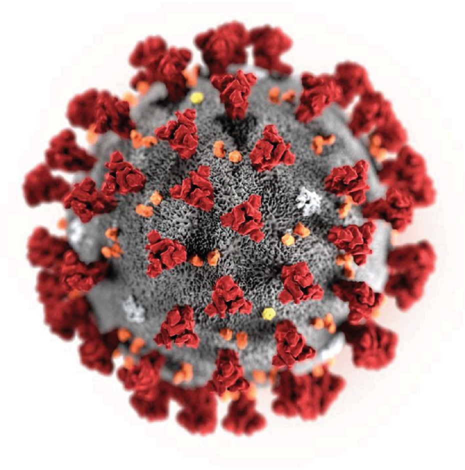
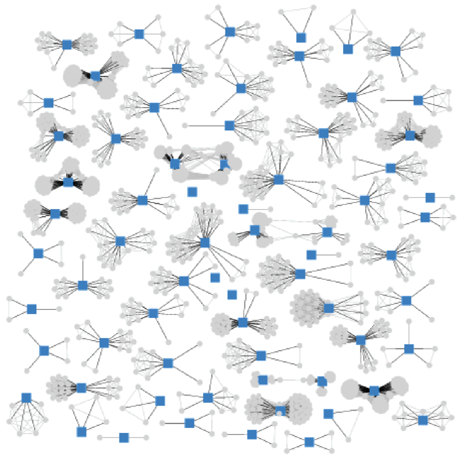




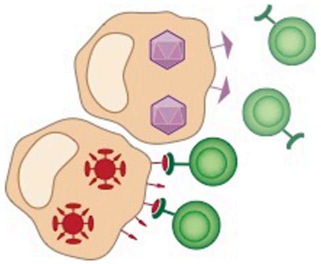
Lecture — February 3rd, 2025

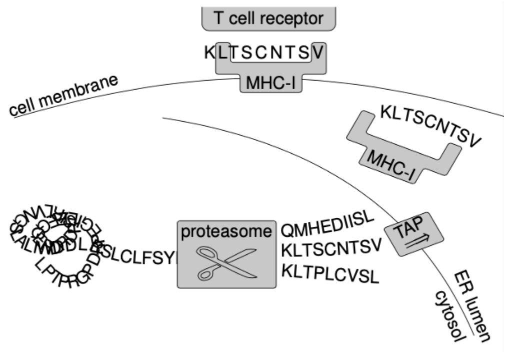
Could T cells learn French?

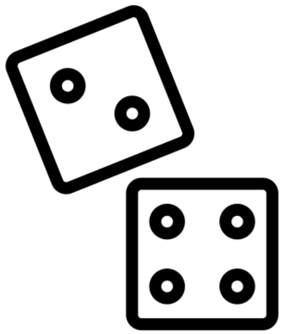
Immunology's "negative selection" as anomaly detector (continued)

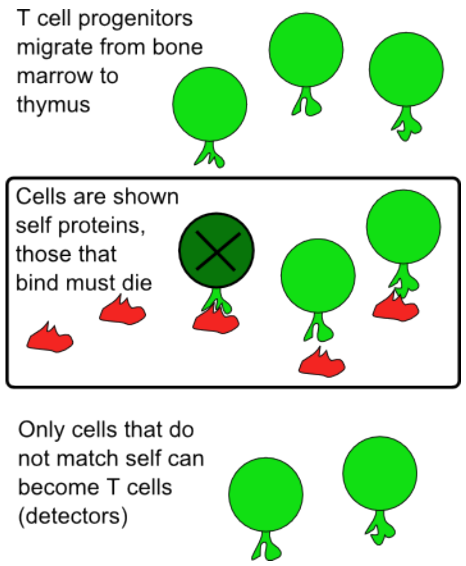
inge.wortel@ru.nl

Recap: T-cell immunology and negative selection









T cells recognize viruses and other anomalies inside other cells.

These cells present **peptides** on the outside via a molecule called MHC. **Each T-cell receptor (TCR)** is specific and detects only 1 : 100,000 peptides.

We need more TCRs than there are genes in the DNA, so a **random** DNA cut-and-paste **process generates diversity**.

Randomized TCRs might falsely **attack self-peptides**. The ones that do are killed during **negative selection**.

Before the break: how can algorithms inspired by this process solve anomaly detection?

Now: can such artificial immune systems help understand the biological process better?

Should you take this bet?

- You get: 1 hour with my (French*) dictionary
- Test time: I show you words, you tell me if they are French (Y/N)
- You win: €10 per correct guess (but: you pay *me* €10 for every wrong one)

Broad yet specific immunity: a numbers problem

T cells must **discriminate** "self" vs "non-self" peptide strings:

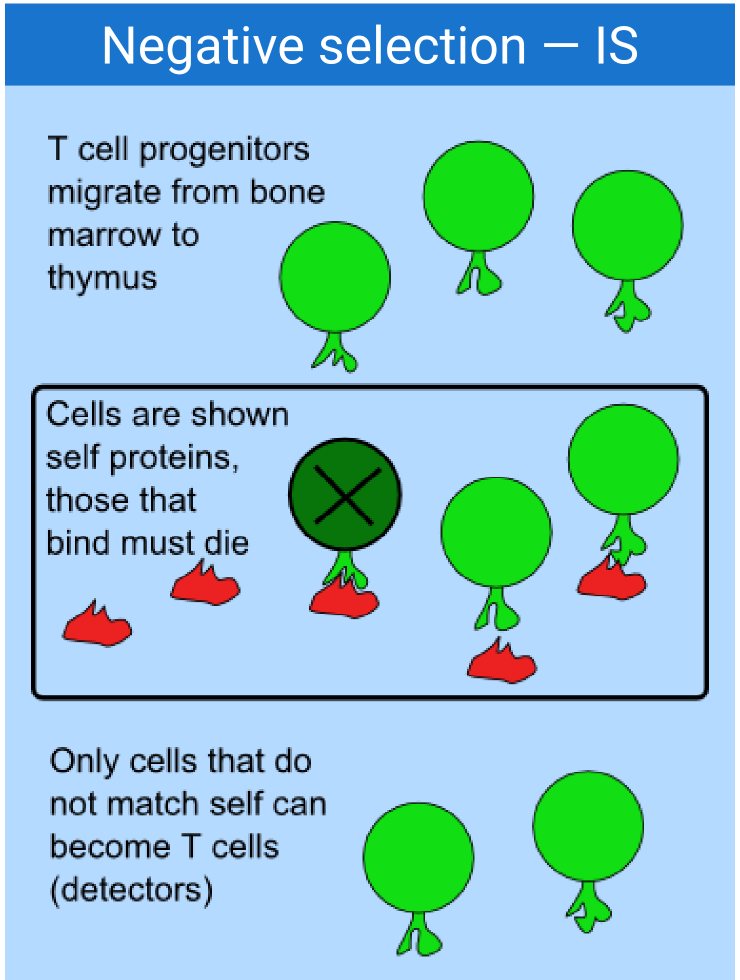
- recognize **$>10^{15}$** potential (and continuously changing) non-self peptides ... [1]
- ... while tolerating **hundreds of thousands** of self peptides

[1] Sewell (2012). *Nature Reviews Immunology.*

***Diversity*, but safe: solution?**

Two-step algorithm:

1. **Make a "repertoire" of many T cells.**
   A **stochastic** process ensures that this repertoire is both broad and diverse.
   ⟶ But: such a stochastic process also generates **self-reactive** T cells.
2. **Negatively select (i.e., kill) harmful cells.**
   Result: a "self-tolerant" T-cell repertoire.



(Forrest, 1997)

So... Problem solved?

Not really.

- Time in thymus: a few days
- Self peptides in "dictionary": half a million

It turns out that negative selection is **incomplete**:

- T cells likely "see" only a fraction of self-peptides in the thymus
- Many self-reactive cells remain even after negative selection.[1]

But if that's true, to what extent is negative selection still useful?

