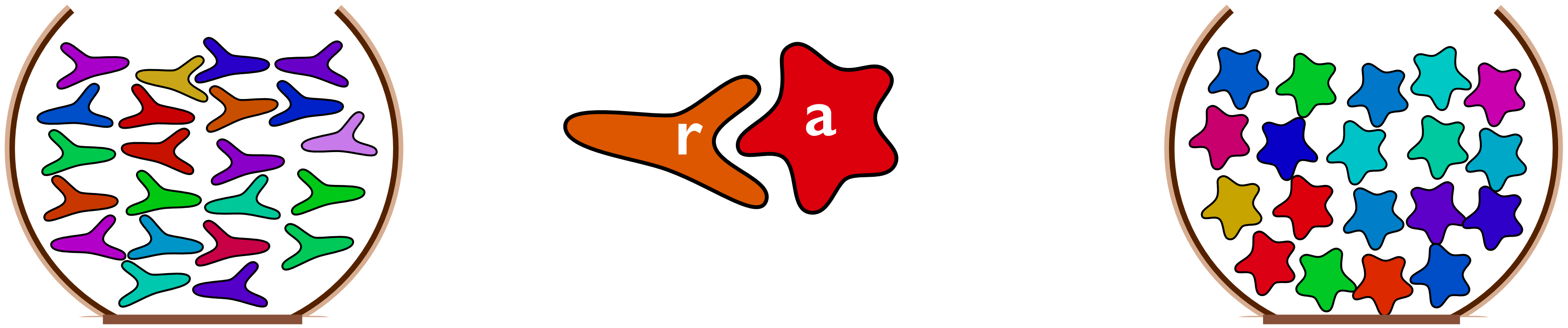


Analysis of immune receptor repertoires

Aleksandra M Walczak

Laboratoire de Physique Théorique - ENS, CNRS

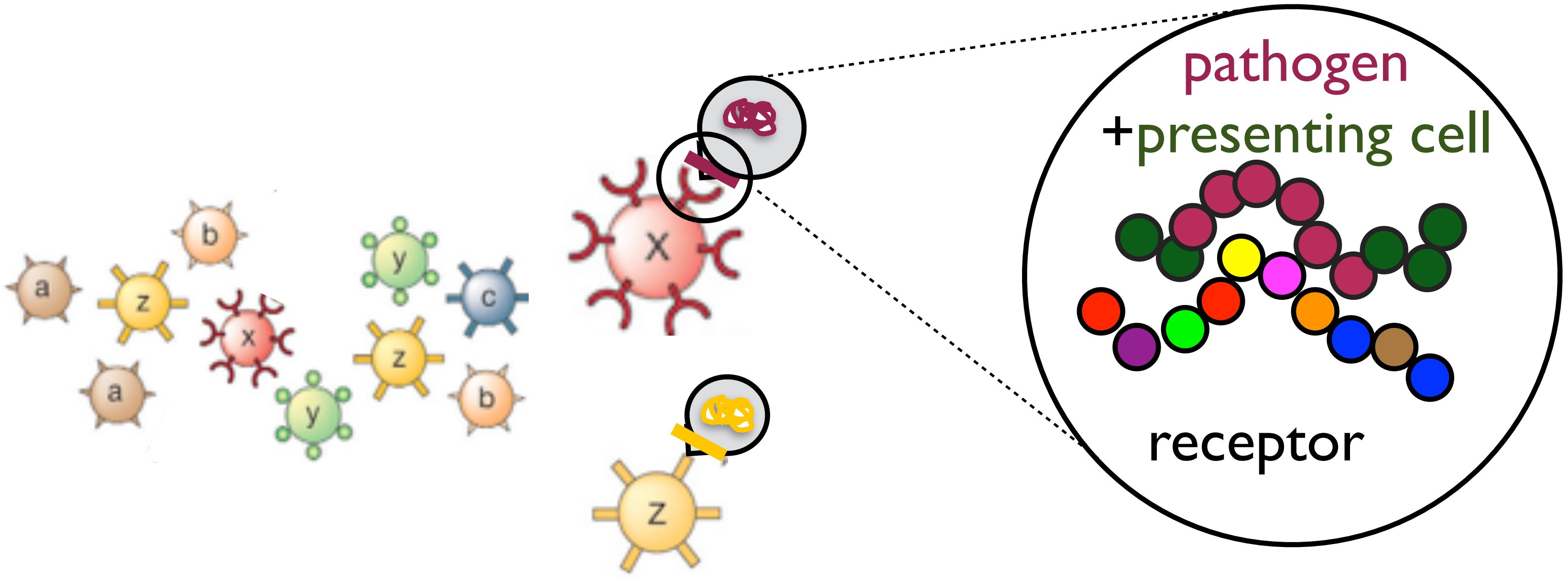
work with
Rhys Adams, ENS
Yuval Elhanati, ENS
Quentin Marcou, ENS
Maximilian Puelma Touzel, ENS
Zachary Sethna, Princeton
Justin Kinney, CSHL
Anand Murugan, Princeton
Curt Callan, Princeton
Thierry Mora, ENS

European Research Council
Established by the European Commission

and
Misha Pogorely, RAS Moscow
Dmitriy Chudakov, RAS Moscow
Ilgar Mamedov, RAS Moscow
Yury Lebedev, RAS Moscow

Immune receptors

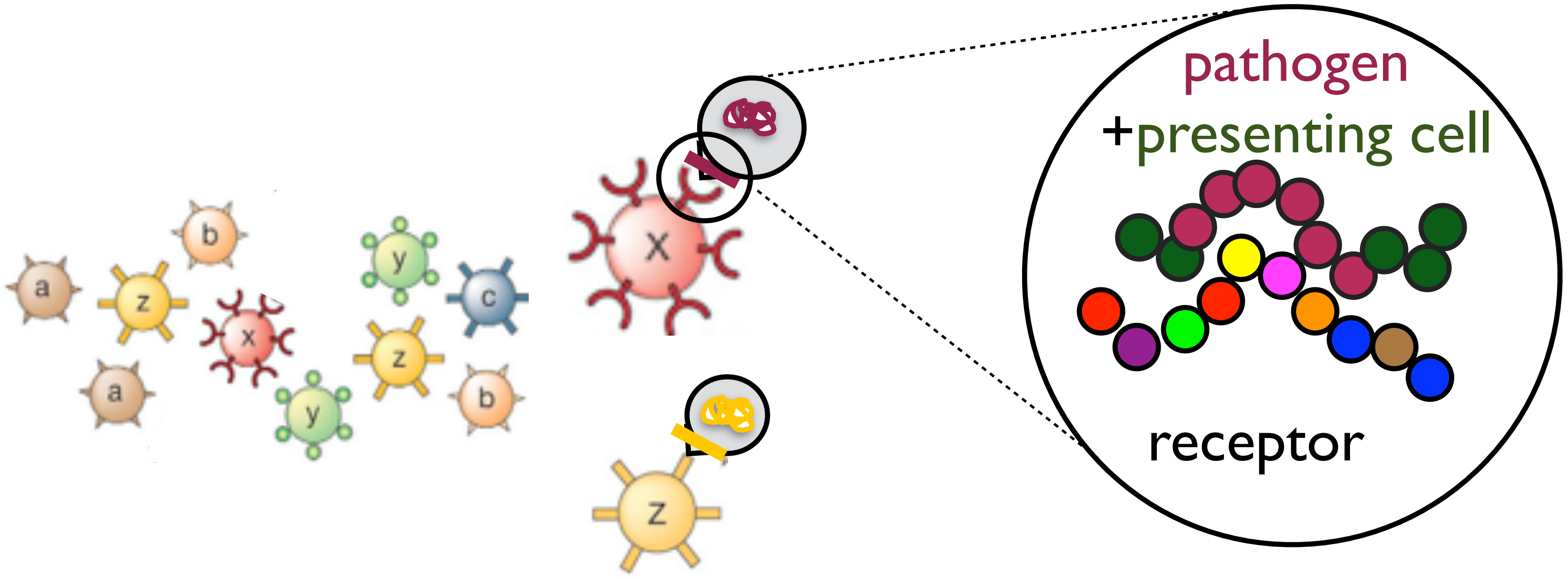
- T-cells important actors of immune system



- triggers immune response

Immune receptors

- T-cells important actors of immune system

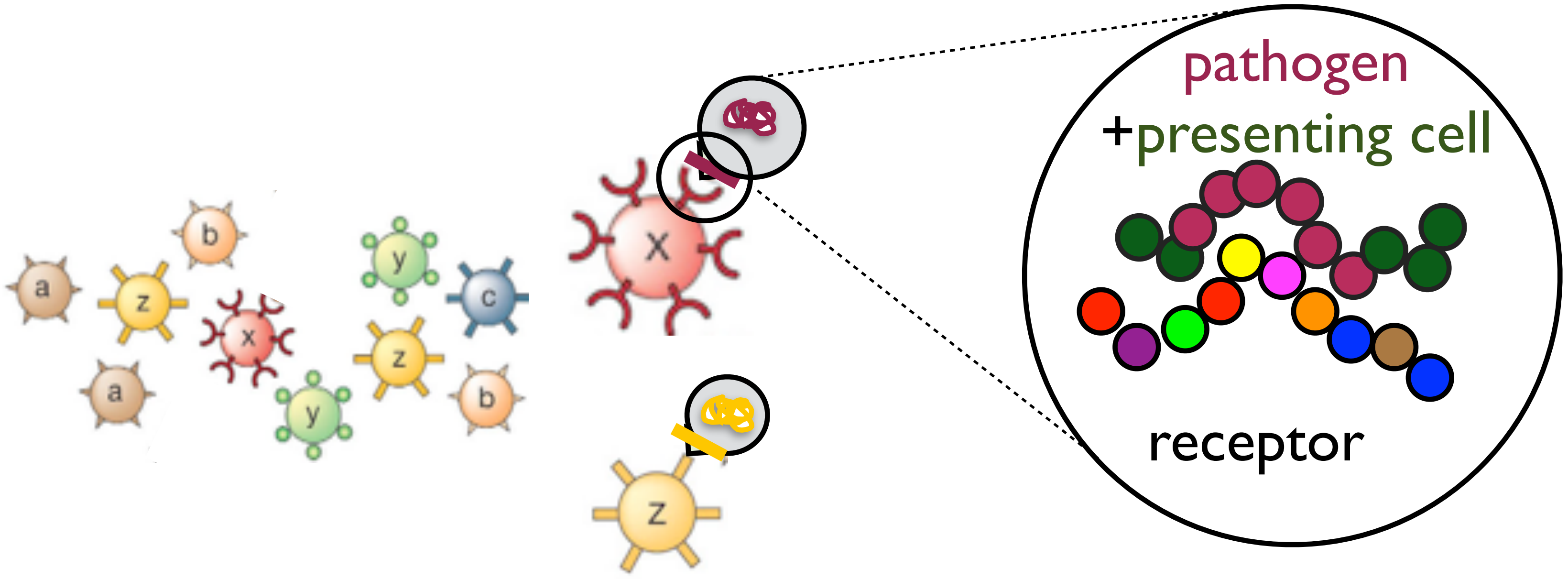


many unique receptors ↔ many different pathogens (virus...)

- triggers immune response

Immune receptors

- T-cells important actors of immune system



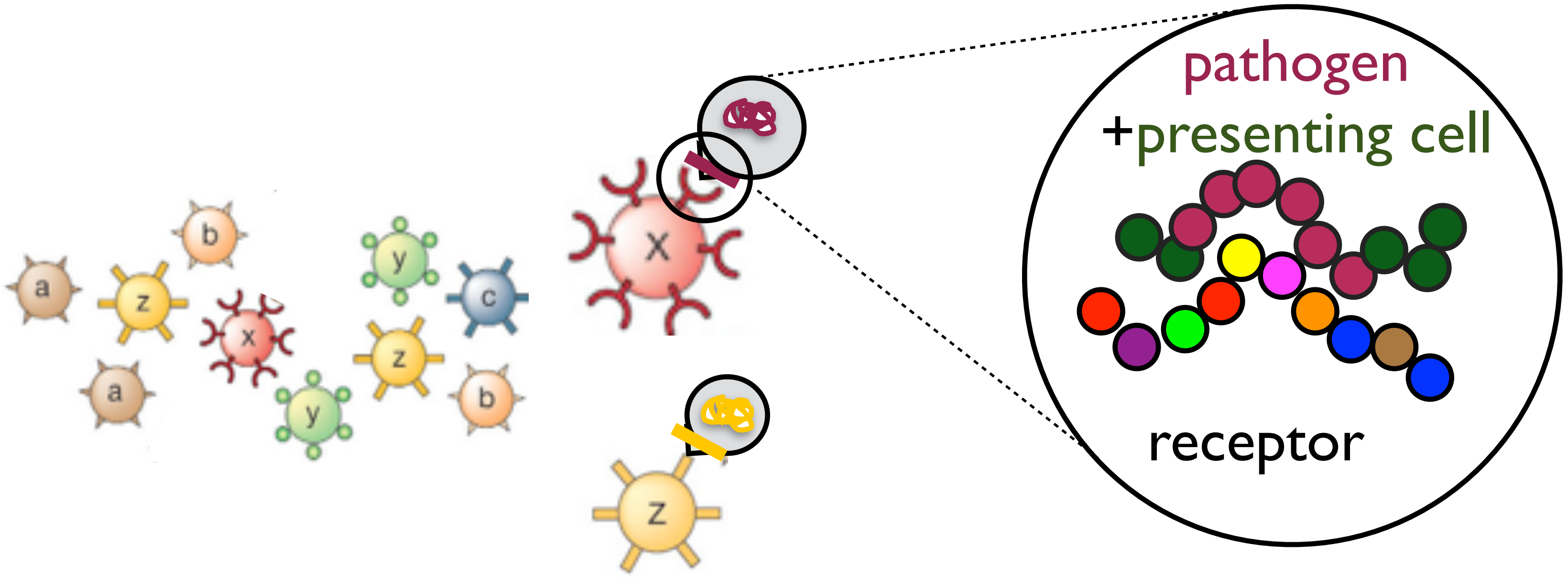
many unique receptors ↔ many different pathogens (virus...)

- triggers immune response

natural, healthy (="normal") diversity of immune receptors?

Immune receptors

- T-cells important actors of immune system



many unique receptors ↔ many different pathogens (virus...)

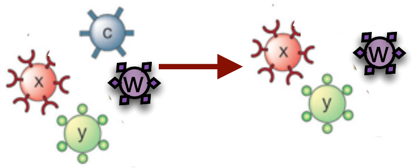
- triggers immune response

natural, healthy (="normal") diversity of immune receptors?

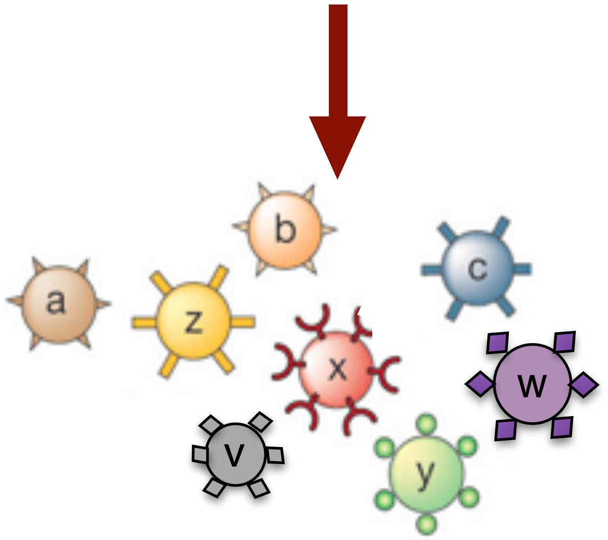
how diverse is it?

predicting repertoires ?

