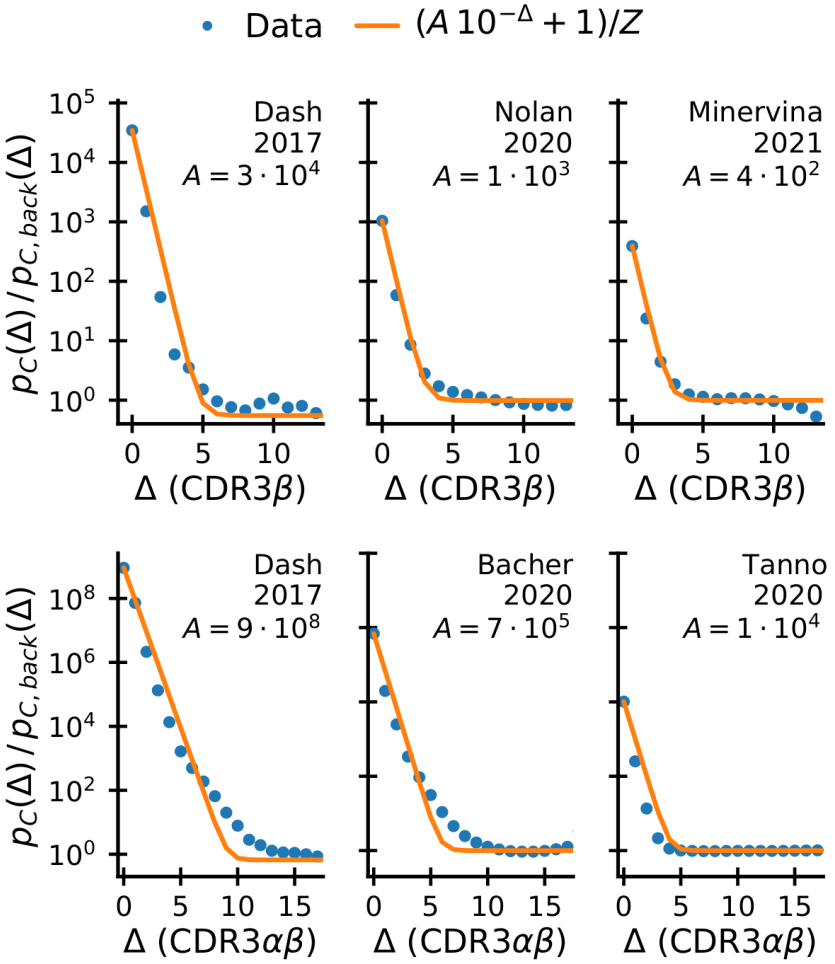
Selection for specific binding:

$$\frac{p_C[QP](\Delta)}{p_C[P](\Delta)} = \frac{p_C[QP](0)}{p_C[P](0)} \langle f_{\boldsymbol{\sigma}}(\Delta)\rangle_{\boldsymbol{\sigma}\in S}$$