Can incomplete negative selection still be useful?

Return to our game. Mission impossible?

Can incomplete negative selection still be useful?

Return to our game. Mission impossible?

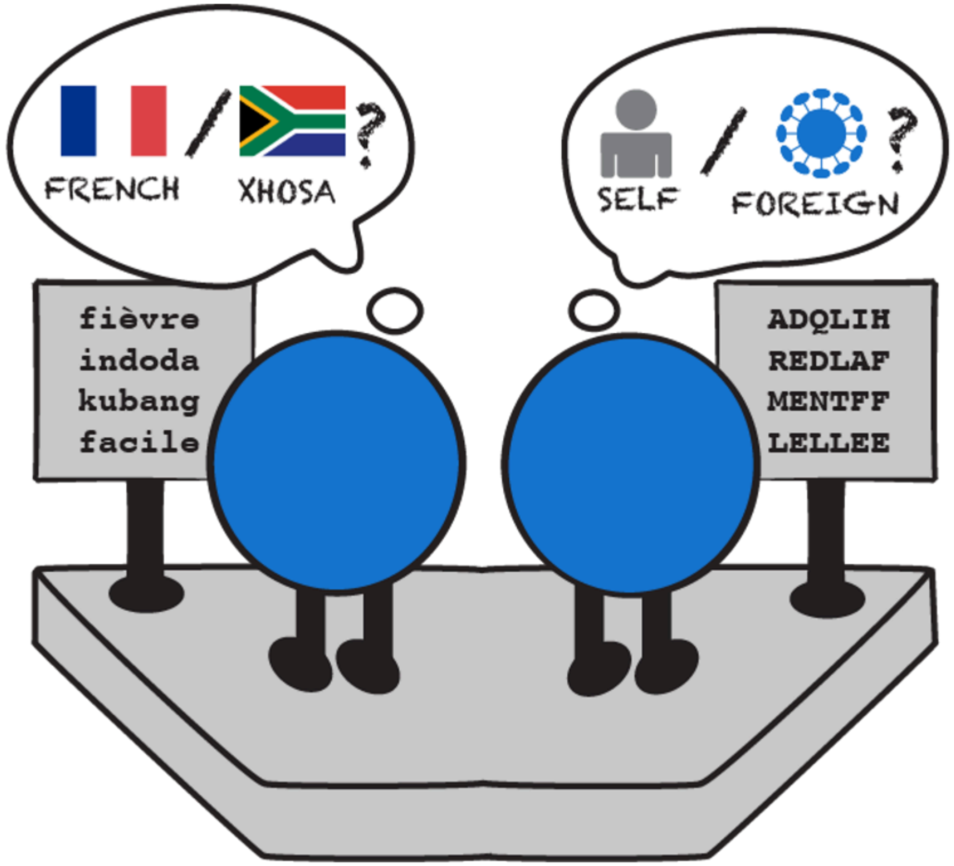
Which word is French?

fièvre
indoda

fièvre
indoda

Goat cheese and generalisation

fi**èvre** sounds/looks like ch**èvre**



If we can learn which **words** look like French, can T cells in the thymus learn which **peptides** look like self?

Tolerance and self-foreign discrimination

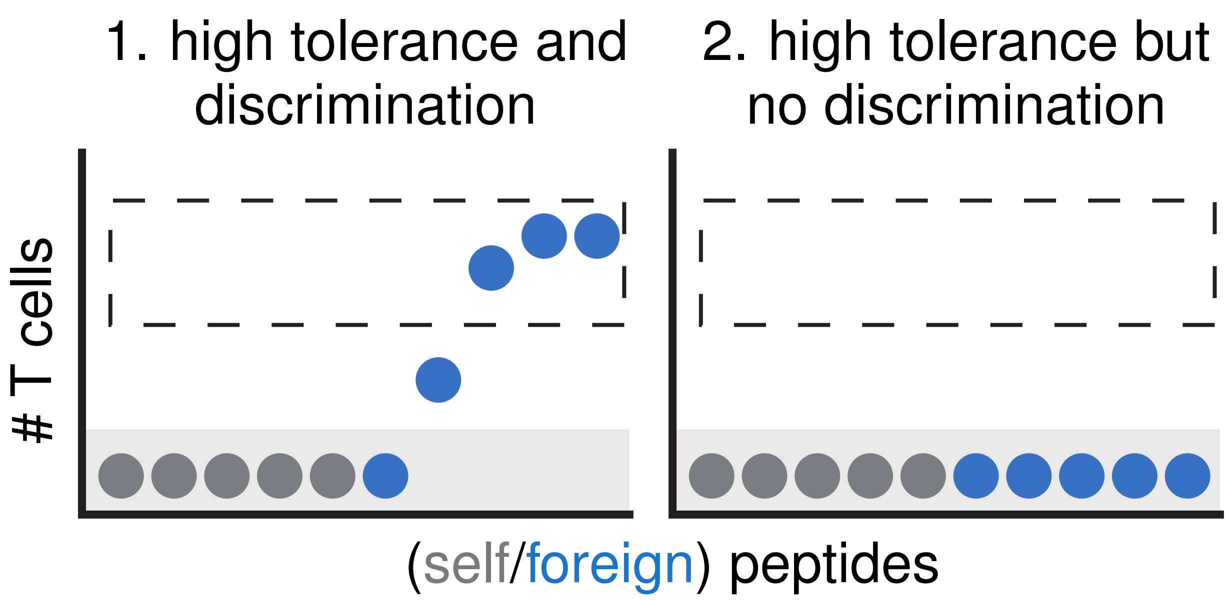
We have talked about "tolerance" (not having T cells against self) and "self-foreign discrimination".

We just saw that negative selection (and therefore tolerance) is incomplete.

Question
Are "tolerance" and "self-foreign discrimination" the same thing? Why (not)?

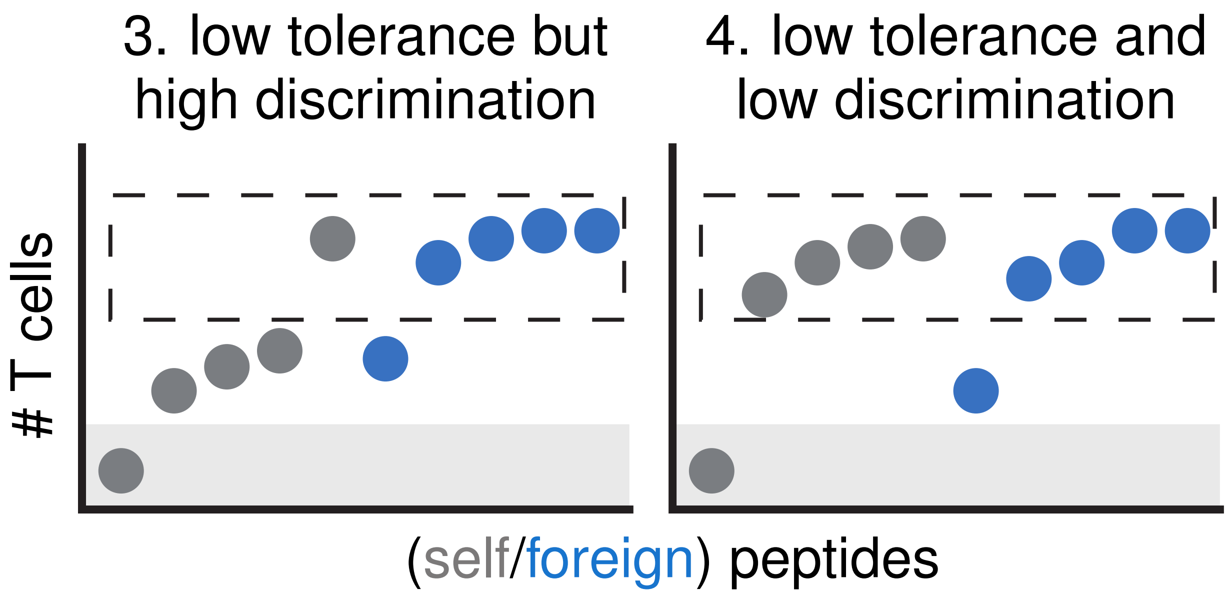
Tolerance ≠ discrimination.