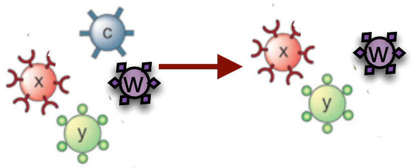
Repertoire evolution



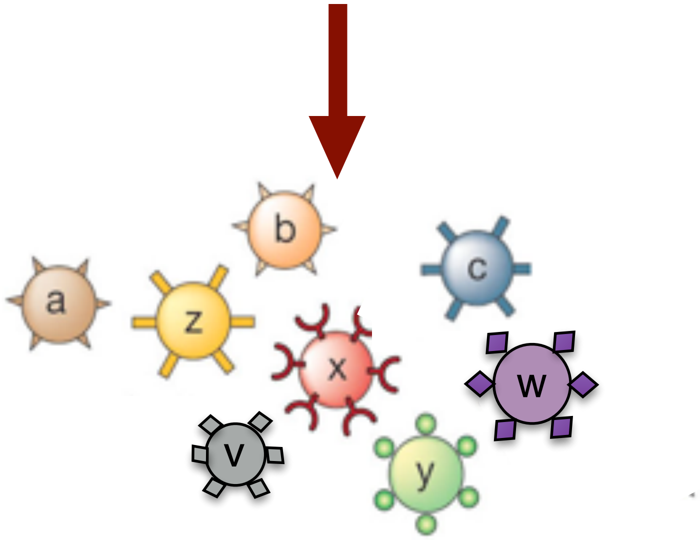
RECEPTOR GENERATION



Repertoire evolution

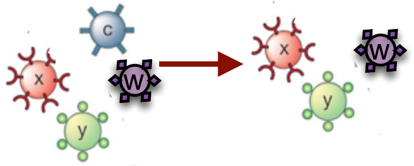


RECEPTOR GENERATION

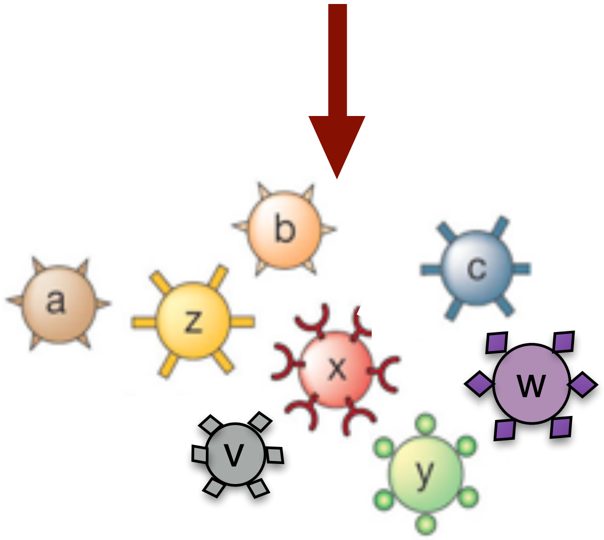


combinatorics + *randomness* → diversity

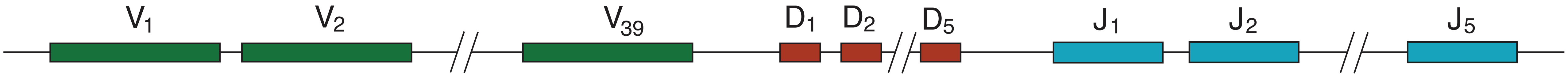
Repertoire evolution

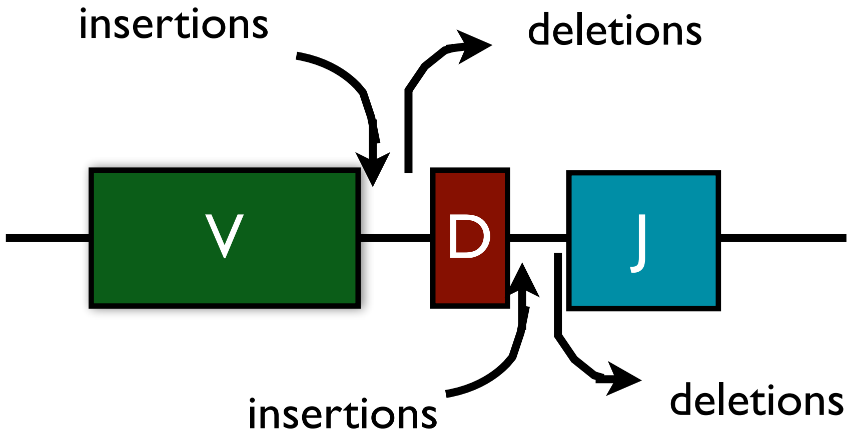


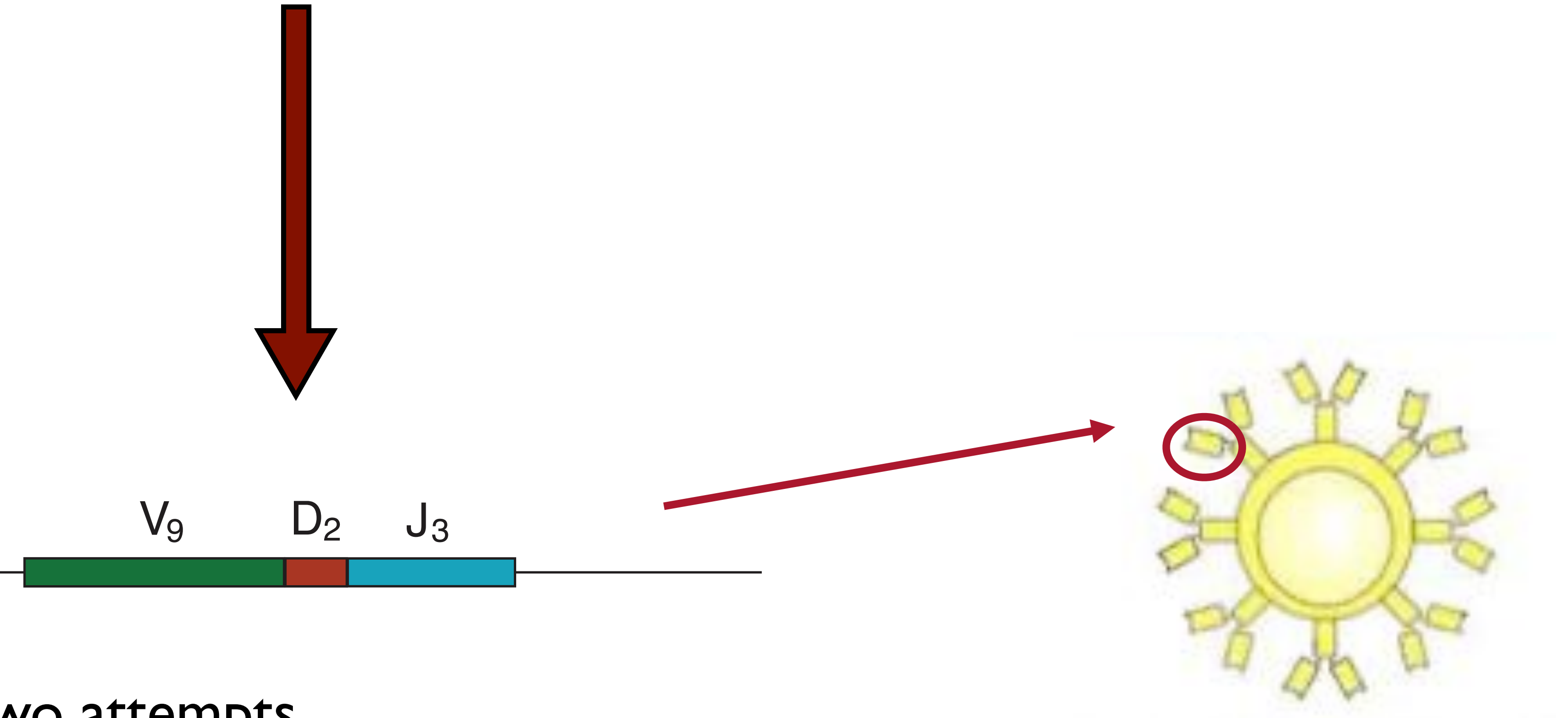
RECEPTOR GENERATION



combinatorics + *randomness* → diversity



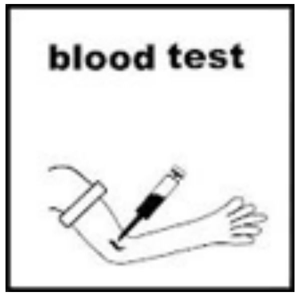




- two chromosomes per cell = two attempts

Sequence data

new data

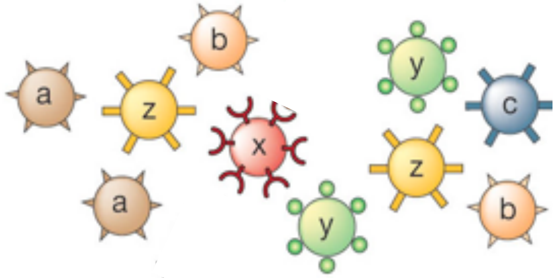


+

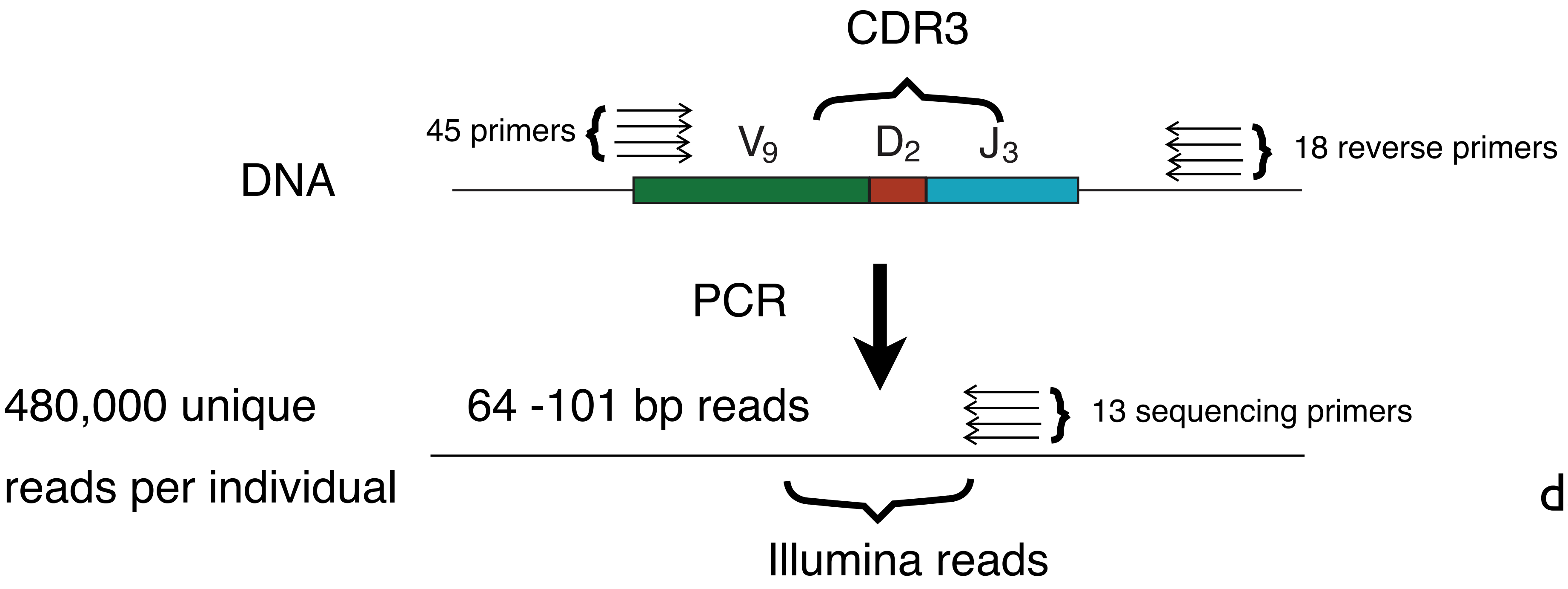
sequencing machine

=

natural diversity distribution



- human T-cell beta chain receptor sequences
- 9 people
- out of frame reads (~14% of each type of cells) ⟶ *generation*
- in frame reads (~235,000 unique reads) ⟶ *selection*

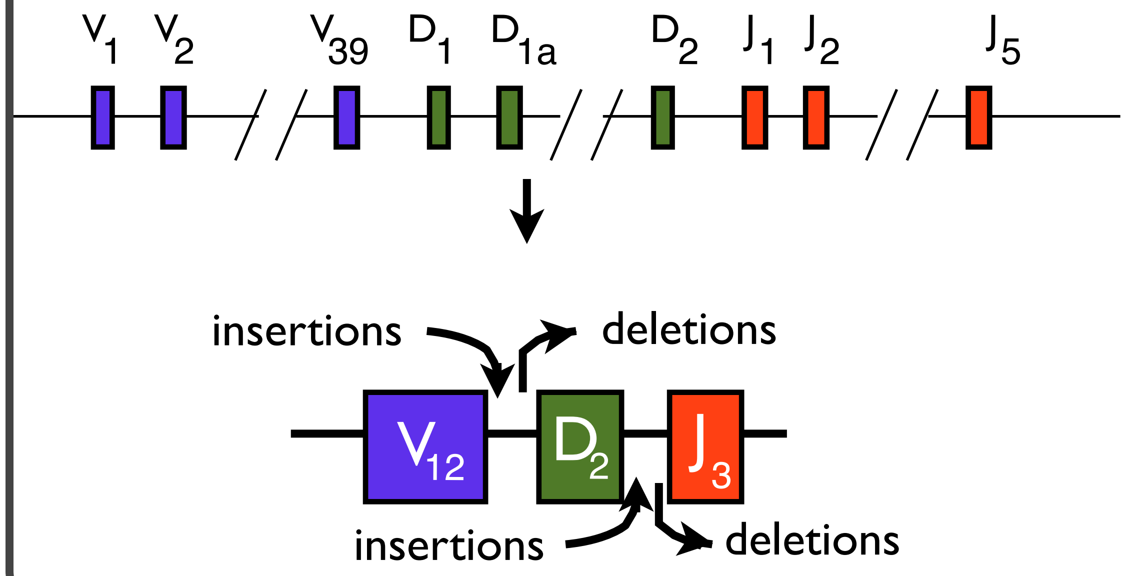


data from Robins lab and Chudakov lab

Probabilistic VDJ recombination annotation

sequence generation:

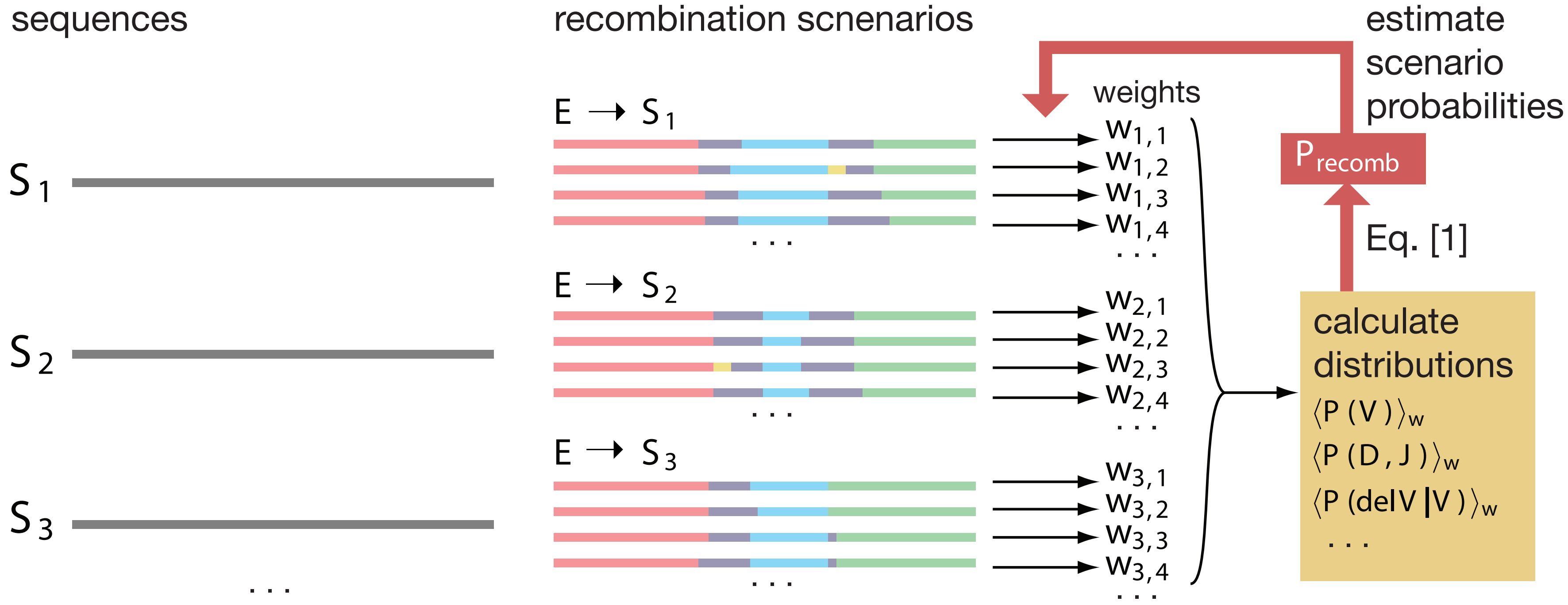
combinatorics + *randomness* → diversity



$\vec{\sigma}$ - receptor DNA sequence

```
TAGGACCTCGGAAACCTCTTCTGCTGGCGCAGAGATACAGGGCTGAGTCTTCTT
TATCTGCCGATGTCGCGAAGGCCCTCCCGCTAAGATCACTGGTGGCACAGAAGT
TAGGGAGAGGTGCCTAACTGCTGGCACAGAAGTACAGAGAGGTCTGGTTGGGGT
TCCGCCGCTAGTCCCTGAAACTACTGGCACAGAGATAGAAAGCTGTCGGGTTCT
CGAAACTGCTGGCACAGAAGTACACAGATGTTTGGGAGGGAGCAGCCGACTCCA
TCTTGGCCGCTAGTCCGAGAAACTGCTGGCACAGAAGTACACAGATGTTTGGGA
TTGTAGGAGCCGGACCGGCCCCCTGTCCCCTTGGCTGCTGGCGCAGAGATACAG
TCATTTAACGTGCGGCCCGCCTGGCACAGAAGTAAAGAGCTGTCTGGTTGTGGT
TAGTAACTCCGCTTCACTGCTGGCACAGAAGTACACAGATGTCTGGGAGGGAGC
TCCCTCCGGTTTGAAGGGTCTGCTGGCACAGAAGTACACAGATGTTTGGGAGGG
CCGGTGTTCGCACAGCCCTGGGGACCCTGGCGCAAACCCCGCTTCCCTCGAGGA
```