Fraction specific at d =Δ

Measuring autocorrelation function of selection factor on sequence space

Common signature across modalities



→ roughly exponential falloff by factors of 10 for each substitution

Mayer, Callan PNAS 2023

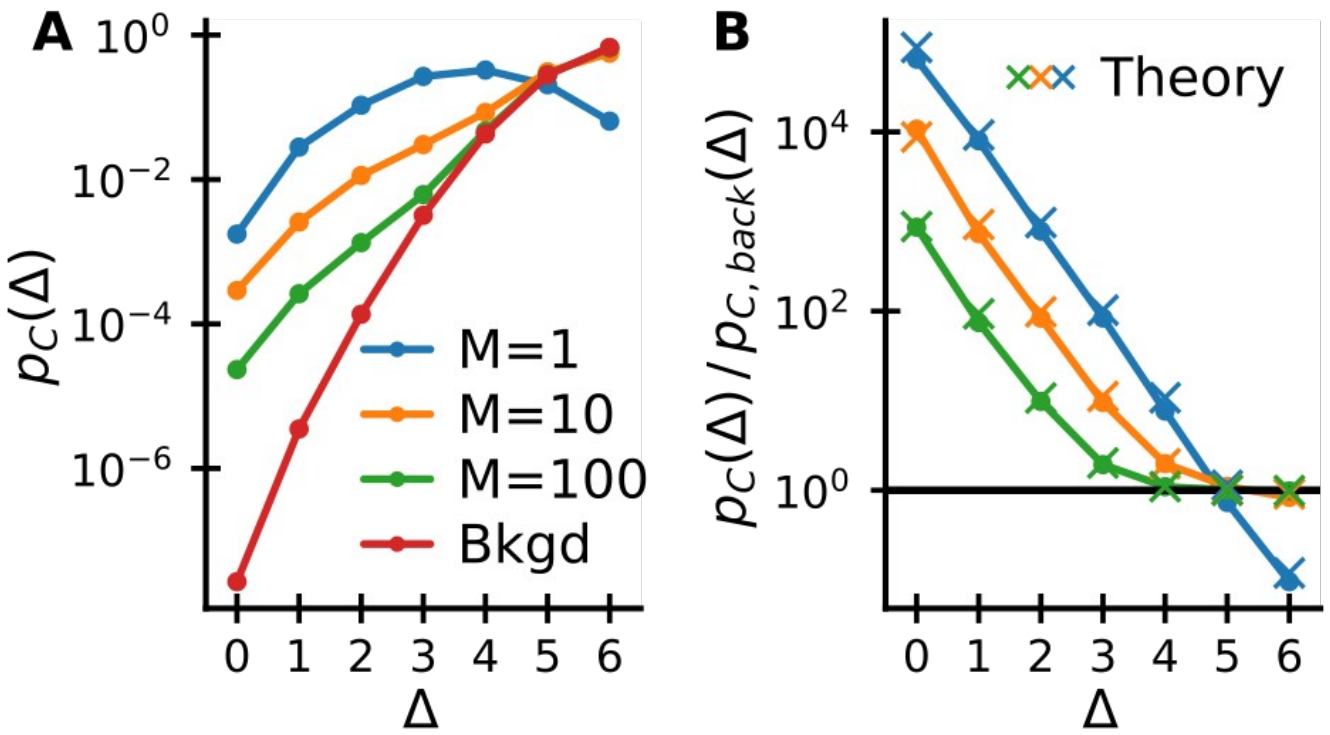
Signature is explained by simple biophysical binding models

Modeling binding energies?

→ Mix of independent site models

$$E_p(\sigma) = \sum_{i=1}^{k} \epsilon_p(i, \sigma_i),$$

$$\epsilon_p(i, \sigma_i) = \begin{cases} -1 & \text{for } \sigma_i \in \mathcal{S}_i^p \\ 0, & \text{otherwise.} \end{cases}$$



Mayer, Callan PNAS 2023

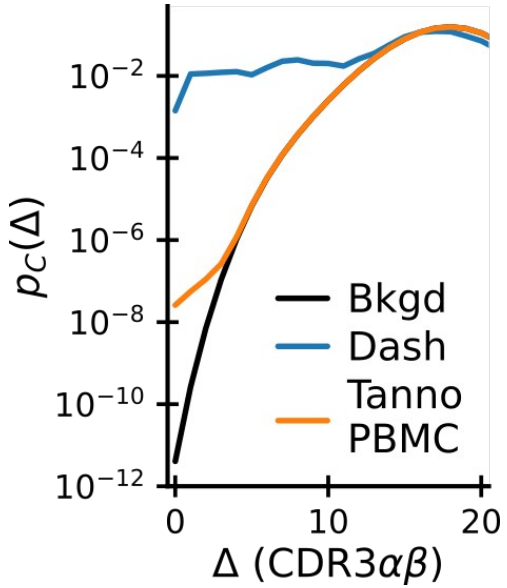
Functional diversity of the immune repertoire