Complete negative selection:

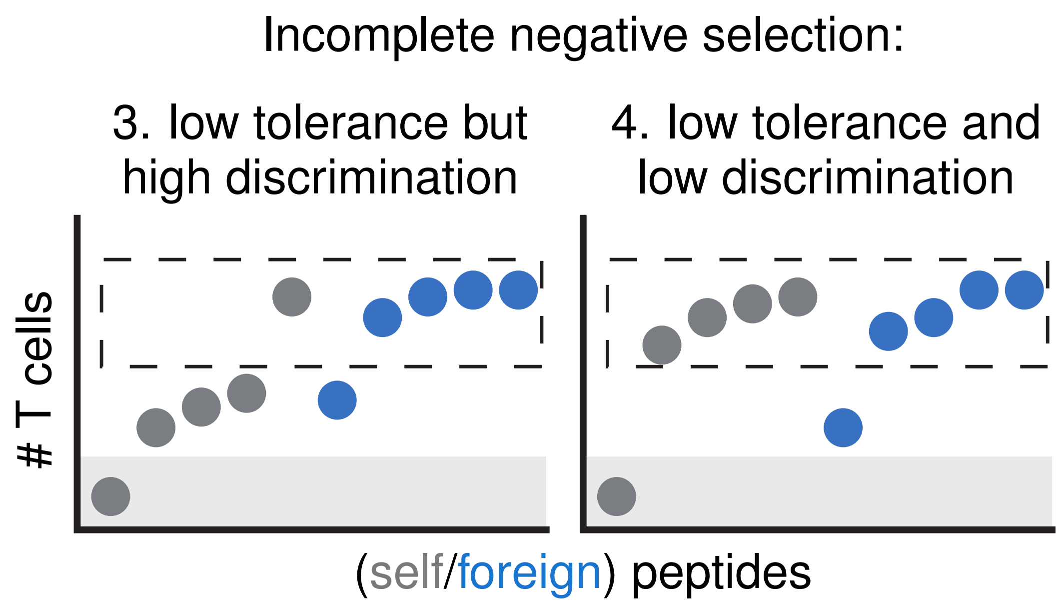


Tolerance ≠ discrimination.

Incomplete negative selection:



Which of these scenarios holds?



Given that negative selection is **incomplete**, can T cells distinguish between self and foreign peptides they **haven't seen** in the thymus?

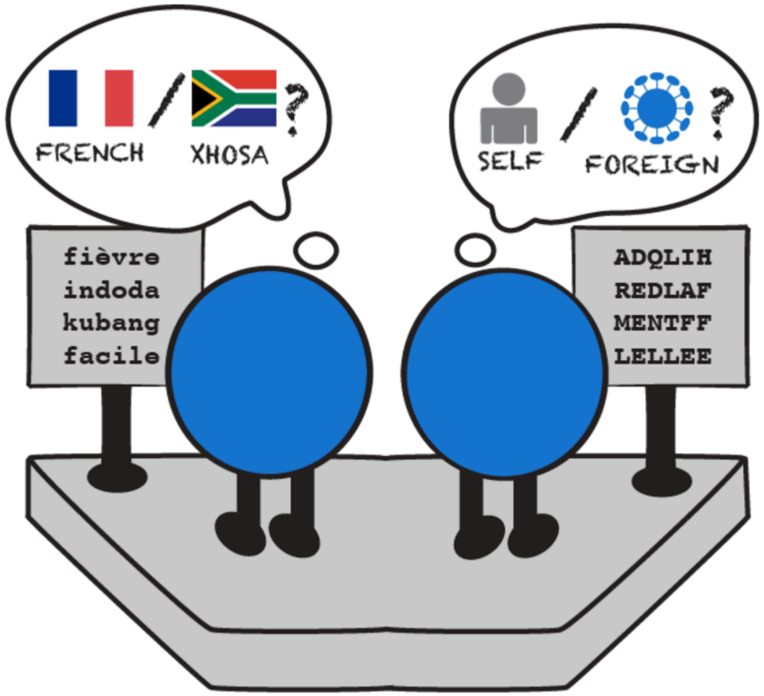
The question

Can T cells distinguish between self and foreign peptides they **haven't seen** in the thymus?

Under which conditions can T cells **"learn self by example"** during negative selection?

1. in principle?
2. in our immune system?

We used an AIS to build repertoires of T-cell receptors (TCRs) and answer this question.



A simple AIS of TCR-sequence recognition

Three ingredients needed:

1. Data (sequence)
2. Detector (TCR)
3. Matching rule (affinity)

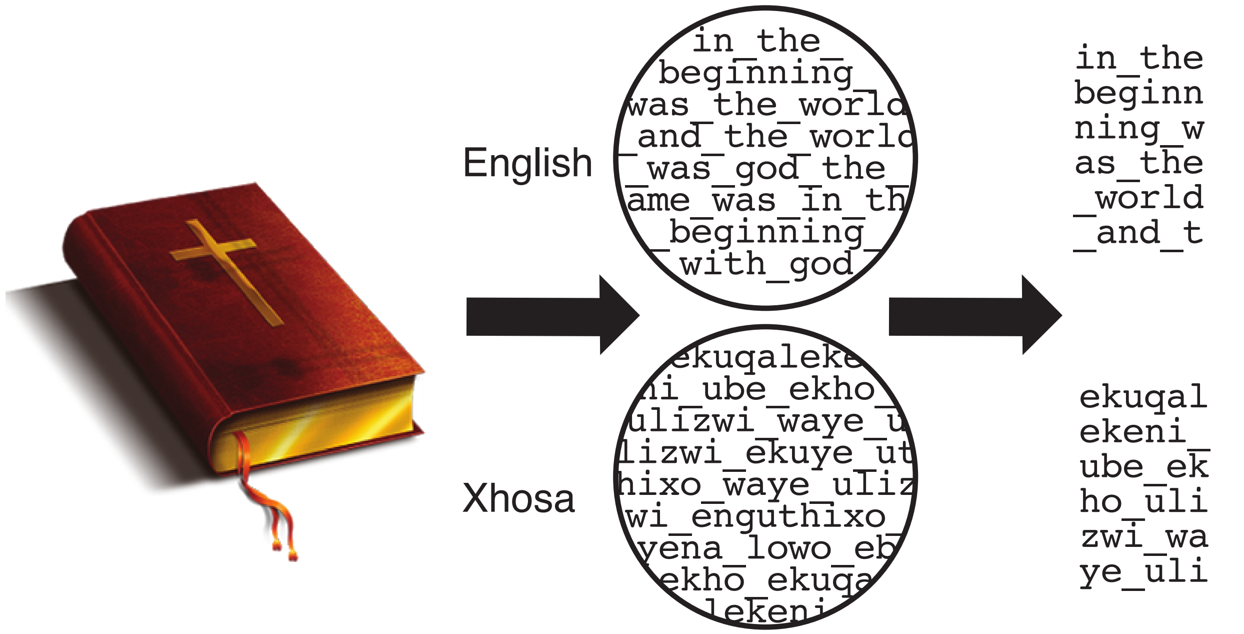
A simple AIS of TCR-sequence recognition

Three ingredients needed:

1 **Data (sequence)** 6-letter strings of text in a certain language

2 Detector (TCR)

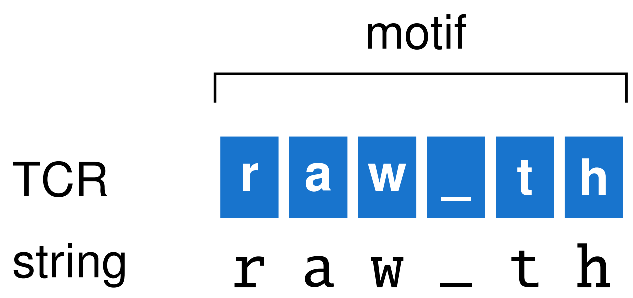
3 Matching rule (affinity)



A simple AIS of TCR-sequence recognition

Three ingredients needed:

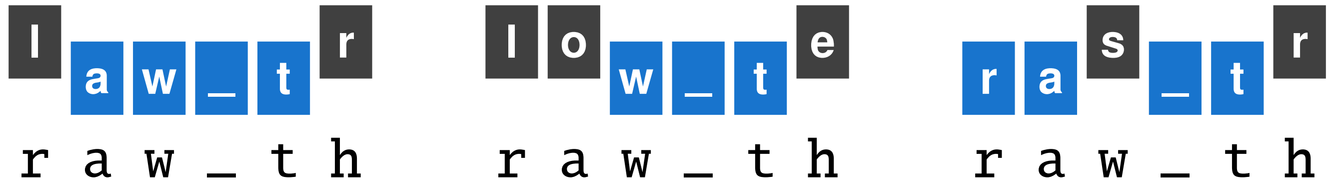
1	Data (sequence)	6-letter strings of text in a certain language
2	**Detector (TCR)**	binding motif, another string
3	Matching rule (affinity)	



A simple AIS of TCR-sequence recognition

Three ingredients needed:

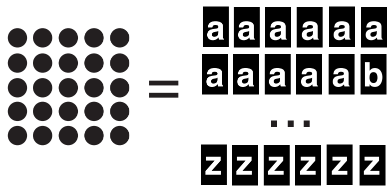
1	Data (sequence)	6-letter strings of text in a certain language
2	Detector (TCR)	binding motif, another string
3	**Matching rule (affinity)**	Longest stretch of adjacent "matches" $\rightarrow$ binding if affinity $\geq$ ***threshold t***



affinity = 4   affinity = 3   affinity = 2

Simulating negative selection *in silico*

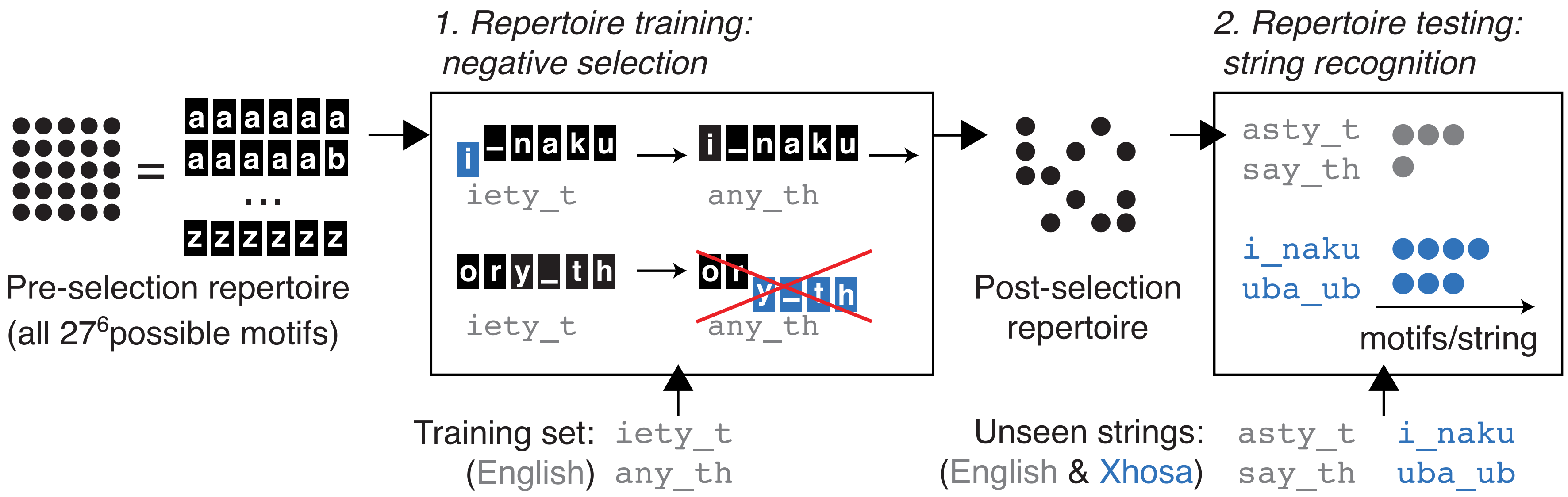
Example: Xhosa recognition after negative selection on English



Pre-selection repertoire
(all 27^6 possible motifs)

Simulating negative selection *in silico*

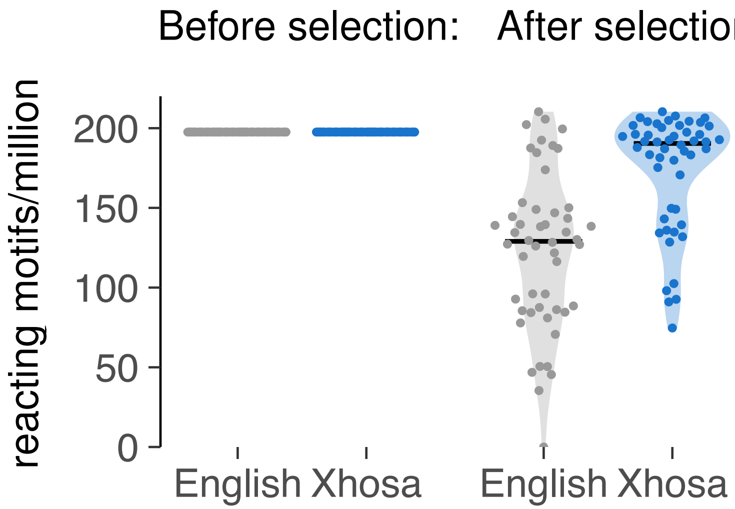
Example: Xhosa recognition after negative selection on English



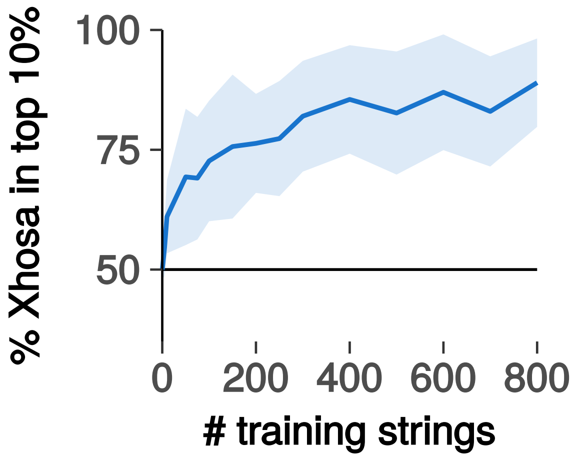
Negative selection allows discrimination

Experiment: negative selection on 500 English strings (t = 3),
Measurement: compare recognition of **different** English and Xhosa strings

Motifs per string, before vs after:



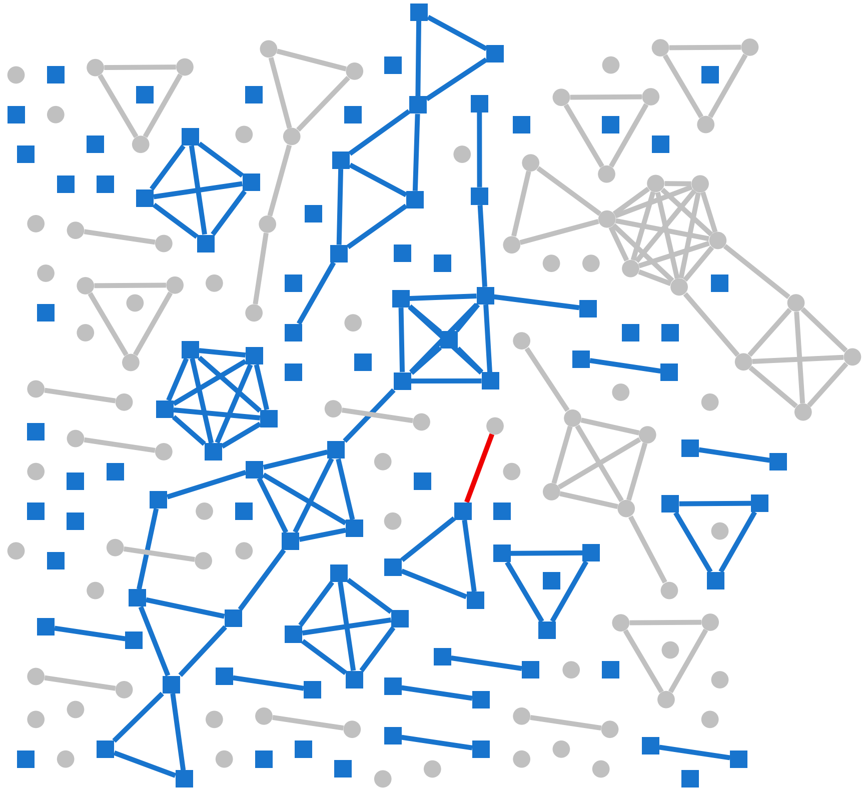
Most frequently recognized:



→ Motifs distinguish strings they **have not seen**!

Why discrimination on unseen strings?

Motifs rarely react to both English and Xhosa:



Nodes:

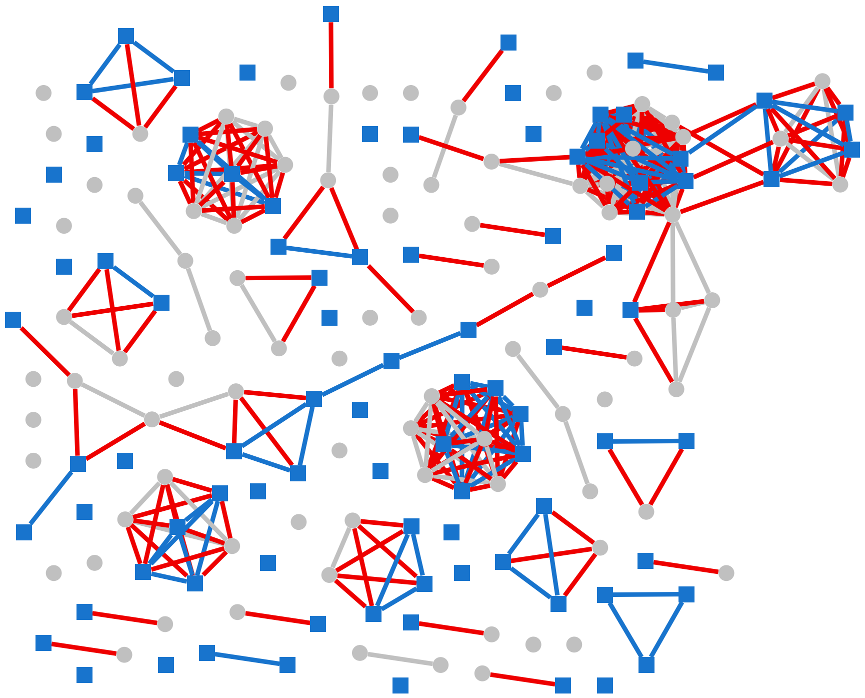
- English
- Xhosa

Edge if >10,000 motifs in common

Concordance (same-language neighbors):
81%

Why discrimination on unseen strings?

Negative control: this only works if strings are truly different:



Nodes:

- English
- More English

Edge if >10,000 motifs in common

Concordance (same-language neighbors):
50%

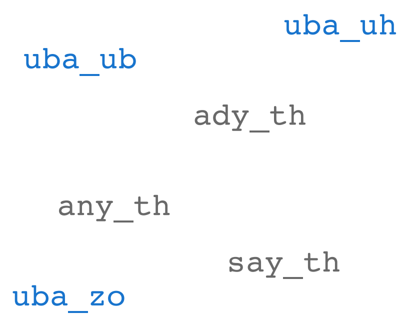
Two key requirements for learning

1. Appropriate specificity/cross-reactivity
2. Sufficient self-foreign dissimilarity

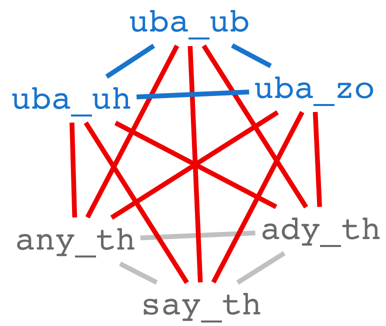
Two key requirements for learning

1. **Appropriate specificity/cross-reactivity**
2. Sufficient self-foreign dissimilarity

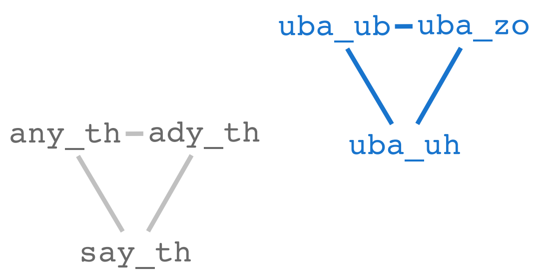
Complete specificity
(t = 6)



Low specificity
(t = 1)

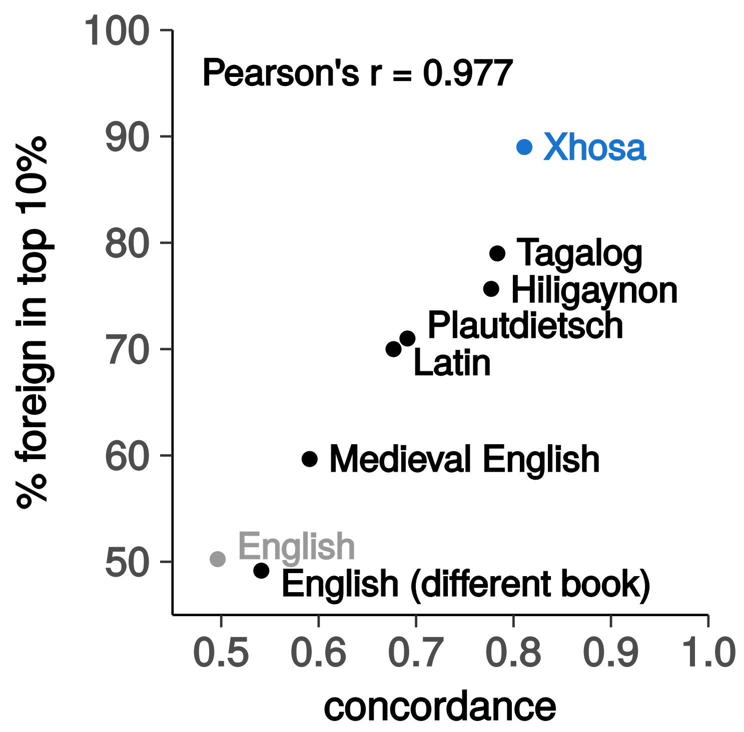


Intermediate specificity
(t = 3)