- probabilistic assignment of:

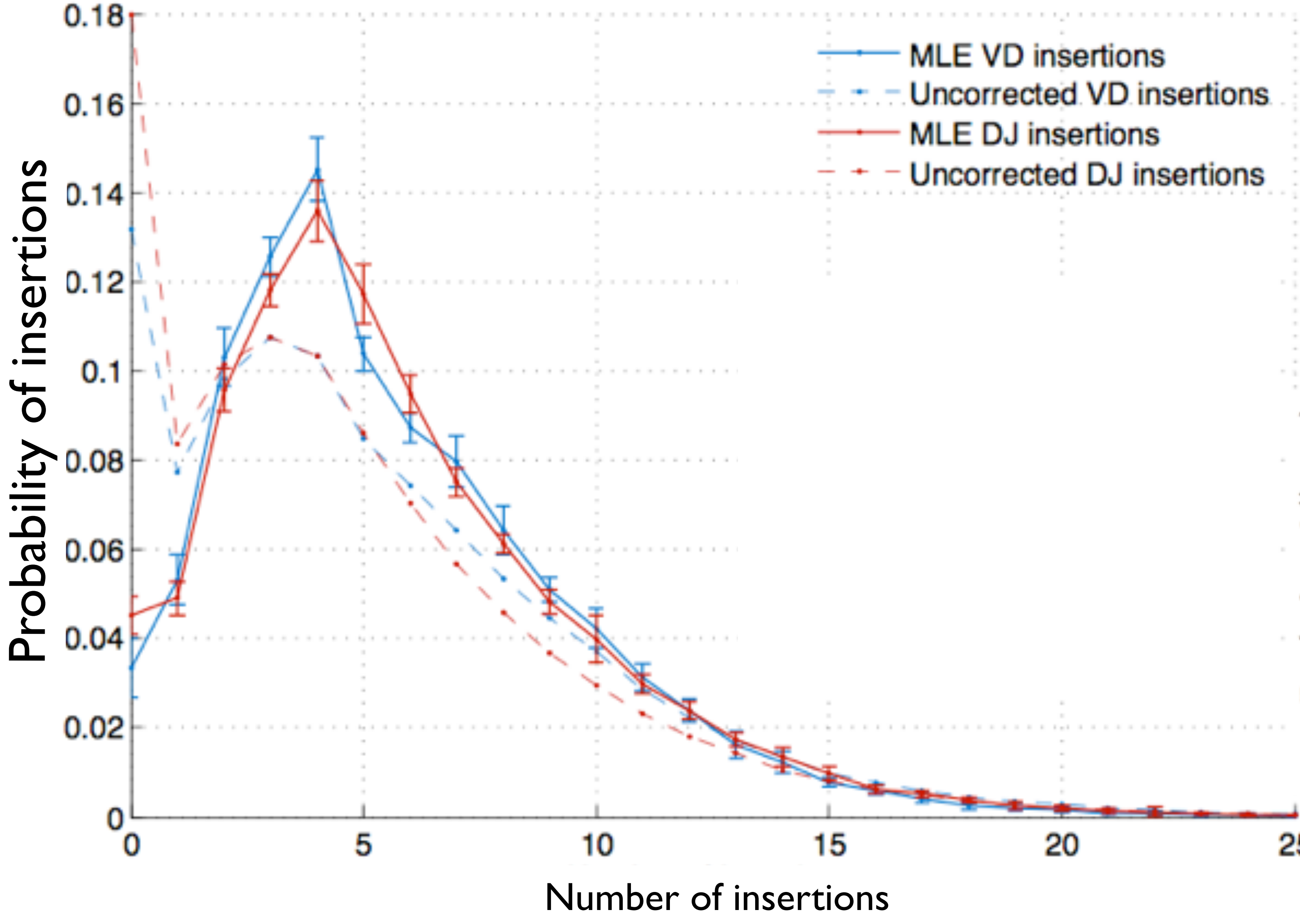


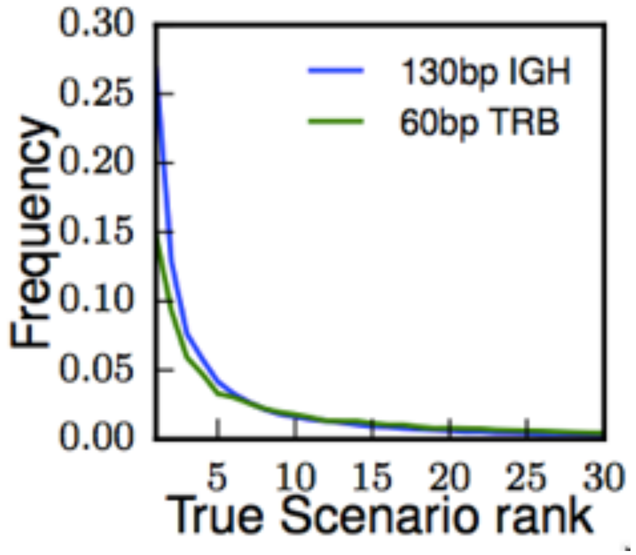
$$P^{\mathrm{recomb}}(\mathrm{scenario}) = P(V)P(D,J)P(\mathrm{deletions}V|V)P(\mathrm{insertions}DJ)... \qquad [1]$$

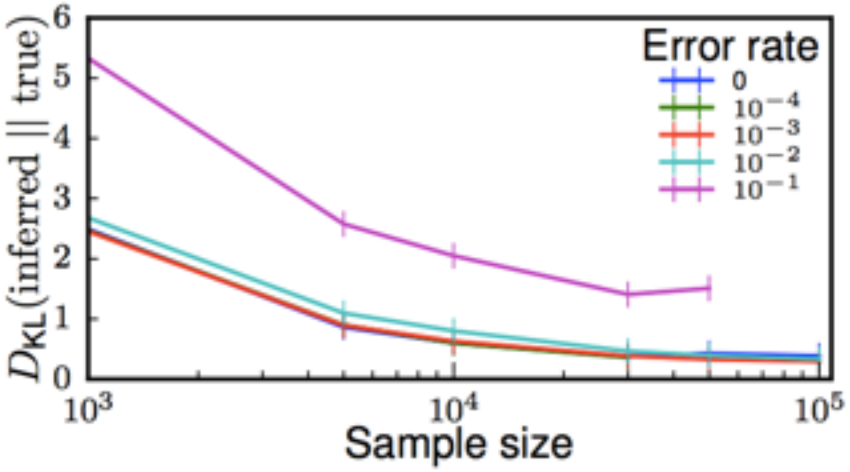
Universal insertion profiles

→ universal mechanism for generation

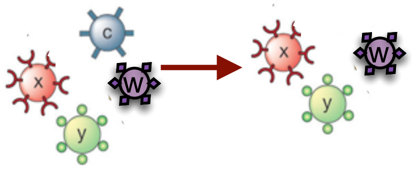
- VD and DJ insertion profiles are identical
- real scenario most likely in less than 30% of cases
- need < 2 ml of blood to learn model

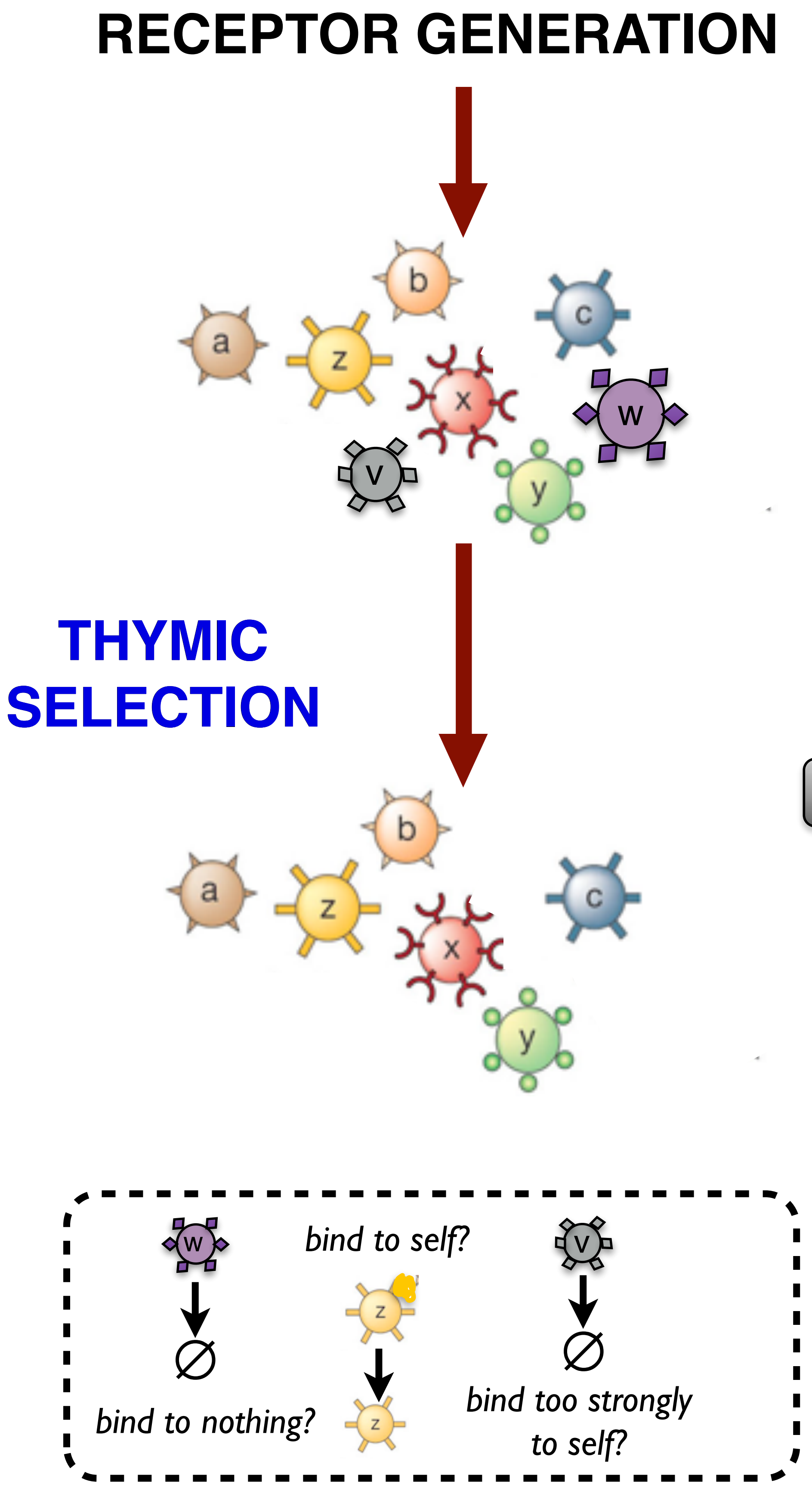






Next step: selection





RECEPTOR GENERATION

THYMIC SELECTION

bind to self?

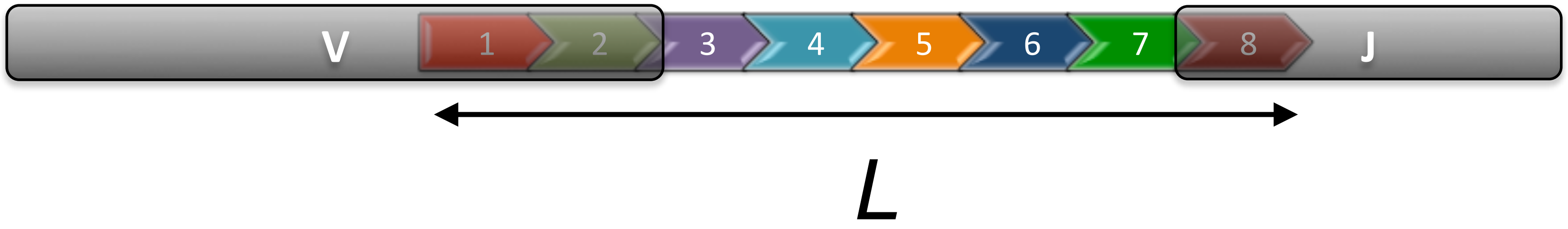
bind to nothing?

bind too strongly to self?

- quantify using selection factors

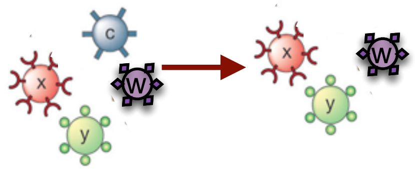
$$Q(\{\sigma\}) = \frac{P_{\text{post-sel}}(\{\sigma\})}{P_{\text{gen}}(\{\sigma\})}$$

- a model for the observed probabilities

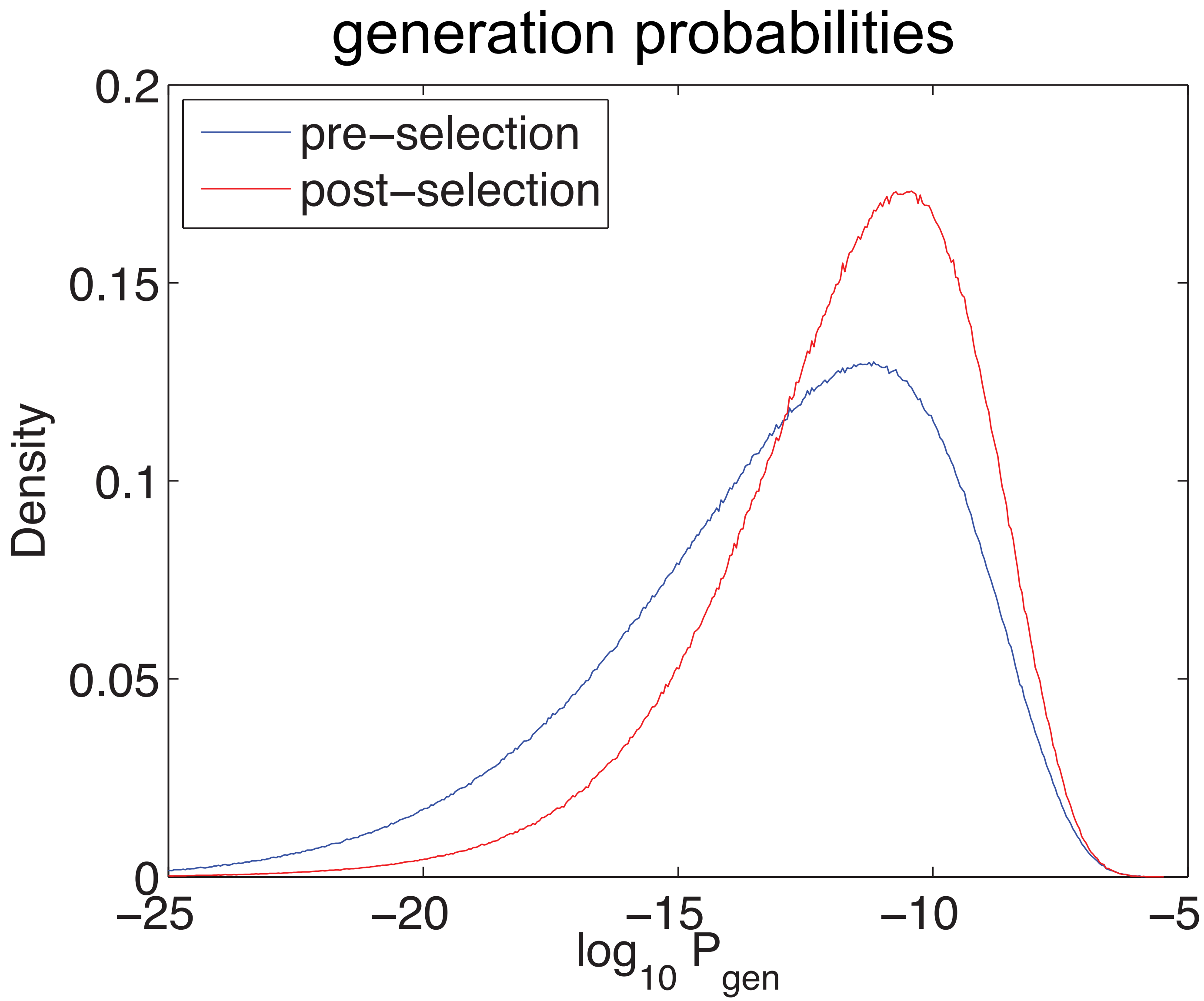


$$Q(\vec{\sigma}, V, J) = \frac{1}{Z}\, q_L \; q_{V,J|L} \prod_{i=1}^{L} q_{i;L}(a_i)$$