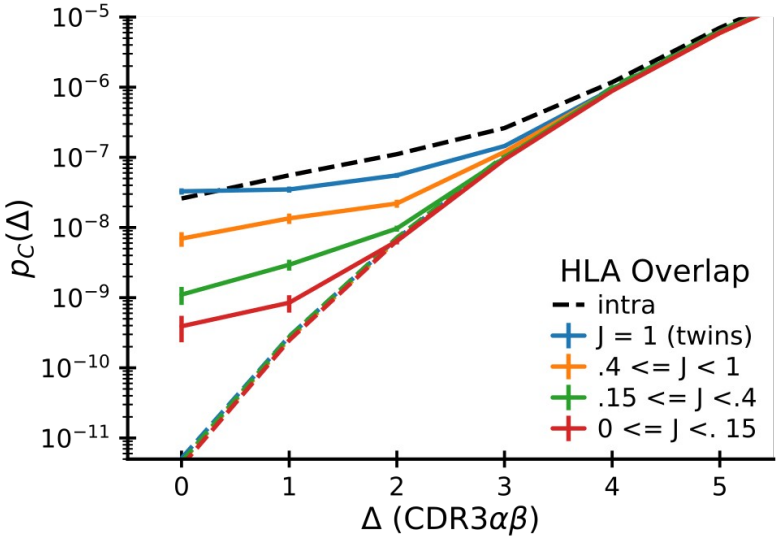
Repertoire = mixture of epitope specific groups

$$P(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\pi} \in \boldsymbol{\Pi}} P(\boldsymbol{\sigma}|\boldsymbol{\pi}) P(\boldsymbol{\pi}),$$

Decomposition theorem:

$$p_C[P(\boldsymbol{\sigma})] = p_C[P(\boldsymbol{\pi})] \, \langle p_C[P(\boldsymbol{\sigma}|\boldsymbol{\pi})] \rangle + (1 - p_C[P(\boldsymbol{\pi})]) \, \langle p_C[P(\boldsymbol{\sigma}|\boldsymbol{\pi}_1), P(\boldsymbol{\sigma}|\boldsymbol{\pi}_2)] \rangle,$$





Mayer, Callan PNAS 2023

Predicting shared specificity?

CASSWNGPTYEQYF – HLA-A2-BMLF1$_{280}$ → CASS**A**NGPTYEQYF –HLA-A2-BMLF1$_{280}$?

So far:

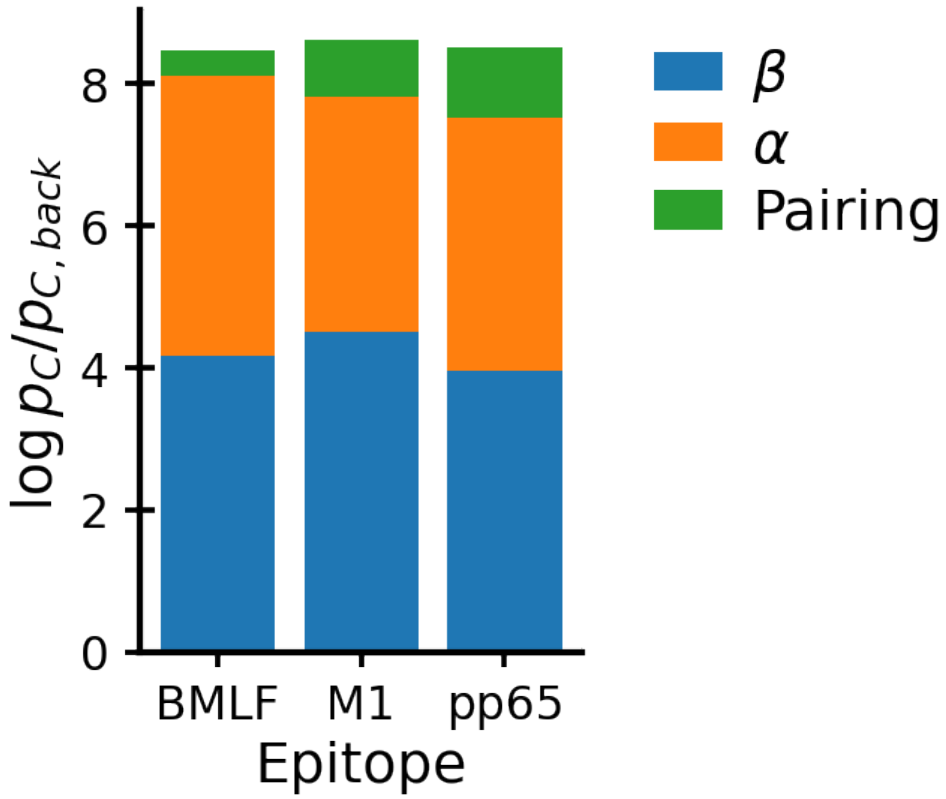
$d(TCR_1, TCR_2) \rightarrow \langle P(TCR_2 \text{ binds X} \mid TCR_1 \text{ binds X})$

In what ways are pMHC-specific TCRs similar?