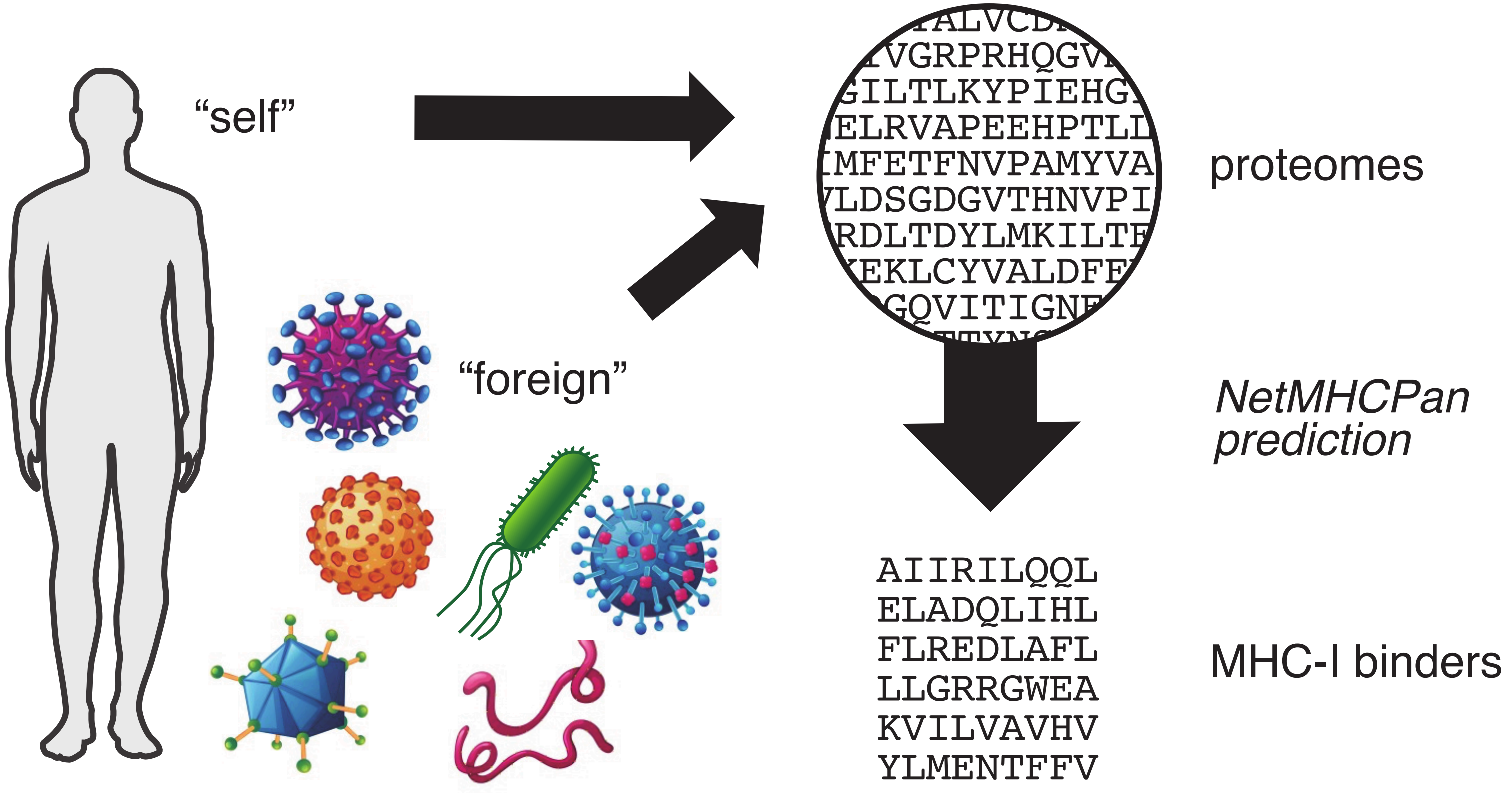


Two key requirements for learning

1. Appropriate specificity/cross-reactivity
2. **Sufficient self-foreign dissimilarity**



Back to self versus foreign peptides



Back to self versus foreign peptides

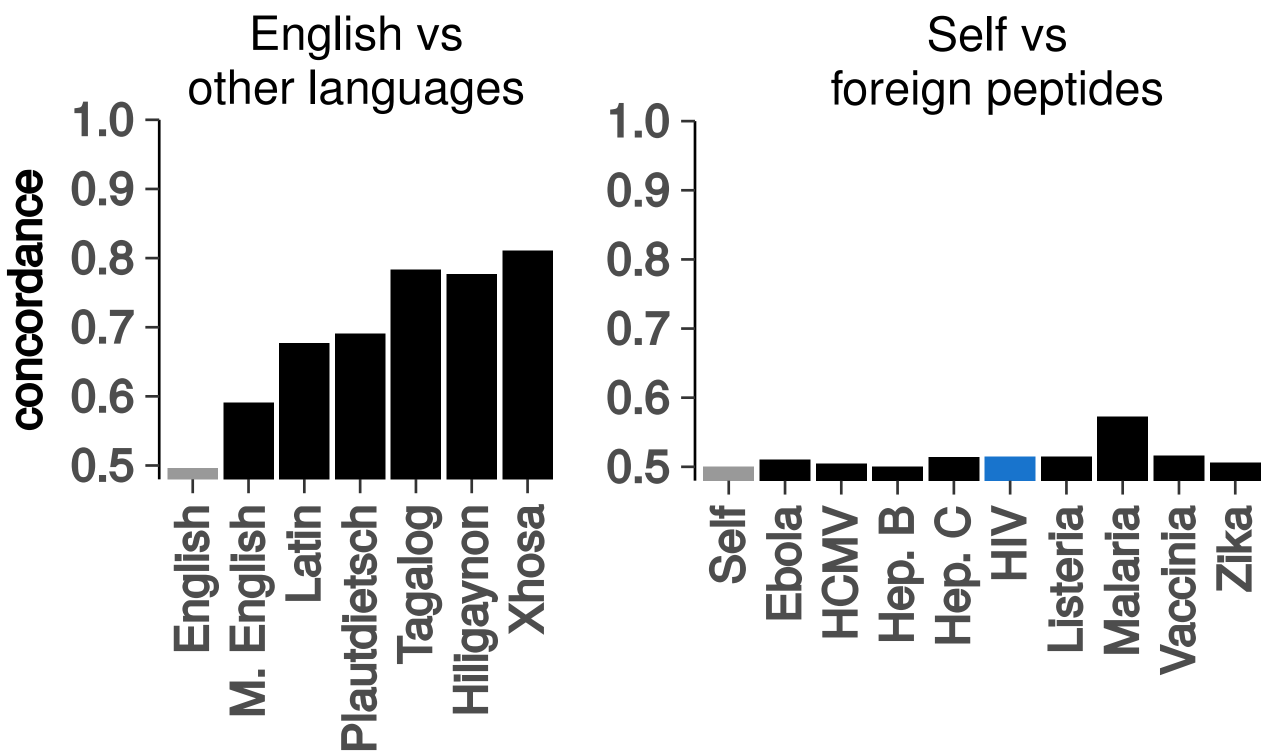
Use the same model for peptides instead of strings:

MHC anchors

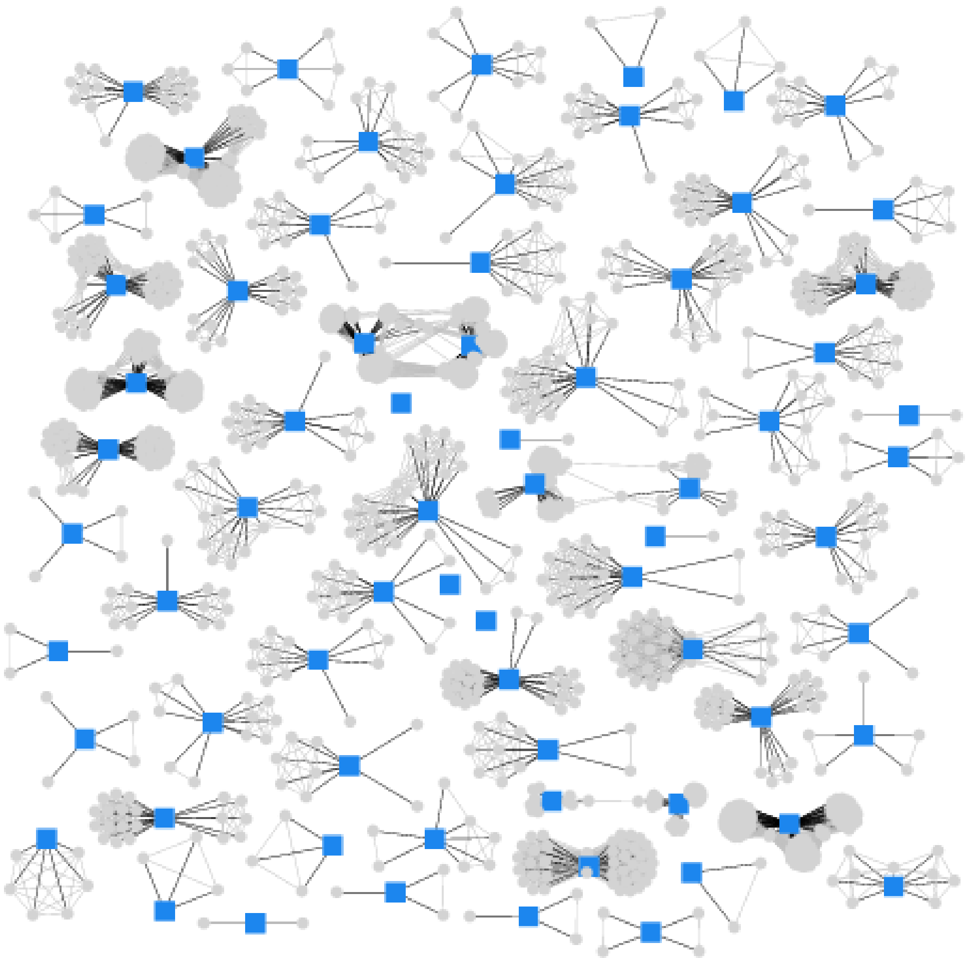
Rarely affects
TCR binding

TCR binding
residues

"Self" and "foreign" peptides are similar



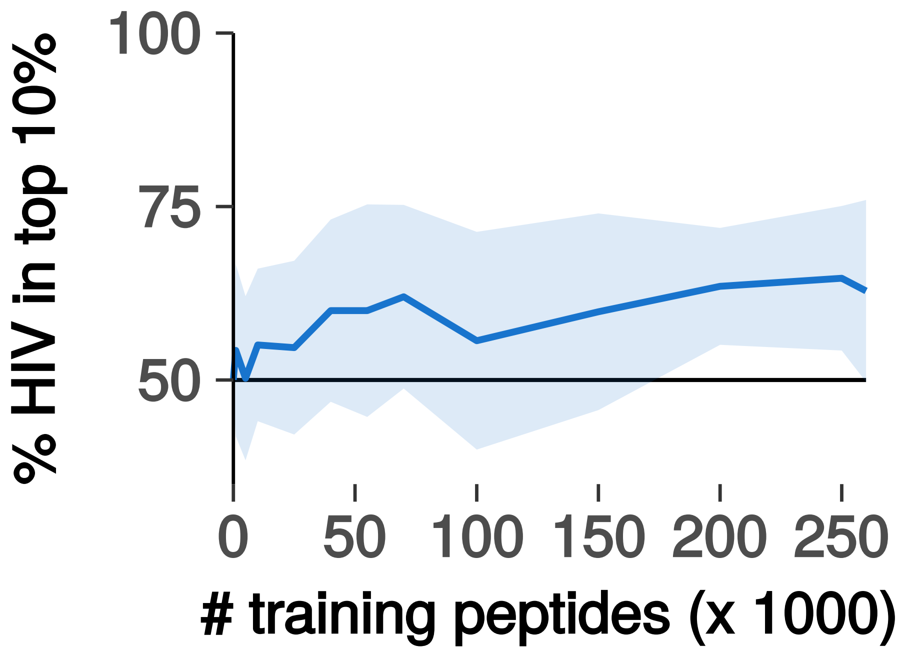
"Self" and "foreign" peptides are similar



HIV peptides are embedded in clusters of self peptides.

Self-foreign discrimination is difficult

... but remains (somewhat) possible:

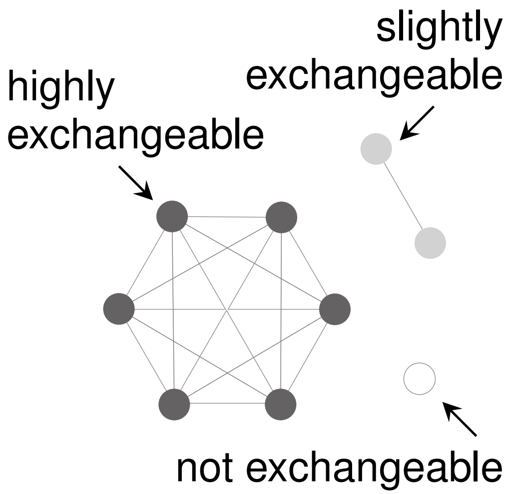


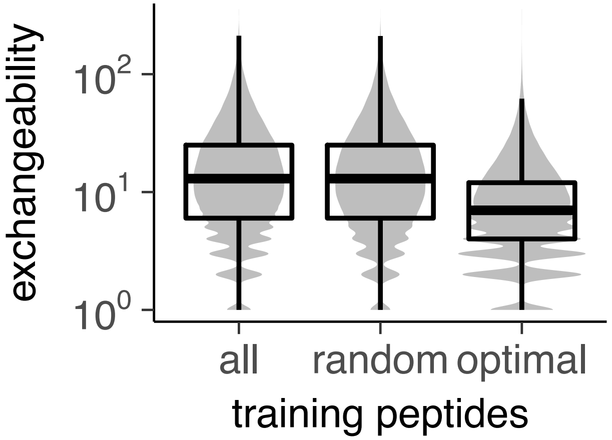
What if thymic self peptides are non-random?

Random self peptides are often similar and delete the same TCRs. This is inefficient. Consider strategies for our French game:

- inefficient: learn fièvre, chèvre, nièvre, mièvre, lièvre, ...
- efficient: learn chèvre, oiseau, cousin, entrée, boiser, ...