Natural selection anticipates somatic selection

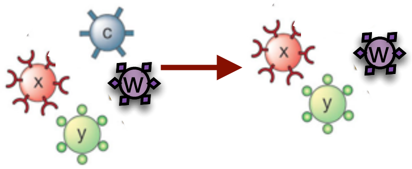


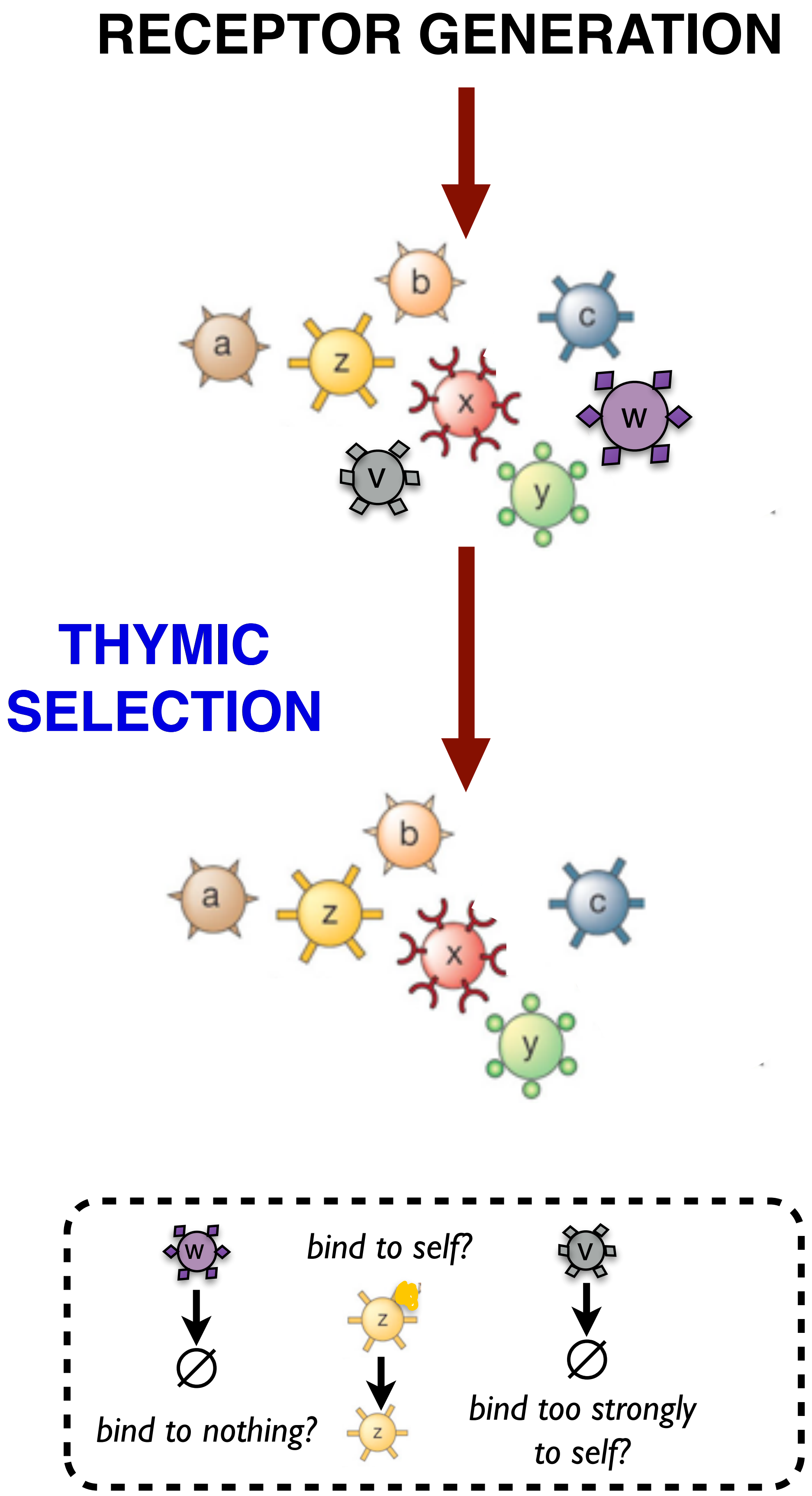
- sequences more likely to be *generated* → more likely to be *selected*



- true for all individuals independently

Receptor sharing





RECEPTOR GENERATION

THYMIC SELECTION

bind to nothing?

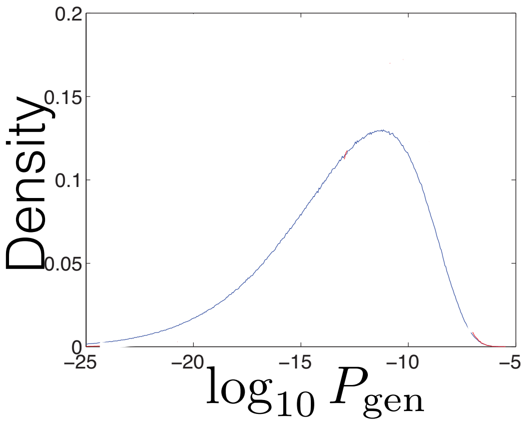
bind to self?

bind too strongly to self?

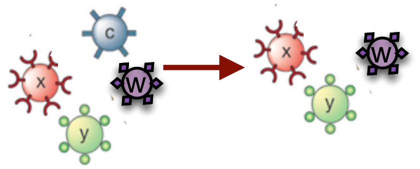
- quantify using selection factors

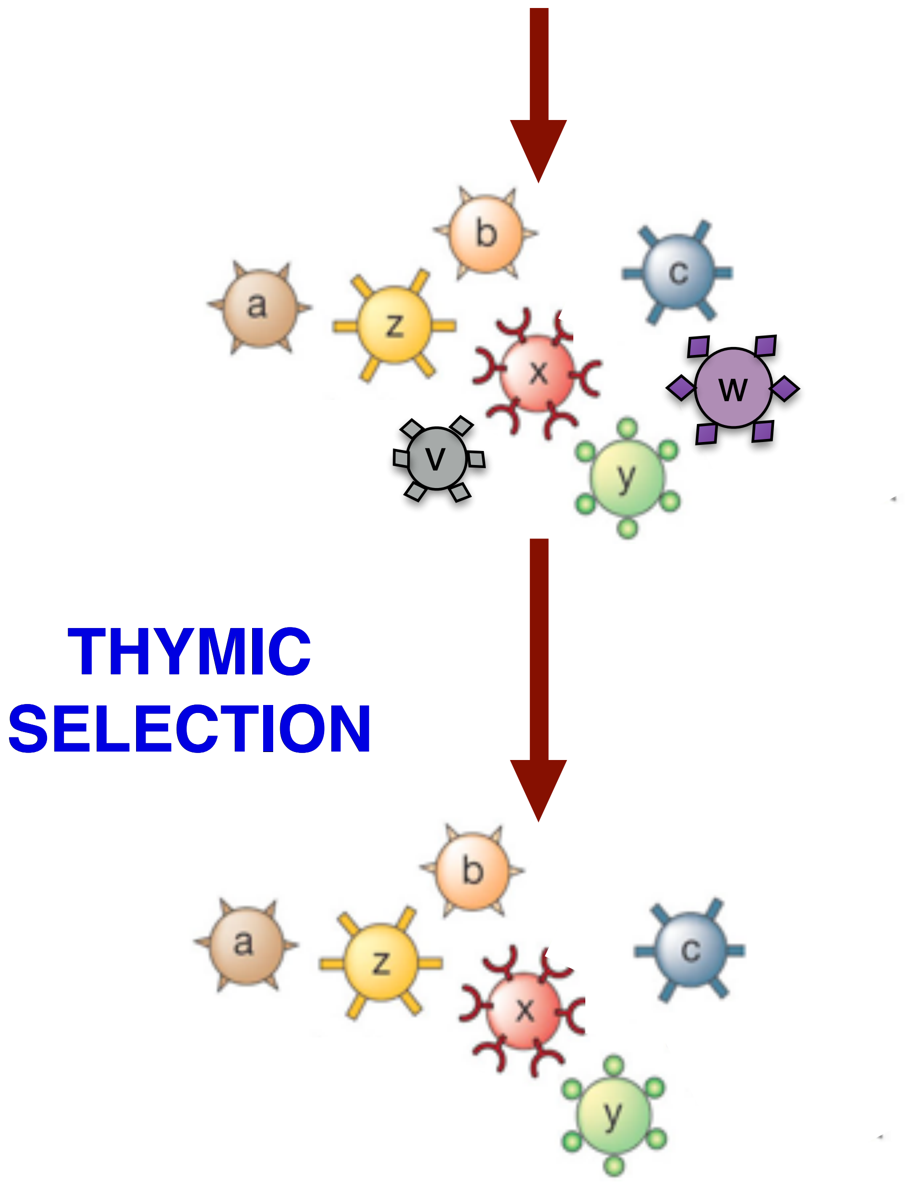
$$Q(\{\sigma\}) = \frac{P_{\text{post-sel}}(\{\sigma\})}{P_{\text{gen}}(\{\sigma\})}$$

human beta TCR data from Robins lab



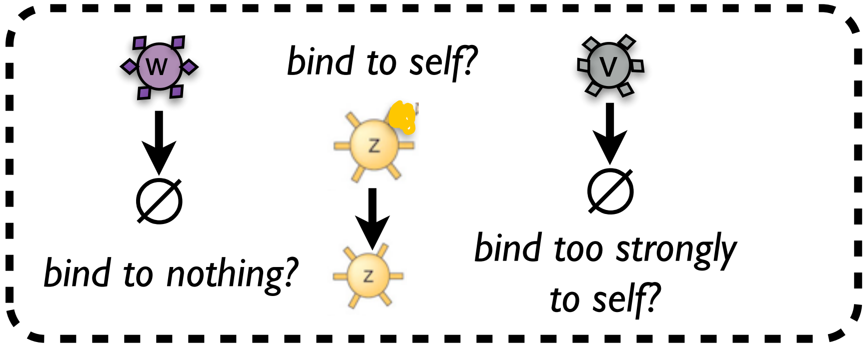
Receptor sharing





RECEPTOR GENERATION

THYMIC SELECTION



bind to self?

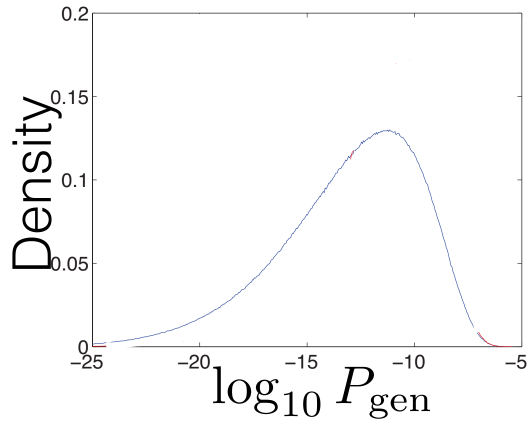
bind to nothing?

bind too strongly to self?

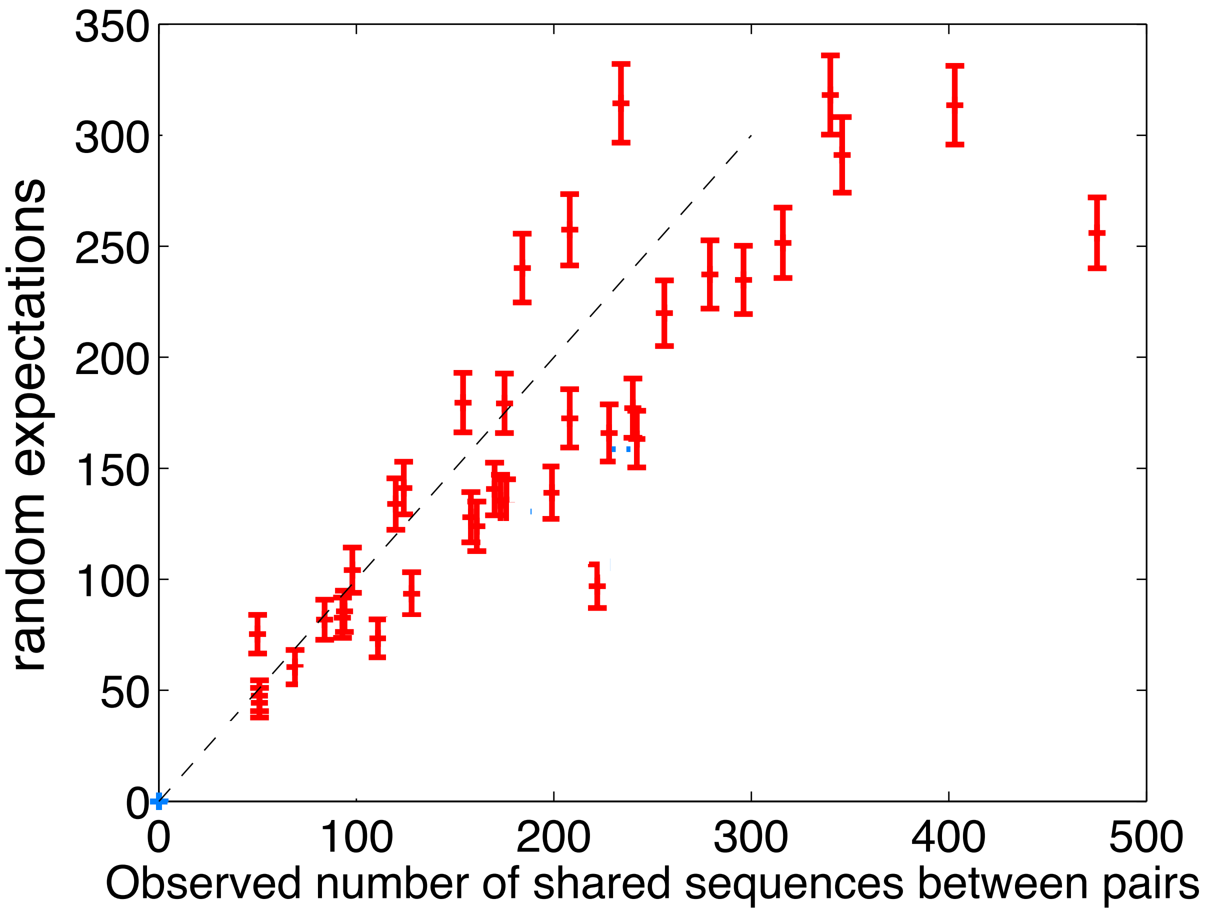
- quantify using selection factors

$$Q(\{\sigma\}) = \frac{P_{\mathrm{post-sel}}(\{\sigma\})}{P_{\mathrm{gen}}(\{\sigma\})}$$

human beta TCR data from Robins lab



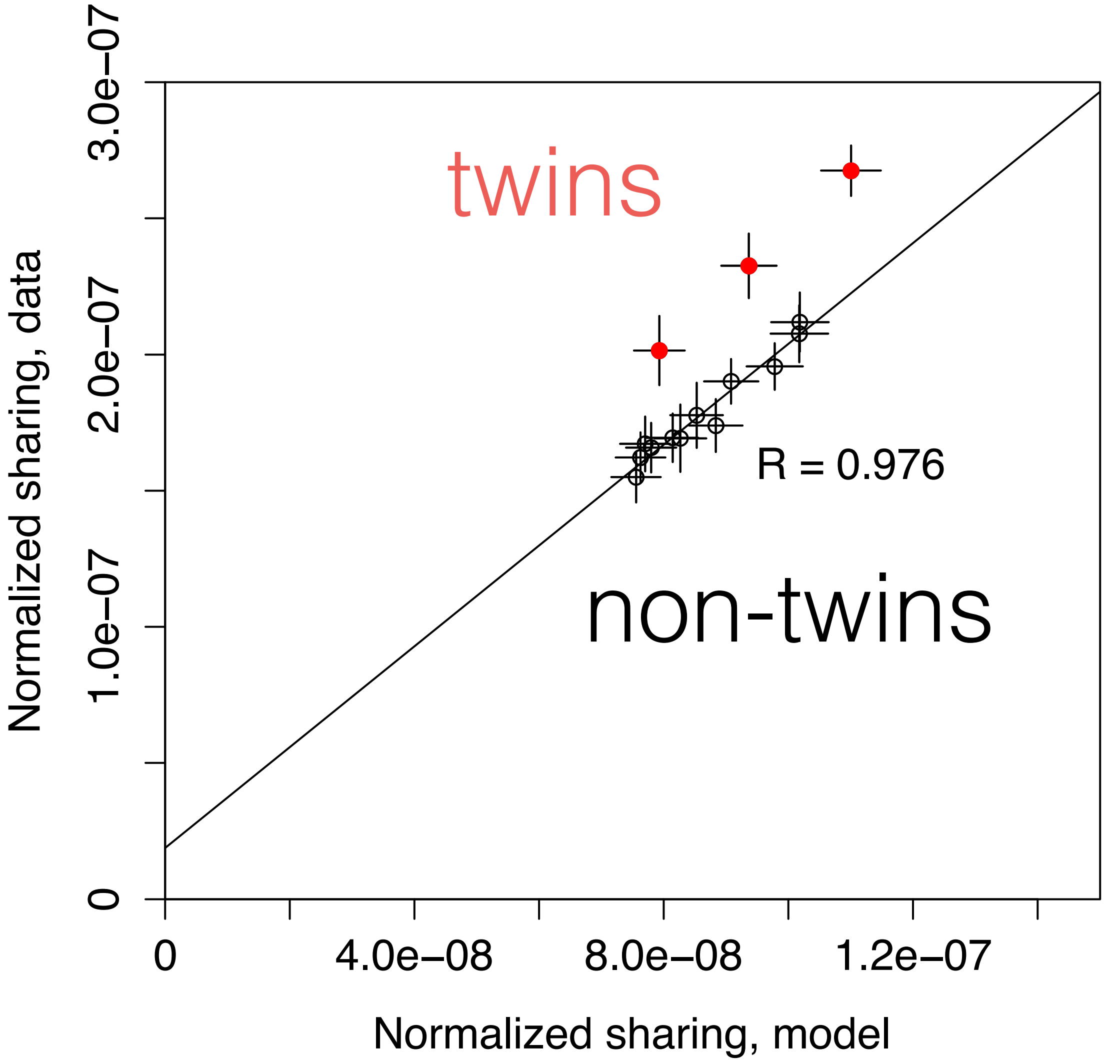
- how many shared receptors between 2 people?