$\langle P(TCR_2 \text{ binds X} \mid TCR_1 \text{ binds X})\rangle \rightarrow d(TCR_1, TCR_2)$

Pyo, Henderson, Wingreen, **Mayer**, in preparation

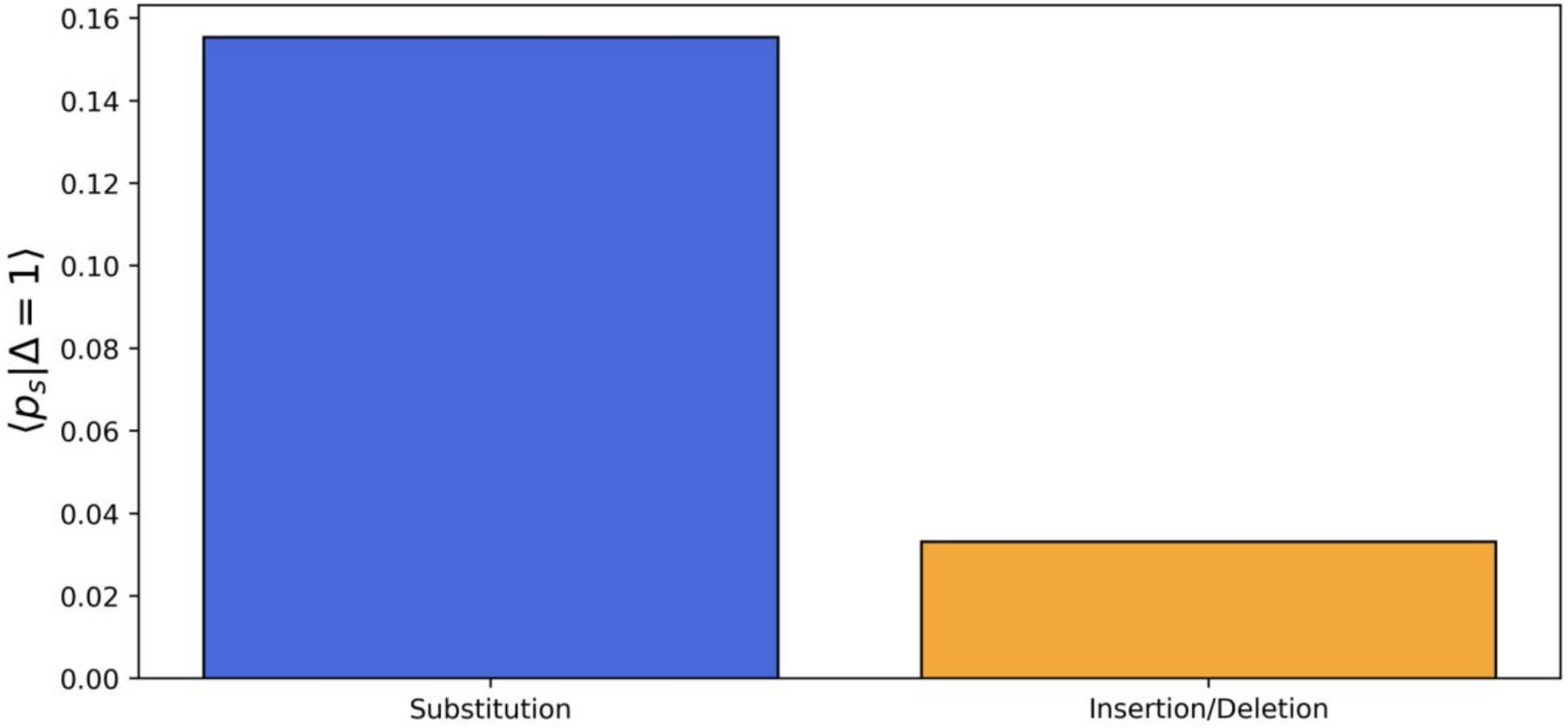
Decomposing selection into parts



Both receptor chains and their (conditional!) pairing contribute

Mayer, Callan PNAS 2023