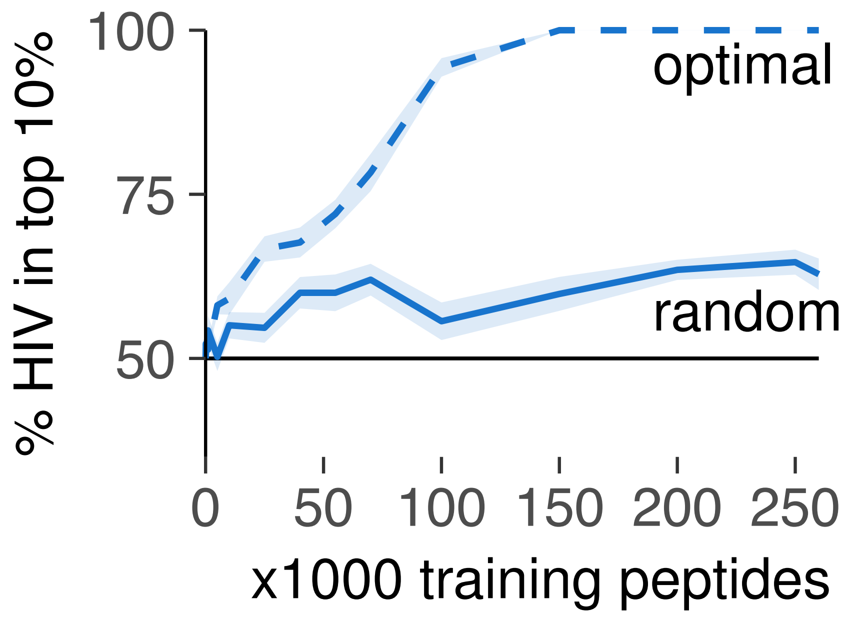
Some peptides are more **exchangeable** than others.
$\longrightarrow$ a training set with more non-exchangeable peptides might do better:





"Optimal" training peptides improve self-foreign discrimination

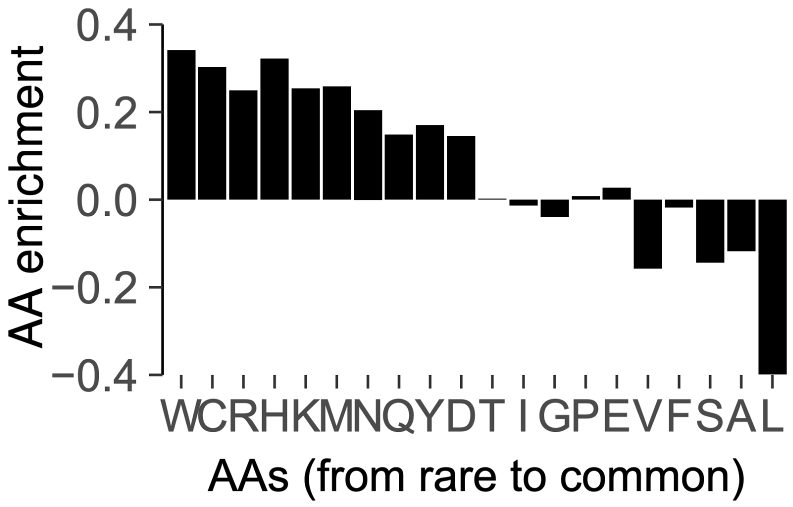
Based on this logic, we define an "optimal" training set of size N as a combination of N self peptides that together "cover" as many distinct TCRs/detectors as possible*.

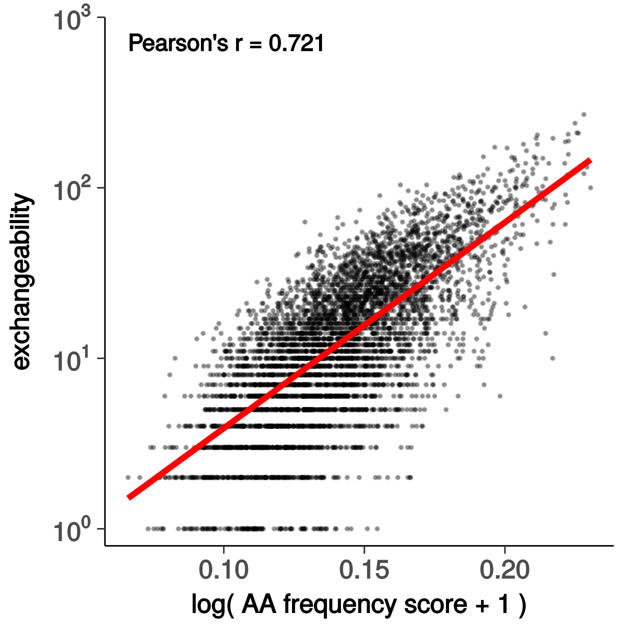


* This combinatorial 'set-cover' optimisation problem is NP-hard, so we approximated a solution with a greedy algorithm.

How could the thymus be "optimal"?

Computed "optimal" set is enriched in rare AAs, depleted of common AAs.
Peptides with rare AAs tend to be less exchangeable:

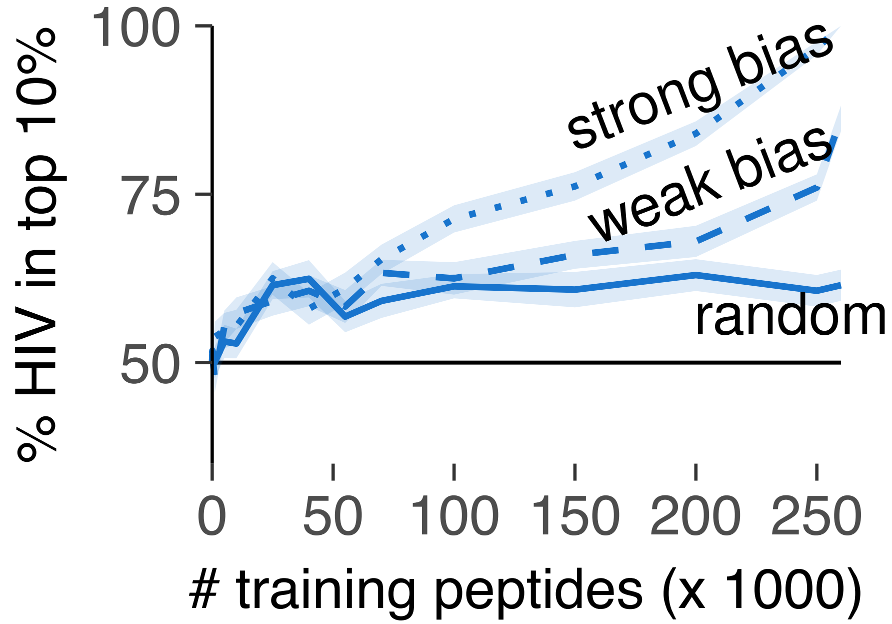




→ Could enrichment of rare AAs help self-foreign discrimination?

How could the thymus be "optimal"?

Choose "training" self peptides with a **bias** for peptides with rare AAs:



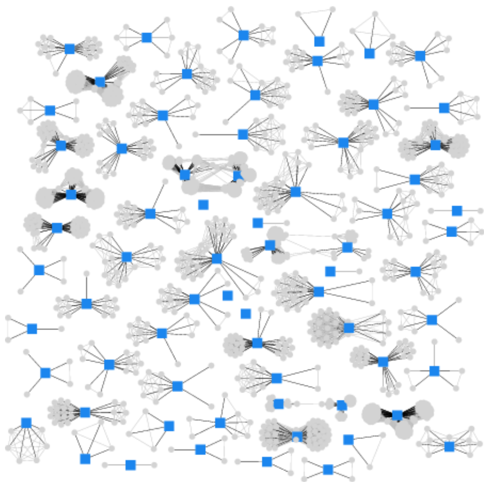
Simple bias already improves self-foreign discrimination.

Summary: self-foreign discrimination in the IS

1. T-cell negative selection is a learning algorithm.
In principle, T cells can "learn self by example" in the thymus.

2. Biological T cells may be learning the wrong thing.
In practice, this is not easy because the "language" of proteins depends more on function than organism.

3. A "smart" thymus may help T cells learn.
Do such biases in peptide presentation pathways truly exist?



Read more

Johannes Textor
Radboud University

Can Keşmir
Utrecht University

Rob de Boer
Utrecht University

Judith Mandl
McGill University, Montréal

Paper:
IMN Wortel, C Keşmir, RJ de Boer, JN Mandl, J Textor (2020). **Is T-cell negative selection a learning algorithm?** *Cells* 9(3), 690; doi:10.3390/cells9030690

Blog post:
What if T cells could learn French? *The Startup on Medium*; https://medium.com/swlh/what-if-t-cells-could-learn-french-8301f852254f?sk=c0dd70ee935cedefb8d35c670392d268