→ close to random expectations

Is everyone the same?

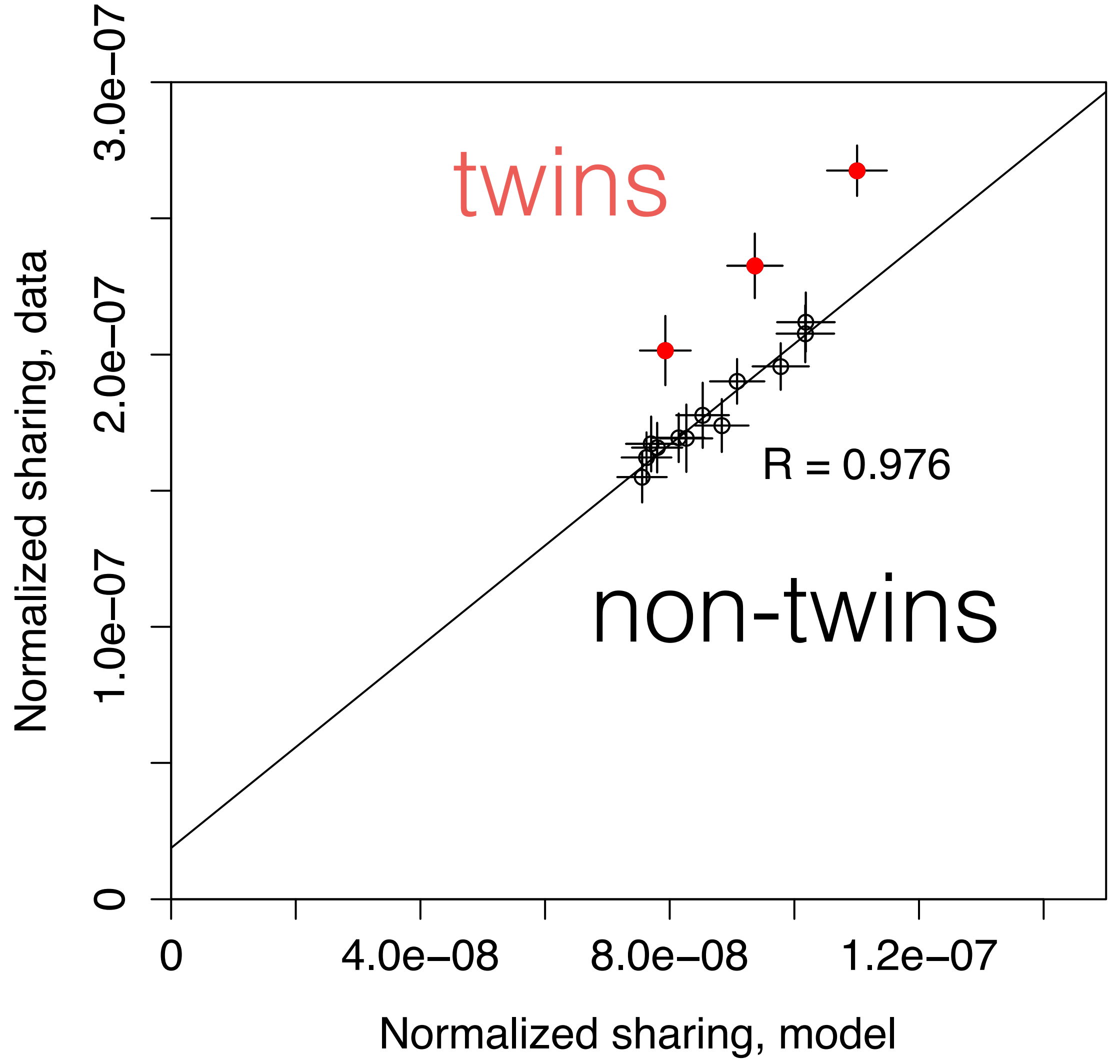
at the level of generation



data from Chudakov lab

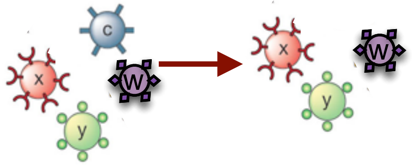
Is everyone the same?

at the level of generation → twins are special

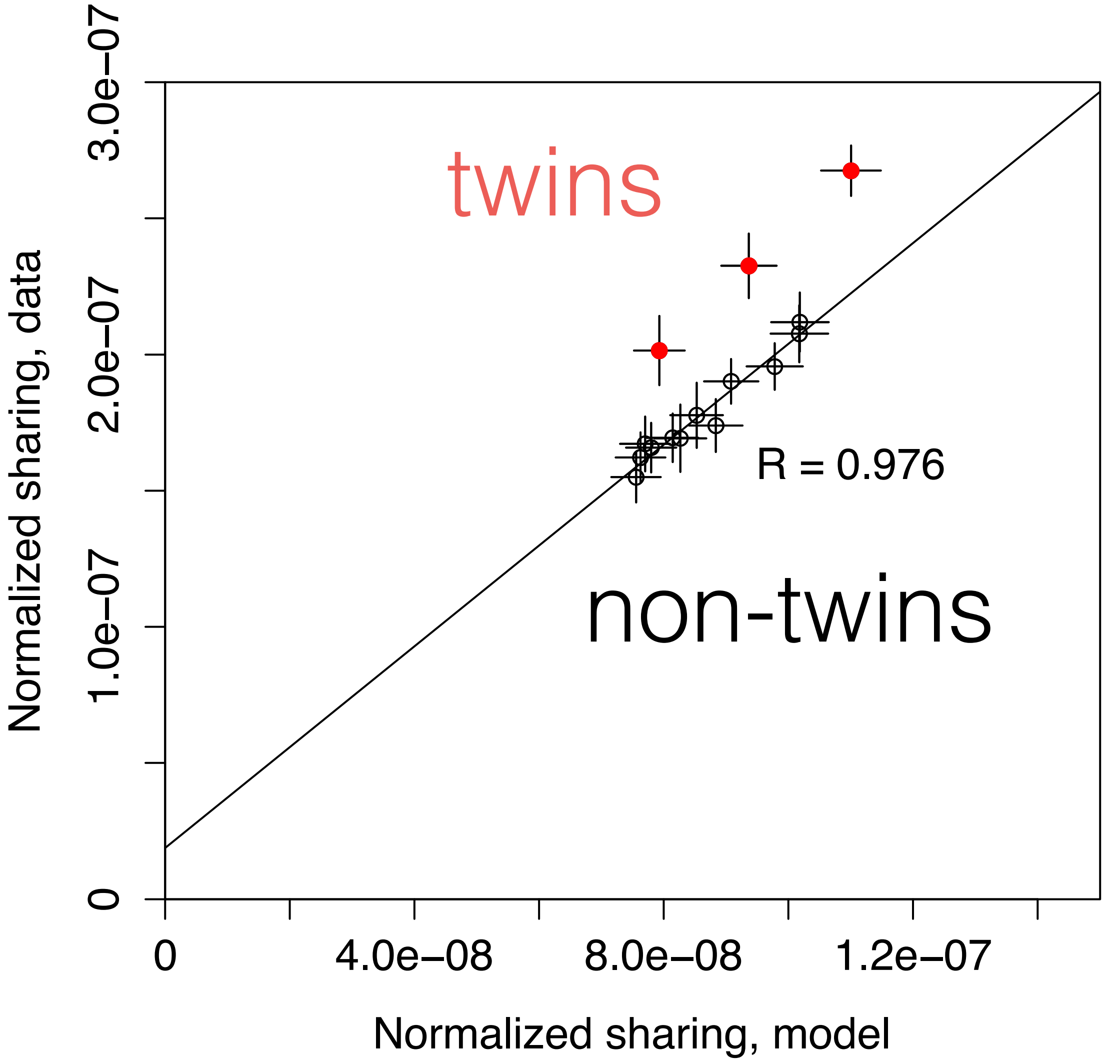


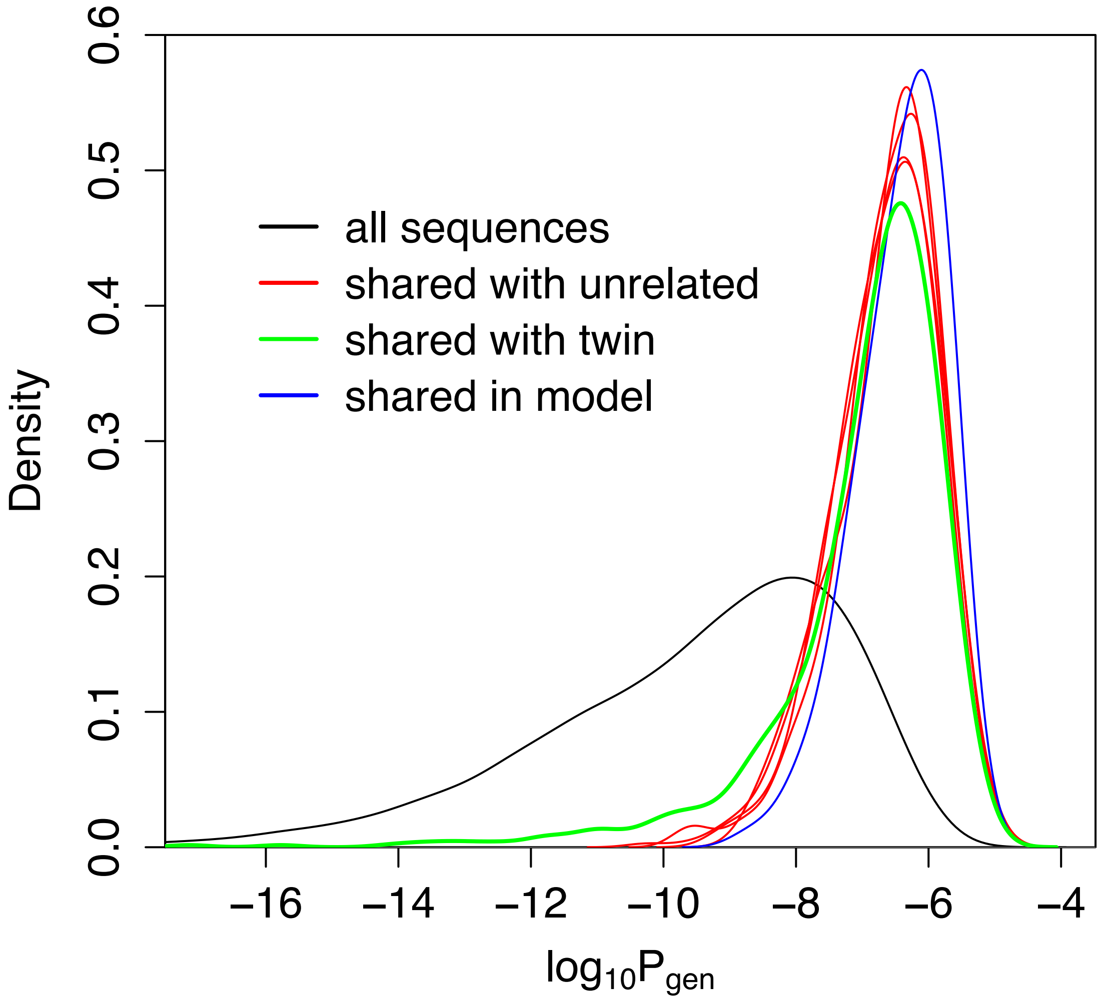
data from Chudakov lab

Is everyone the same?



at the level of generation → twins are special

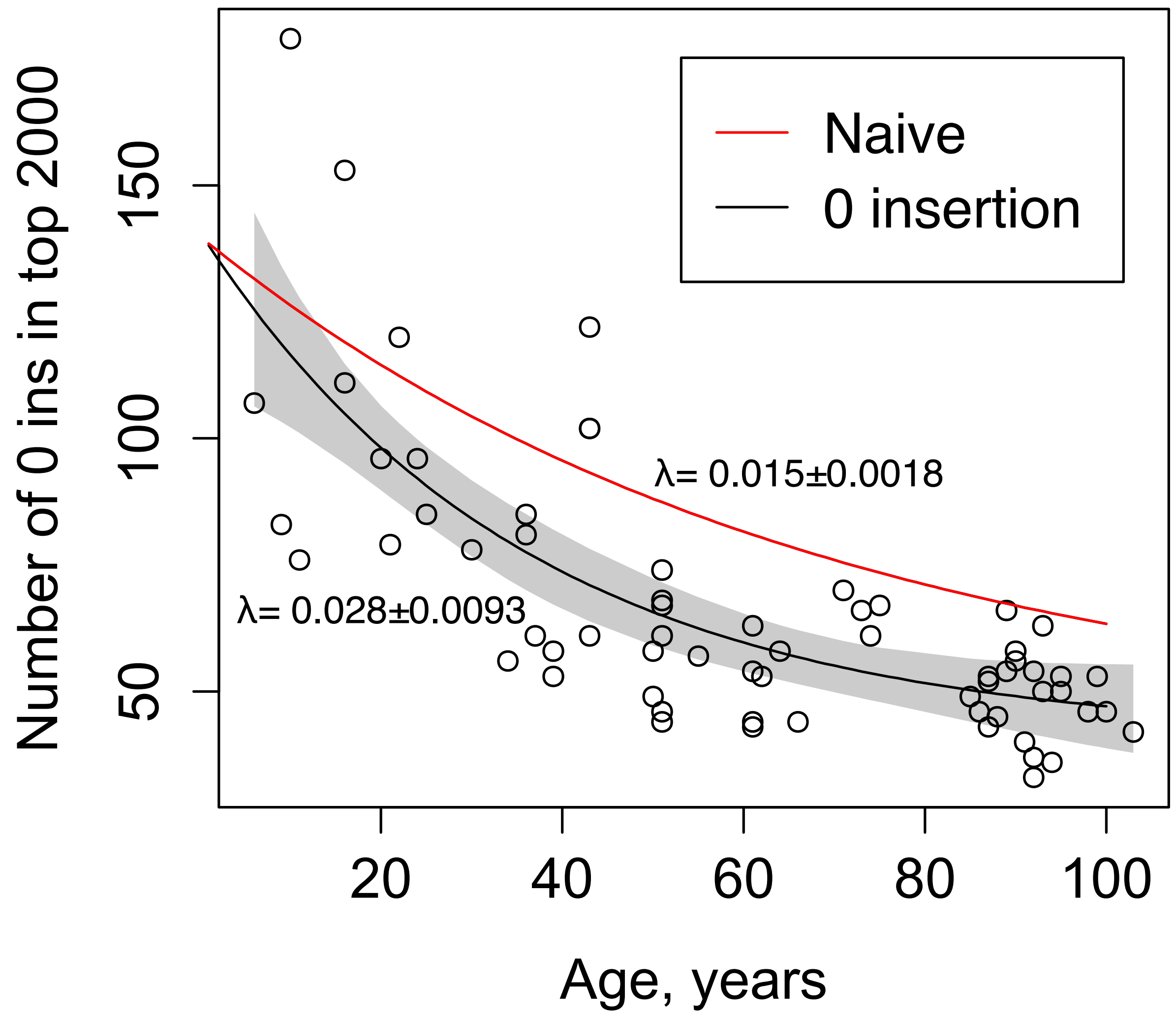


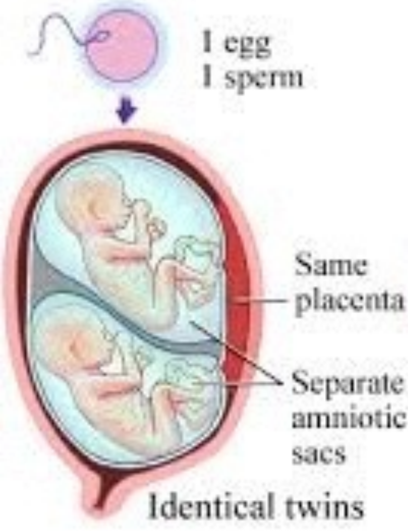


data from Chudakov lab

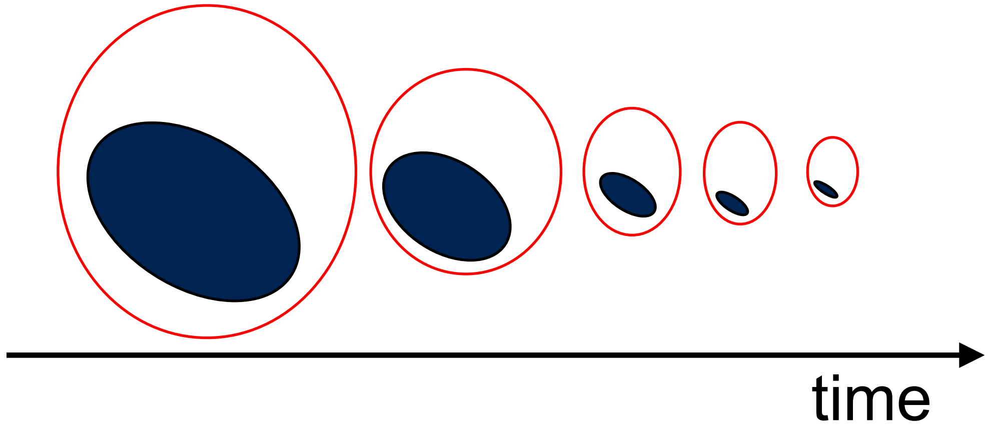
The source: long lived sequences

- last time twins shared blood: before birth
- insertions enzyme less active before birth