Substitutions less disruptive than indels



Pyo, Henderson, Wingreen, **Mayer**, in preparation

Generalized as metric learning problem

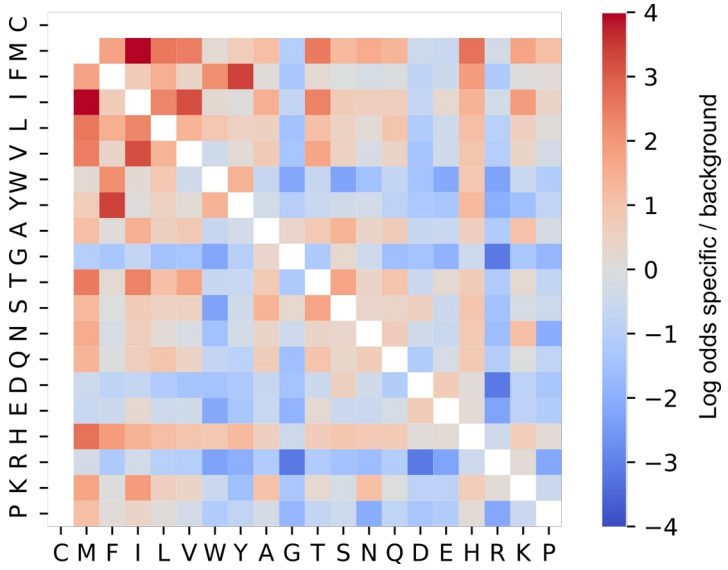
d_θ( , ) < d_θ( , )

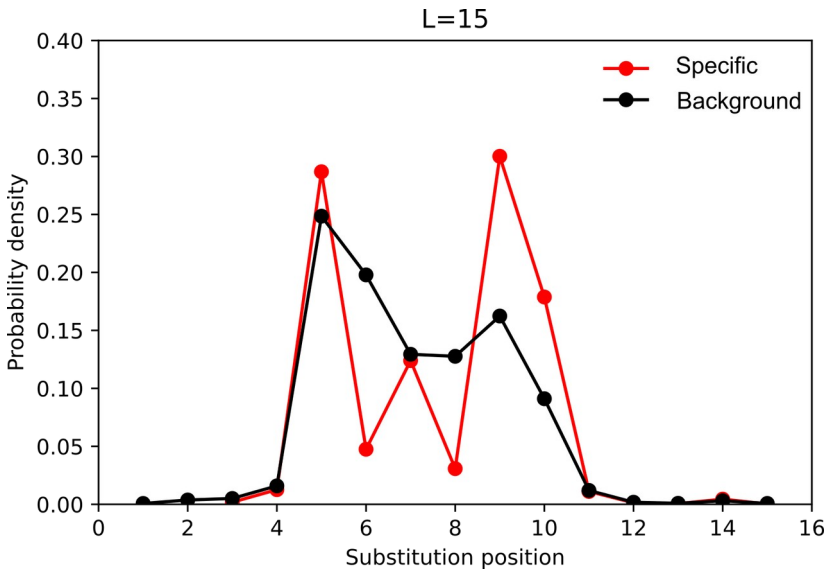
Determine optimal θ by gradient descent from training data