clone lifetime ~ 36 years

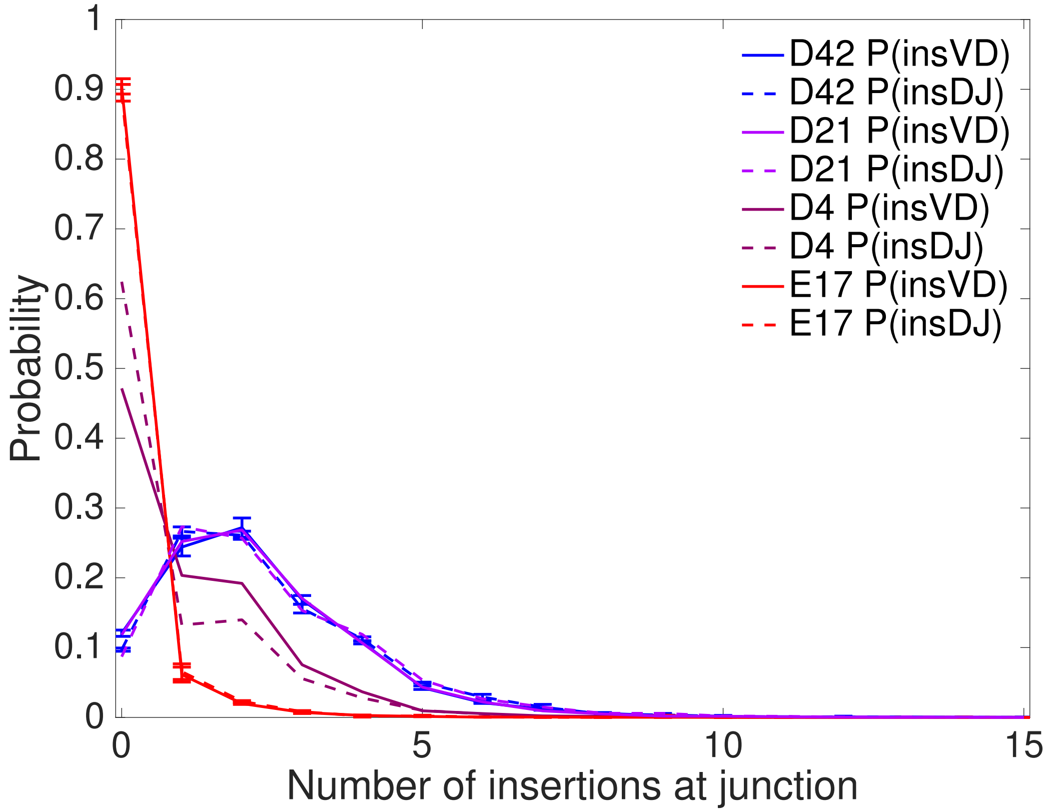


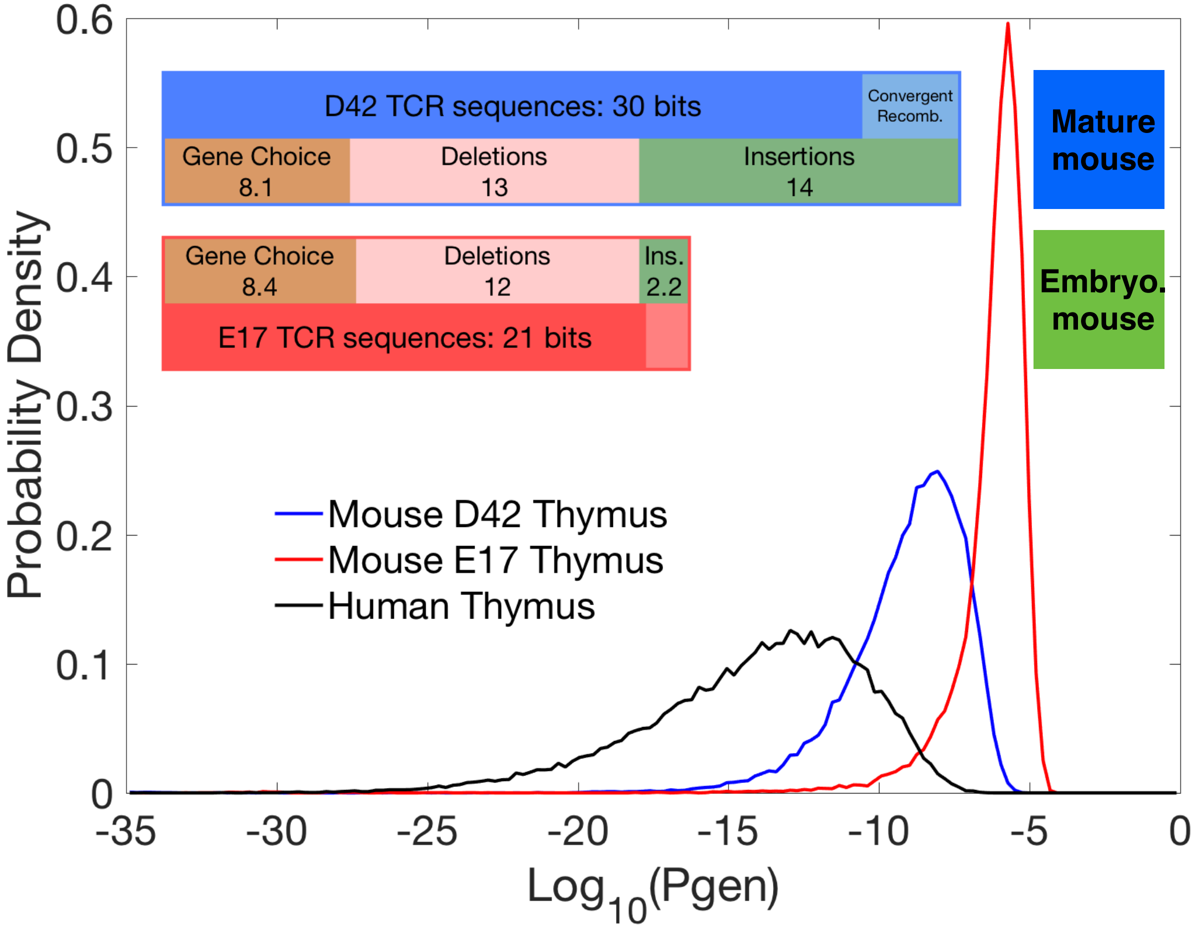
Decay of zero insertion clonotypes:

- zero insertion clonotypes within the naive pool
- size of the total naive pool

Insertions increase diversity

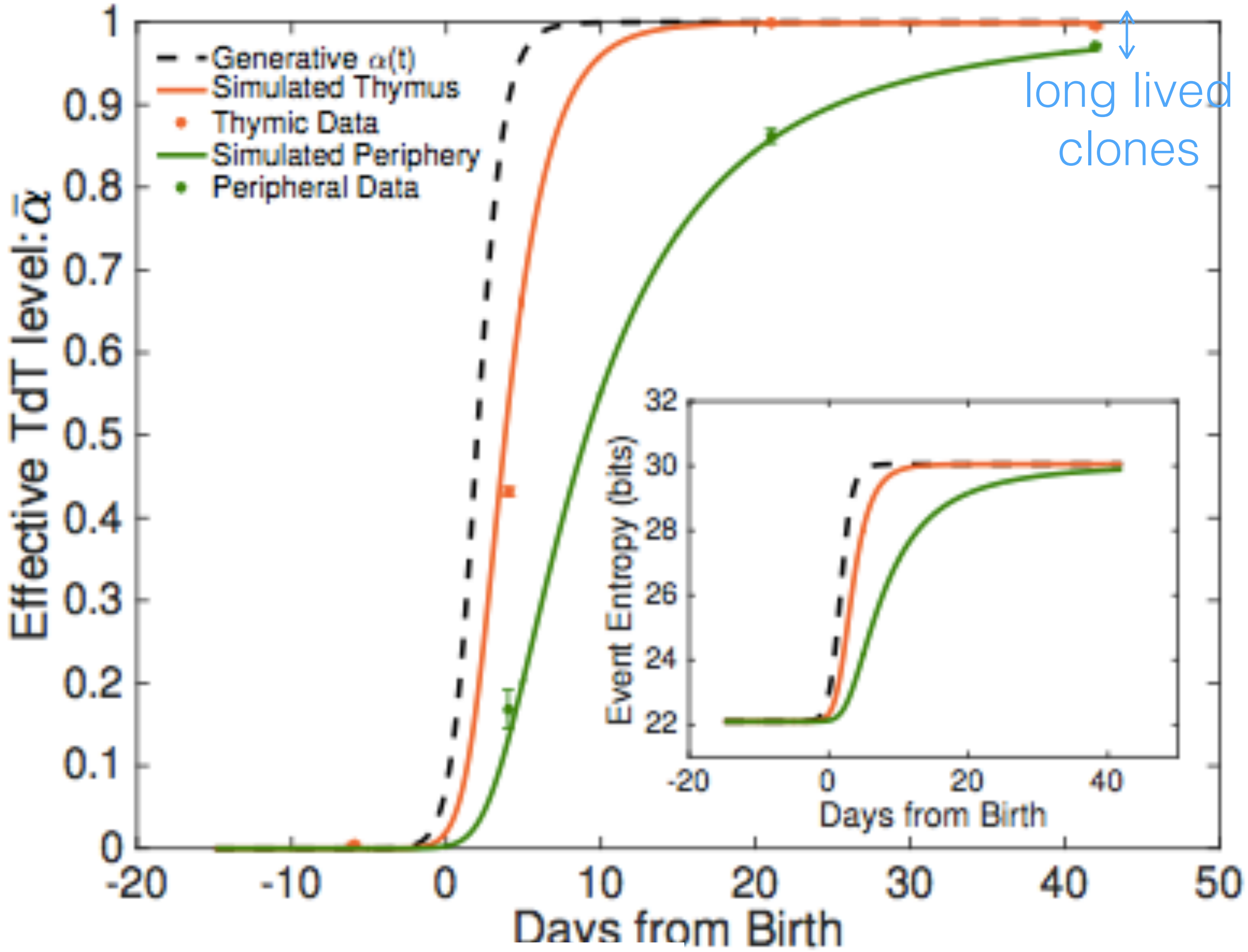
- mouse thymus T-cells
- insertions quickly increase during post-natal development
- diversity increase only due to insertions



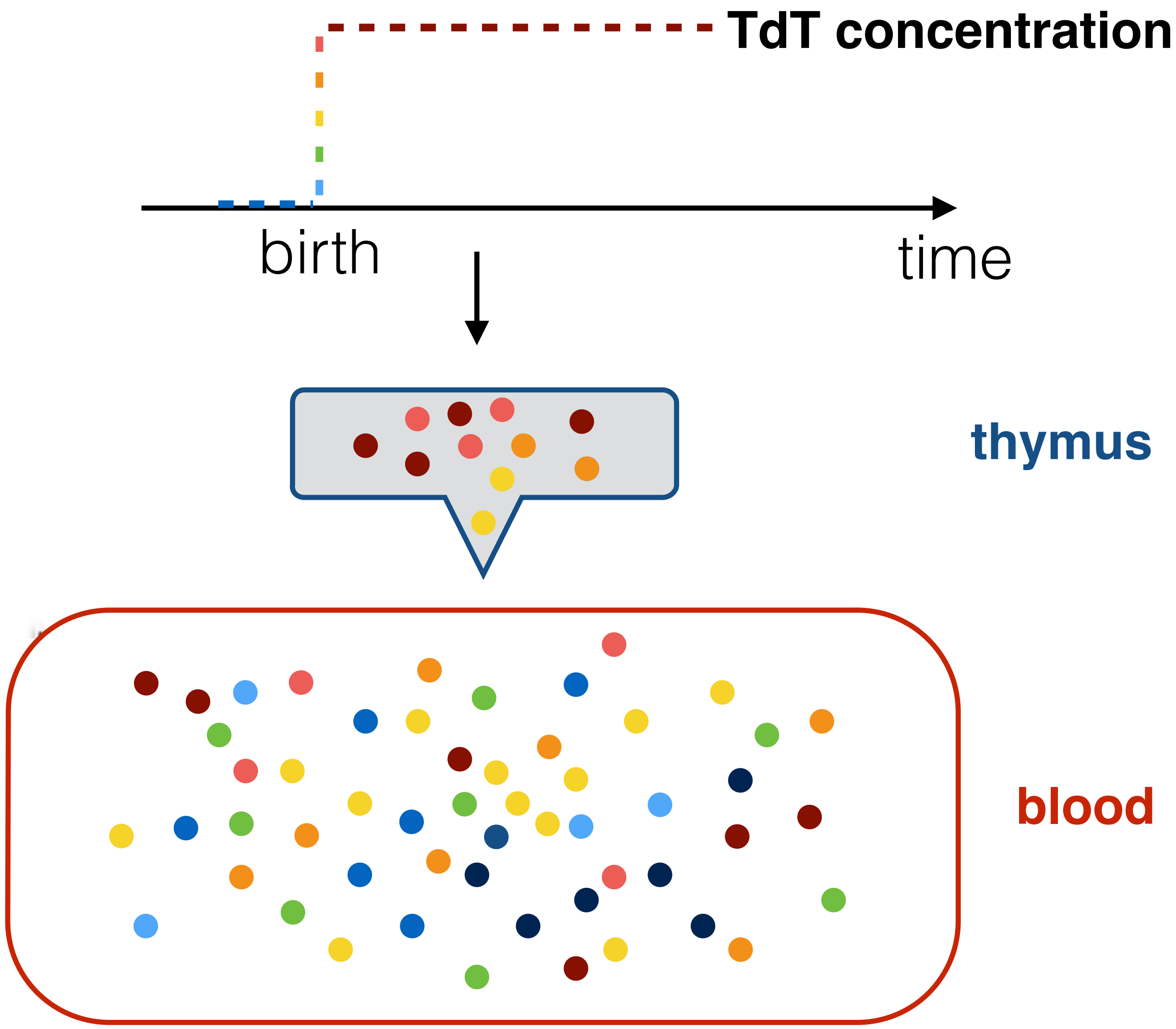


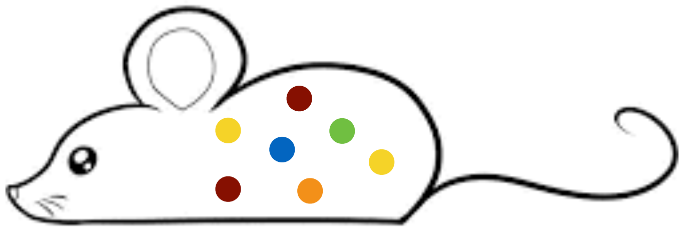
mice: Levine lab
sequencing: Robins lab

Blood tracks thymus

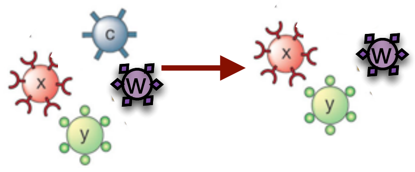


- blood is delayed with respect to thymus
- blood integrates different TdT levels