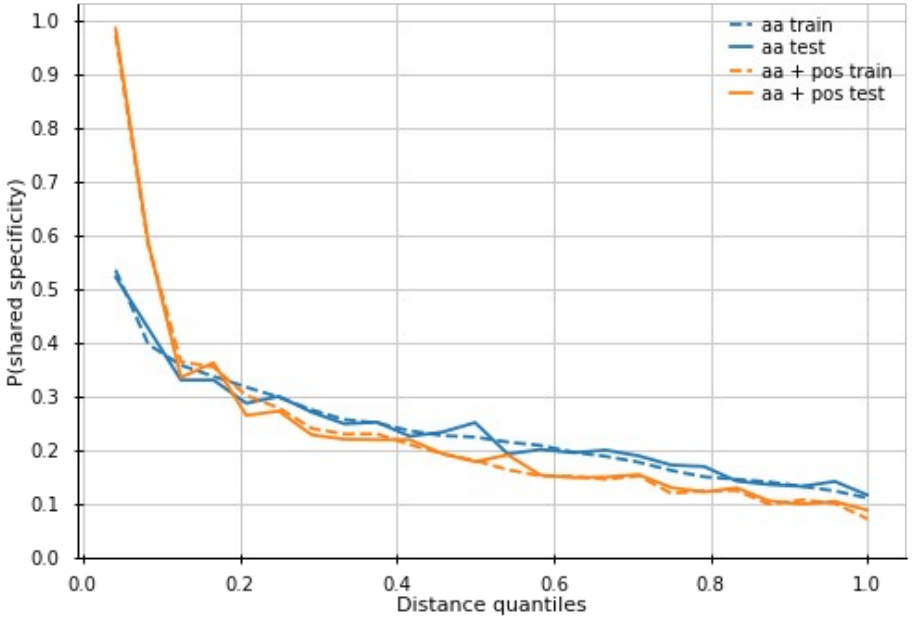
→ From sequences to biophysical features that represent specificity

Pyo, Henderson, Wingreen, **Mayer**, in preparation

Learned biophysical rules predict binding specificity

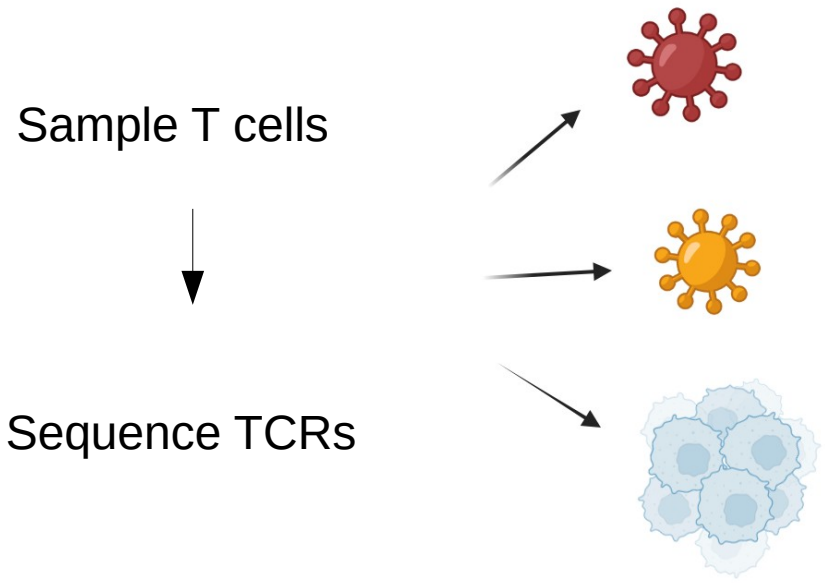






Pyo, Henderson, Wingreen, **Mayer**, in preparation

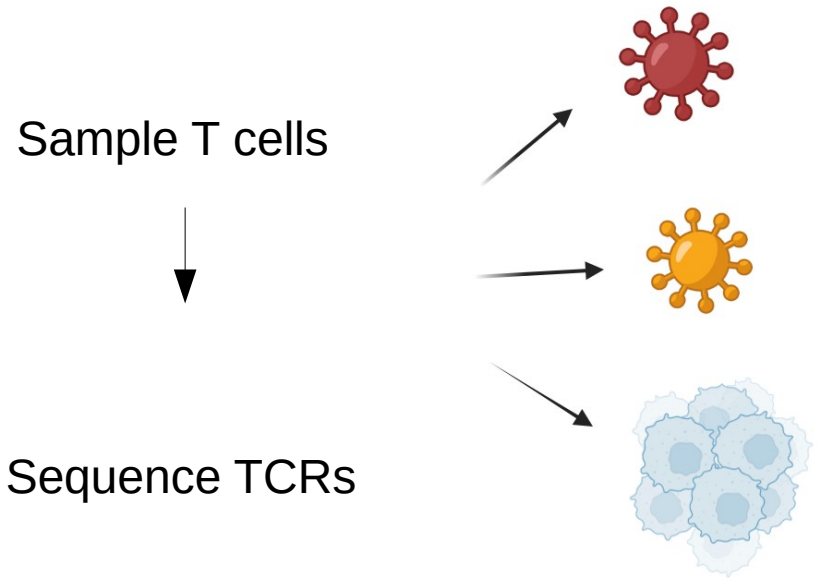
Promise of AI for TCR – pMHC binding prediction for immunology



Made with Biorender

- Unbiased **profiling of T cell responses** at epitope level & reverse epitope discovery
- Multiplexed **biomarkers** in infectious disease, autoimmunity, cancer
- Rational **TCR engineering**
- Predict **crossreactivity**

Our focus: Metric learning as a feasible stepping stone



Made with Biorender

- Unbiased **profiling of T cell responses** at epitope level & reverse epitope discovery
- Multiplexed **biomarkers** in infectious disease, autoimmunity, cancer
- Rational **TCR optimization** ~~engineering~~
- Predict **crossreactivity**

Collaborators

Team

James Henderson, MSci
Touchchai Chotisorayuth, MSc

Co-supervised:
Andrew Pyo, Princeton PhD
Yuta Nagano, UCL PhD

UCL

Benny Chain
Mahdad Noursadeghi
Gabriele Pollara
Gillian Tomlinson

Leo Swadling
Hans Stauss
Jennifer Cowan

Princeton

Bill Bialek
Curtis Callan
Ned Wingreen

Sloan-Kettering

Ben Greenbaum
Vinod Balachandran
Chrysothemis Brown

UPenn
Vijay Balasubramanian

UC Irvine / WUSTL

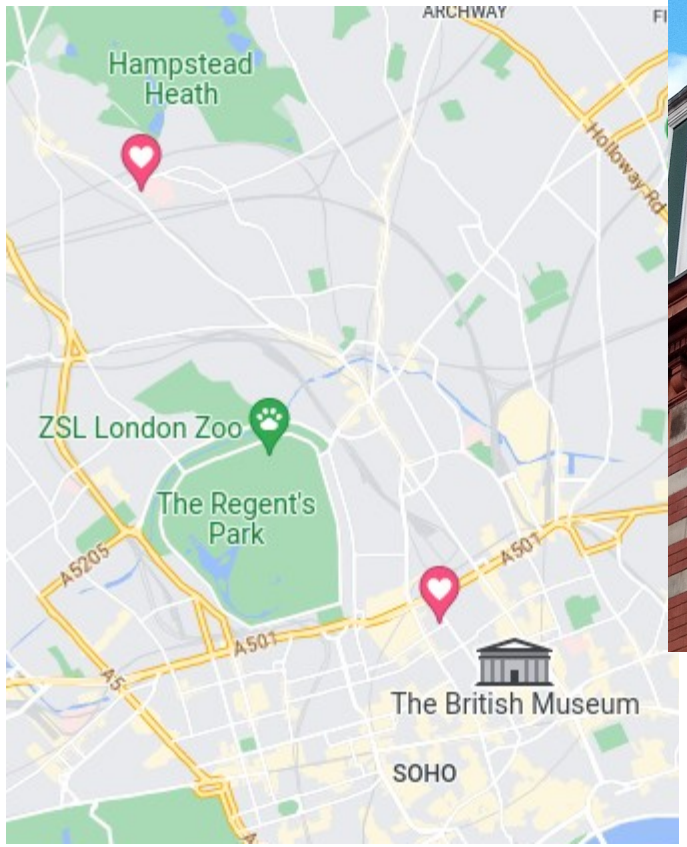
Lisa Wagar
Naresha Saligrama

Monash University

Nicole La Gruta

Oxford
Omer Dushek

Come visit!



@andimscience

Interested in joining the group? Reach out!

www.qimmuno.com