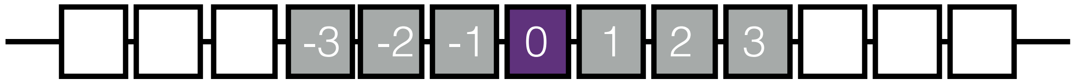


Somatic hypermutations



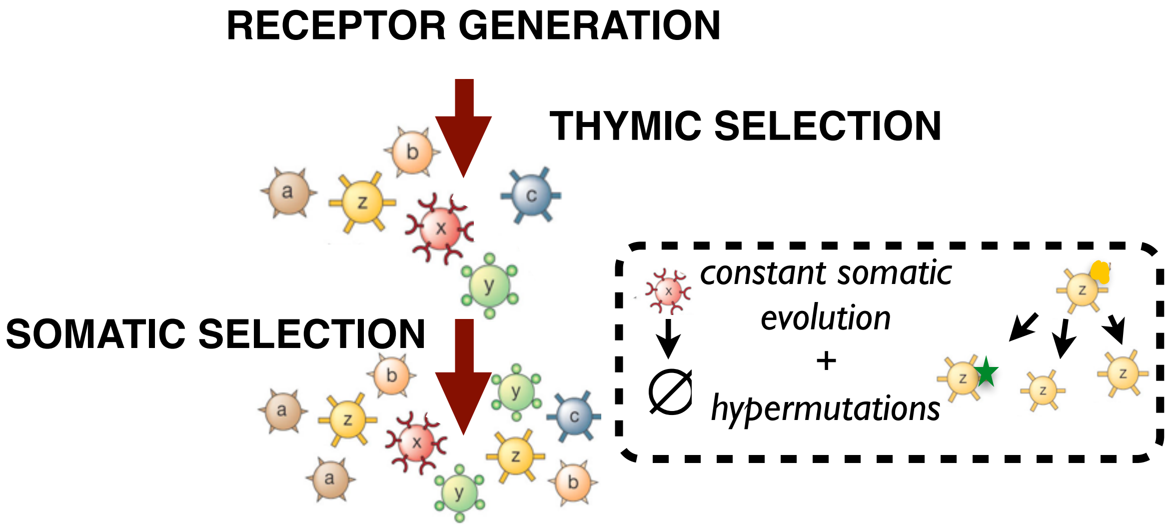
- identifying hypermutation hotspots

7-mer



$$p_{\mathrm{SHM}}(\sigma) \propto p_{\mathrm{bg}}(\sigma) \exp\Big[\sum_{i=-3,3} e_i(\sigma_i)\Big]$$

background probability
of that 7-mer

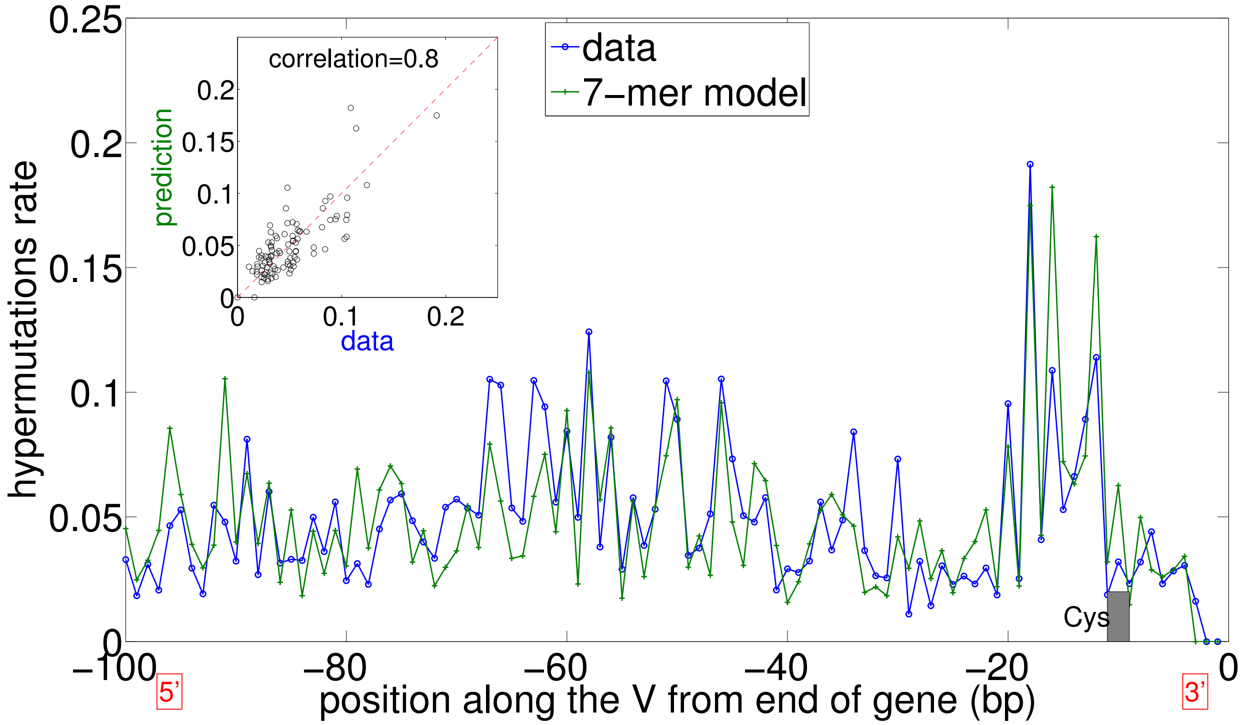


RECEPTOR GENERATION

THYMIC SELECTION

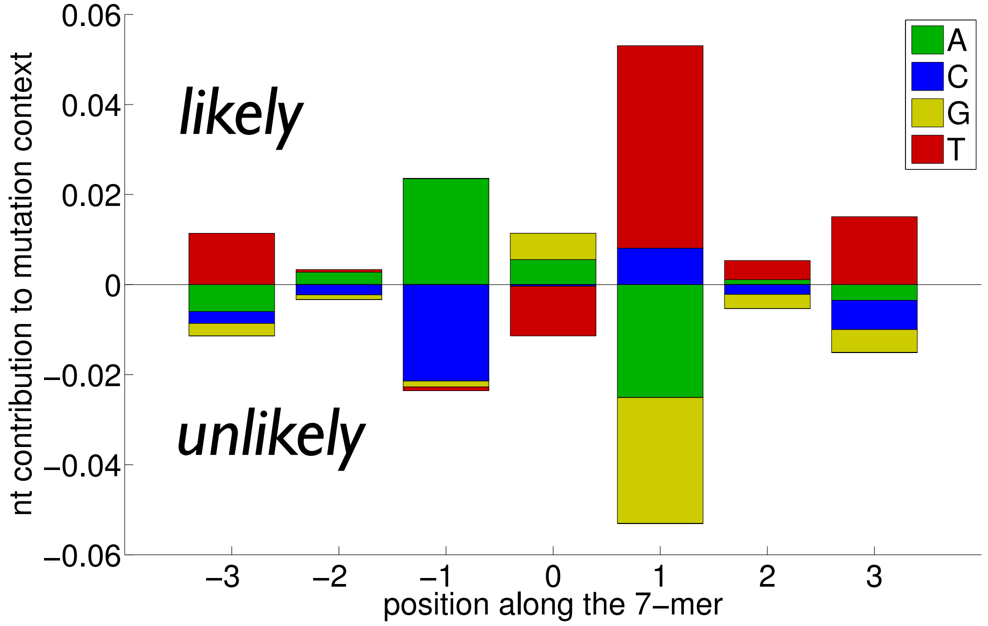
SOMATIC SELECTION

constant somatic evolution
+
hypermutations

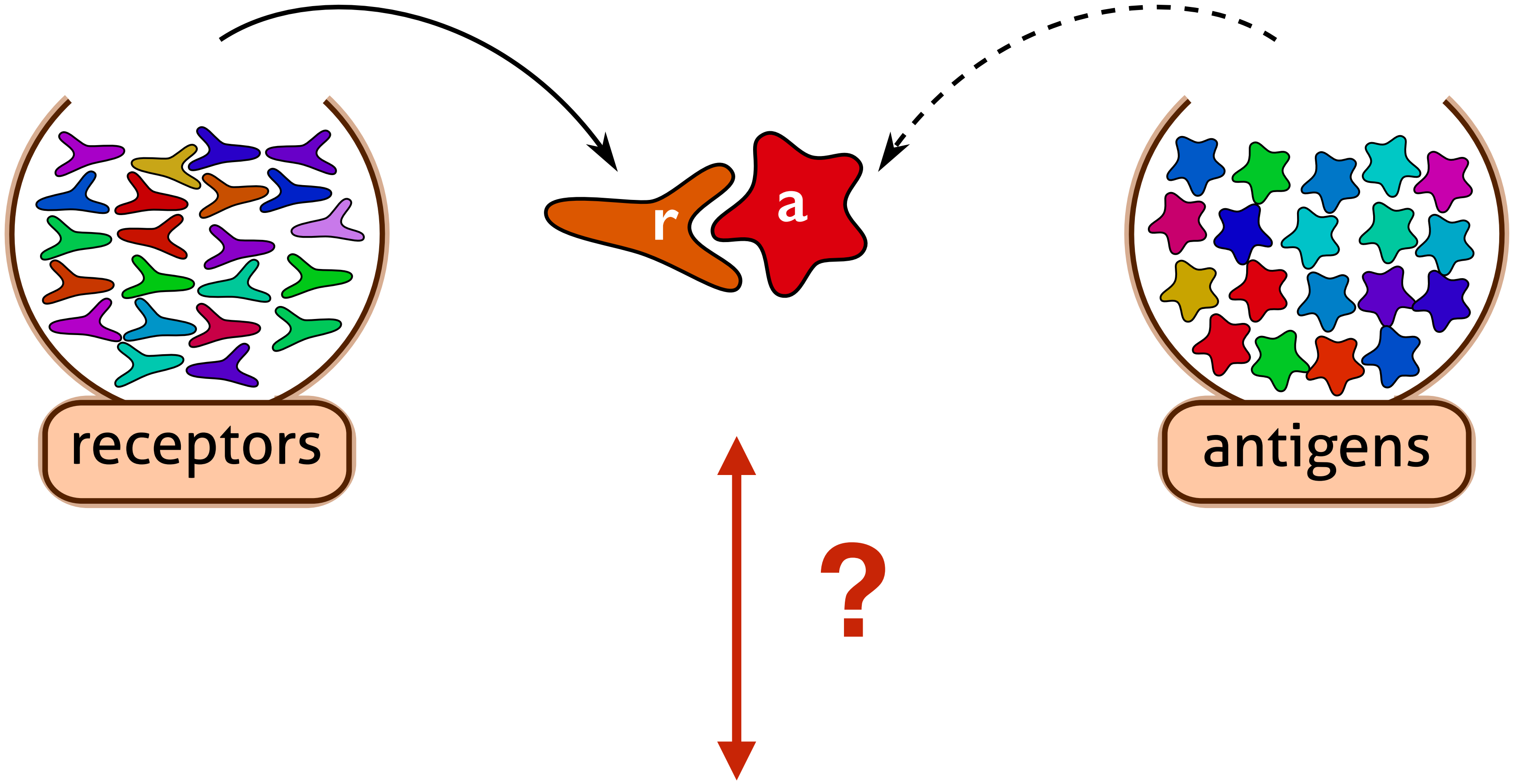


- nt that are likely to hypermutate

hypermutation hotspot 7-mer signature



Linking sequence to function

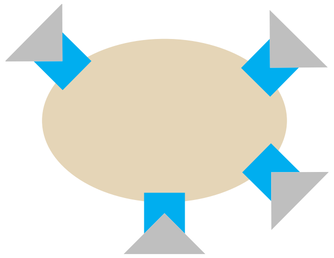


what matters

```
TAGGACCTCGGAAACCTCTTCTGCTGGCGCAGAGATACAGGGCTGAGTCTTCTT
TATCTGCCGATGTCGCGAAGGCCCTCCCGCTAAGATCACTGGTGGCACAGAAGT
TAGGGAGAGGTGCCTAACTGCTGGCACAGAAGTACAGAGAGGTCTGGTTGGGGT
TCCGCCGCTAGTCCCTGAAACTACTGGCACAGAGATAGAAAGCTGTCGGGTTCT
CGAAACTGCTGGCACAGAAGTACACAGATGTTTGGGAGGGAGCAGCCGACTCCA
TCTTGGCCGCTAGTCCGAGAAACTGCTGGCACAGAAGTACACAGATGTTTGGGA
TTGTAGGAGCCGGACCGGCCCCCTGTCCCCTTGGCTGCTGGCGCAGAGATACAG
TCATTTAACGTGCGGCCCGCCTGGCACAGAAGTAAAGAGCTGTCTGGTTGTGGT
TAGTAACTCCGCTTCACTGCTGGCACAGAAGTACACAGATGTCTGGGAGGGAGC
TCCCTCCGGTTTGAAGGGTCTGCTGGCACAGAAGTACACAGATGTTTGGGAGGG
CCGGTGTTCGCACAGCCCTGGGGACCCTGGCGCAAACCCCGCTTCCCTCGAGGA
```

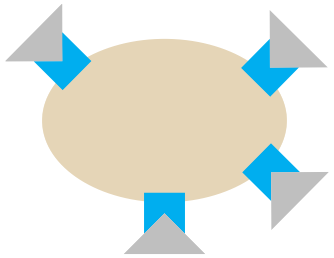
what we have

Tite-Seq



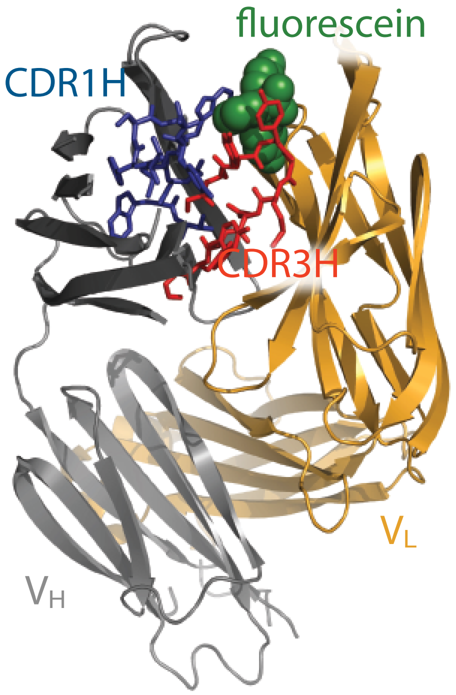
yeast display + FAC**Sort** + **seq**uence

Tite-Seq

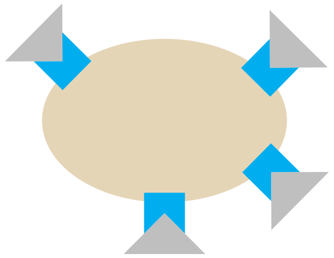


yeast display + FAC**Sort** + **seq**uence

- proof of principle: antigen = fluorescein
  antibody = scFv



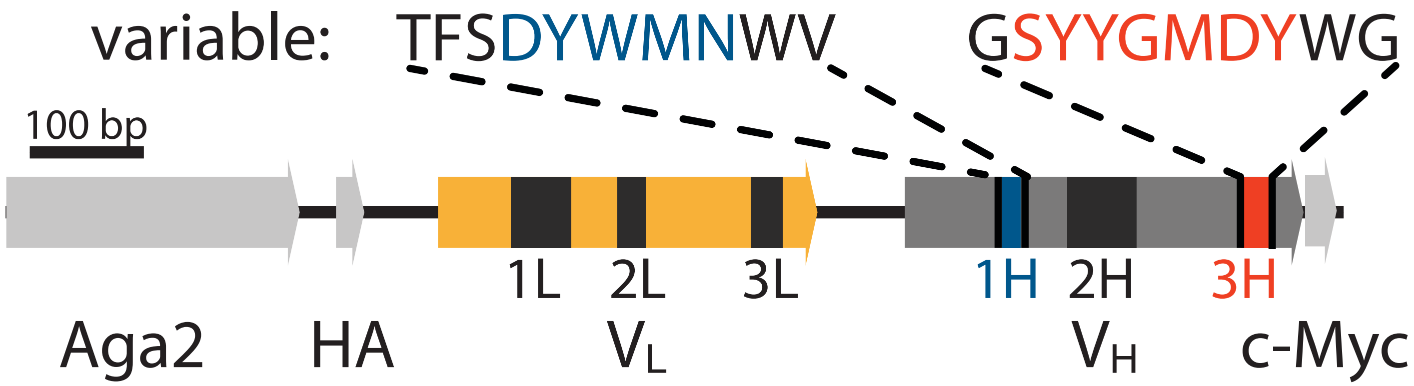
Tite-Seq



yeast display + FAC**Sort** + **seq**uence

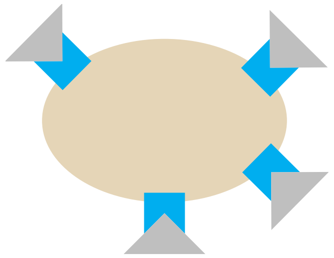
- proof of principle: antigen = fluorescein
  antibody = scFv

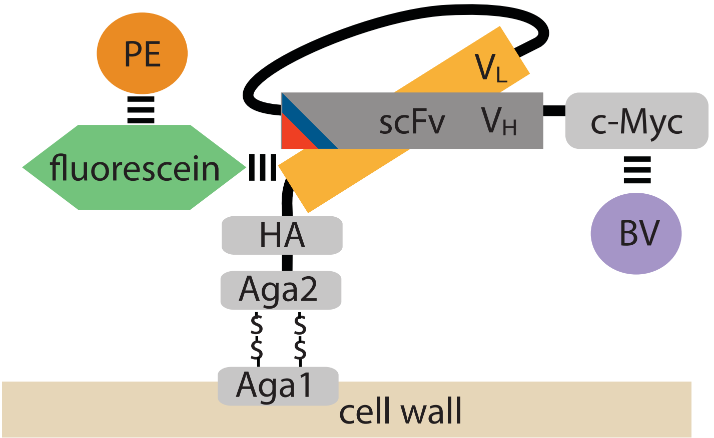
- exhaustive single mutations in CDR1H and CDR3H



variation	1H	3H
1 codon	600	600
2 codon	1100	1100
3 codon	150	150

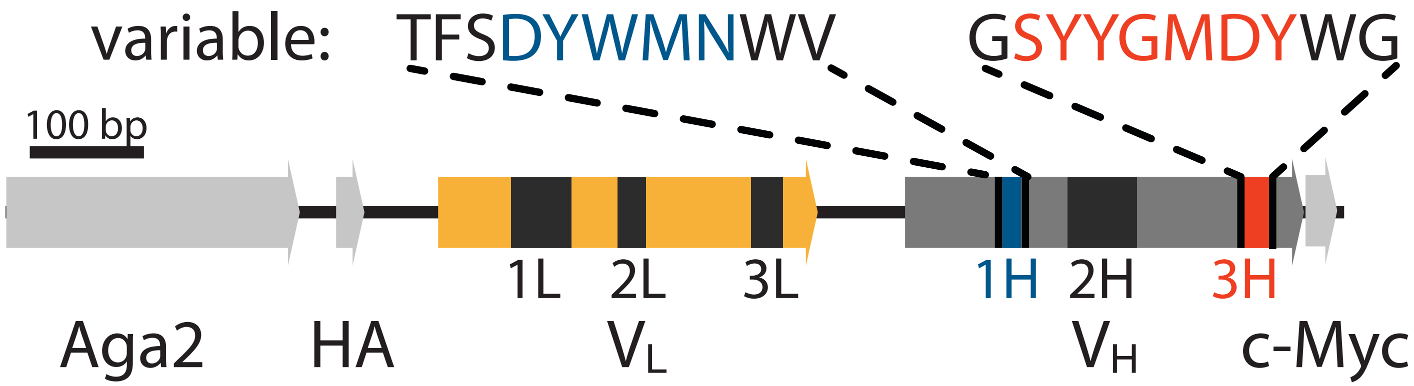
Tite-Seq





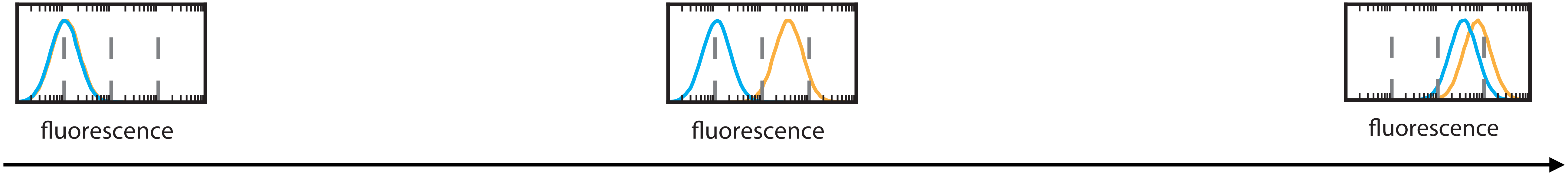
yeast display + FACSort + **seq**uence

- proof of principle: antigen = fluorescein
  antibody = scFv
- simultaneously measure affinity and expression
- exhaustive single mutations in CDR1H and CDR3H



variation	1H	3H
1 codon	600	600
2 codon	1100	1100
3 codon	150	150

Tite-Seq

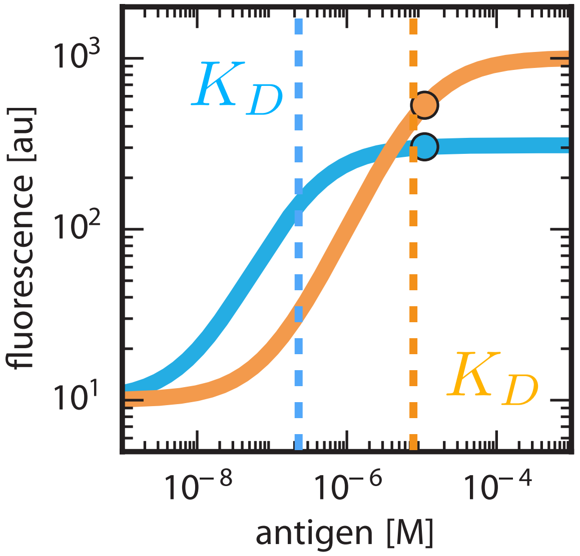


fluorescence

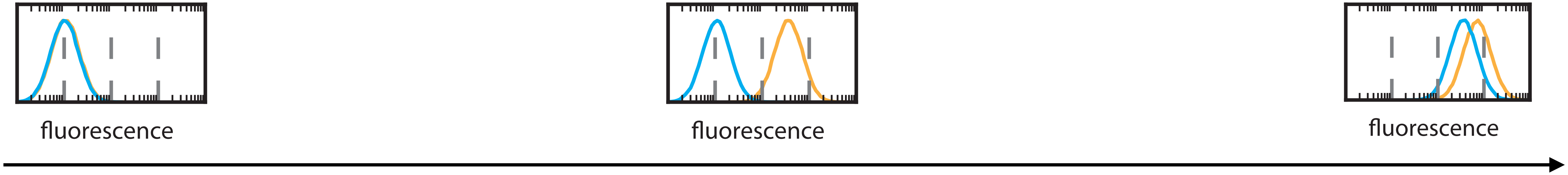
fluorescence

fluorescence

antigen concentration



Tite-Seq

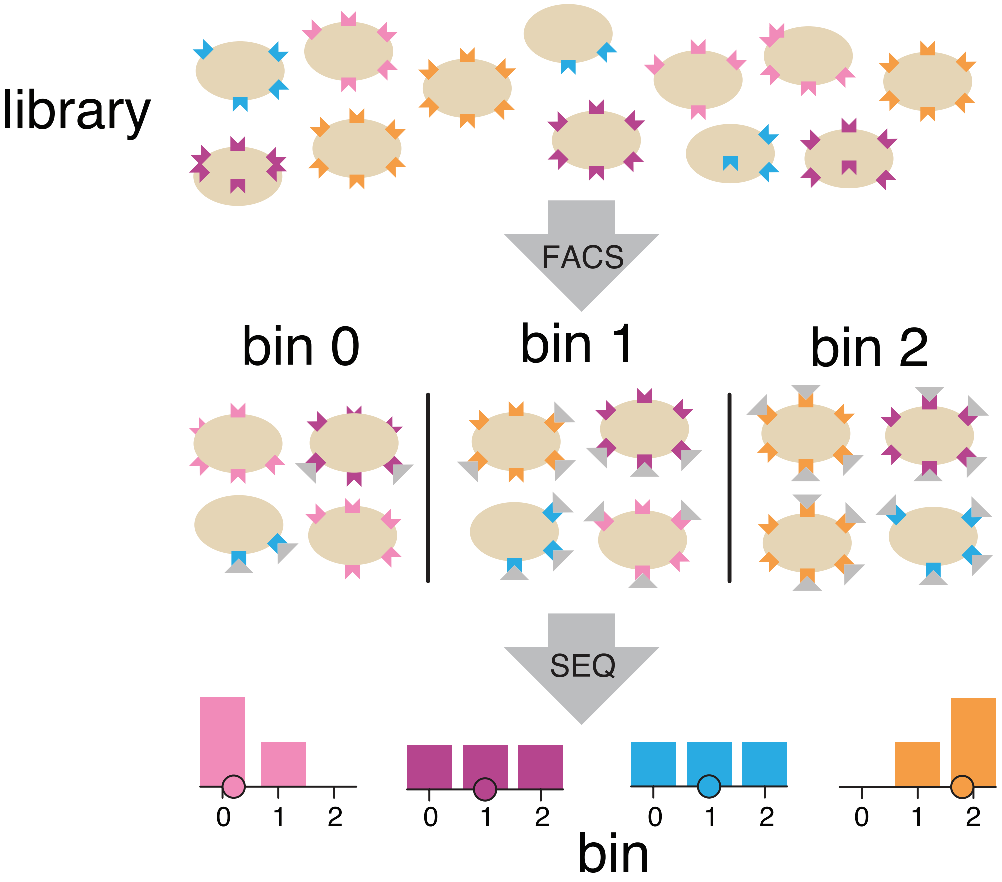


fluorescence

fluorescence

fluorescence

antigen concentration



library

FACS

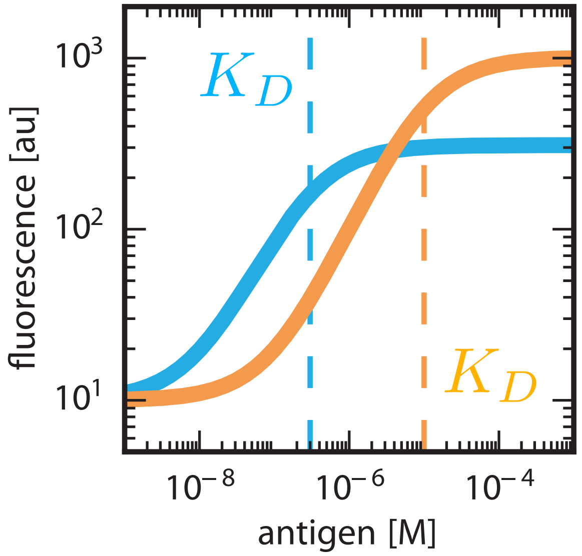
bin 0

bin 1

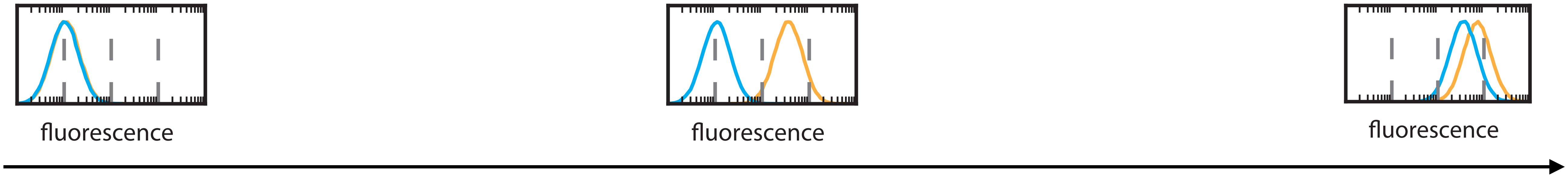
bin 2

SEQ

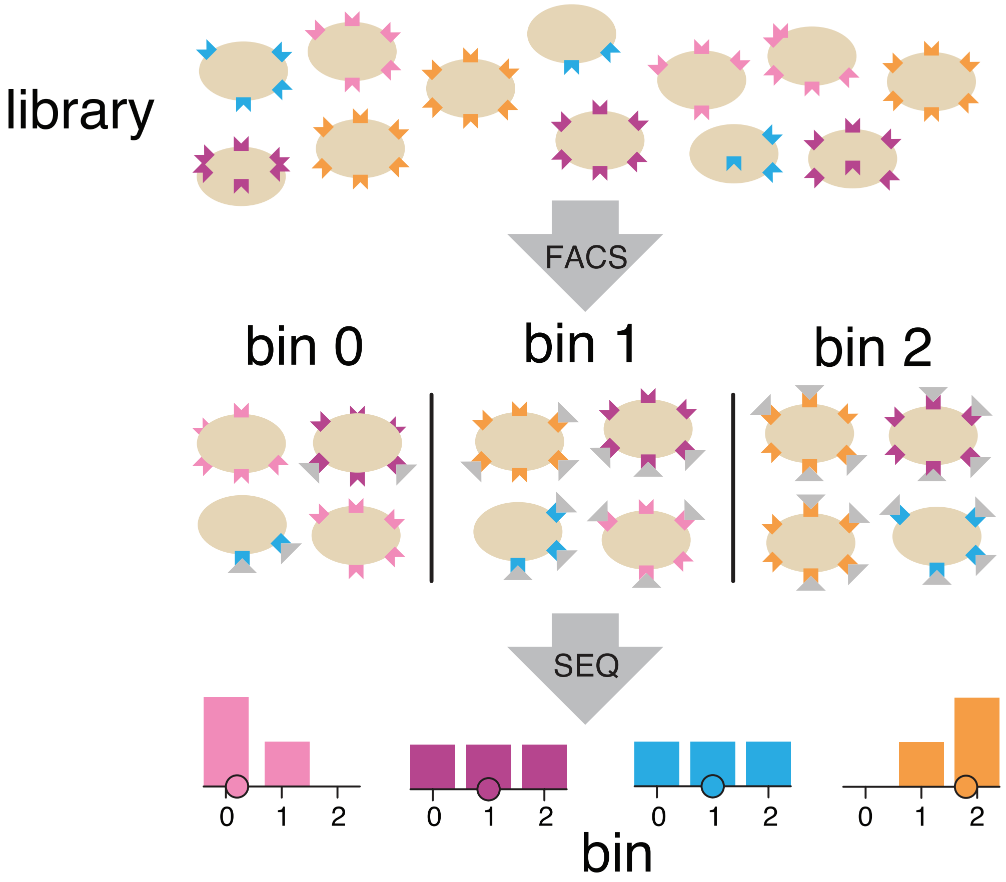
bin



Tite-Seq



antigen concentration

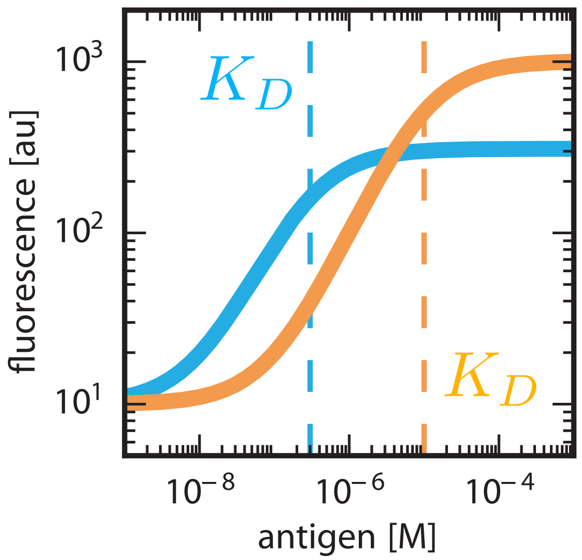

library

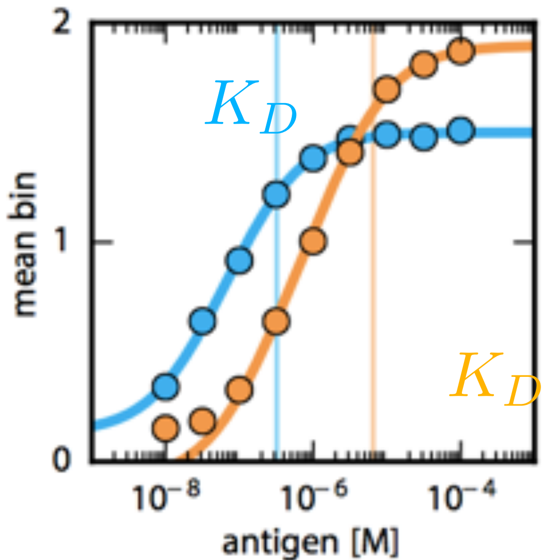
FACS

bin 0 bin 1 bin 2

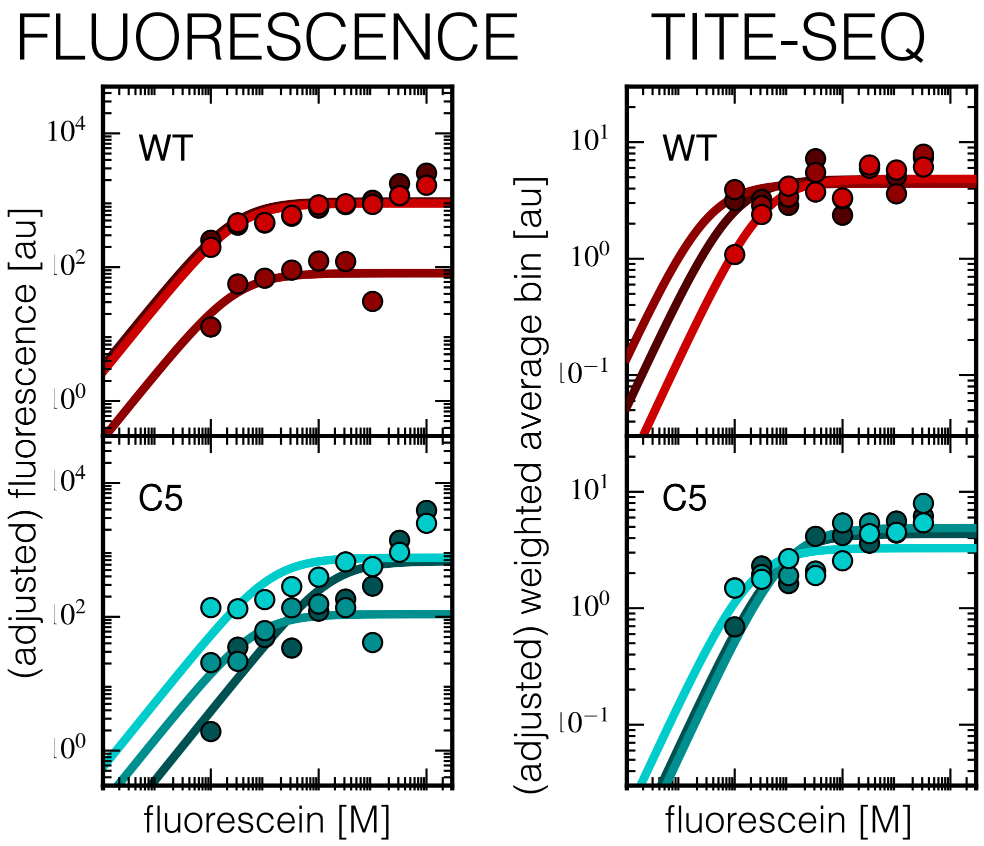
SEQ

bin

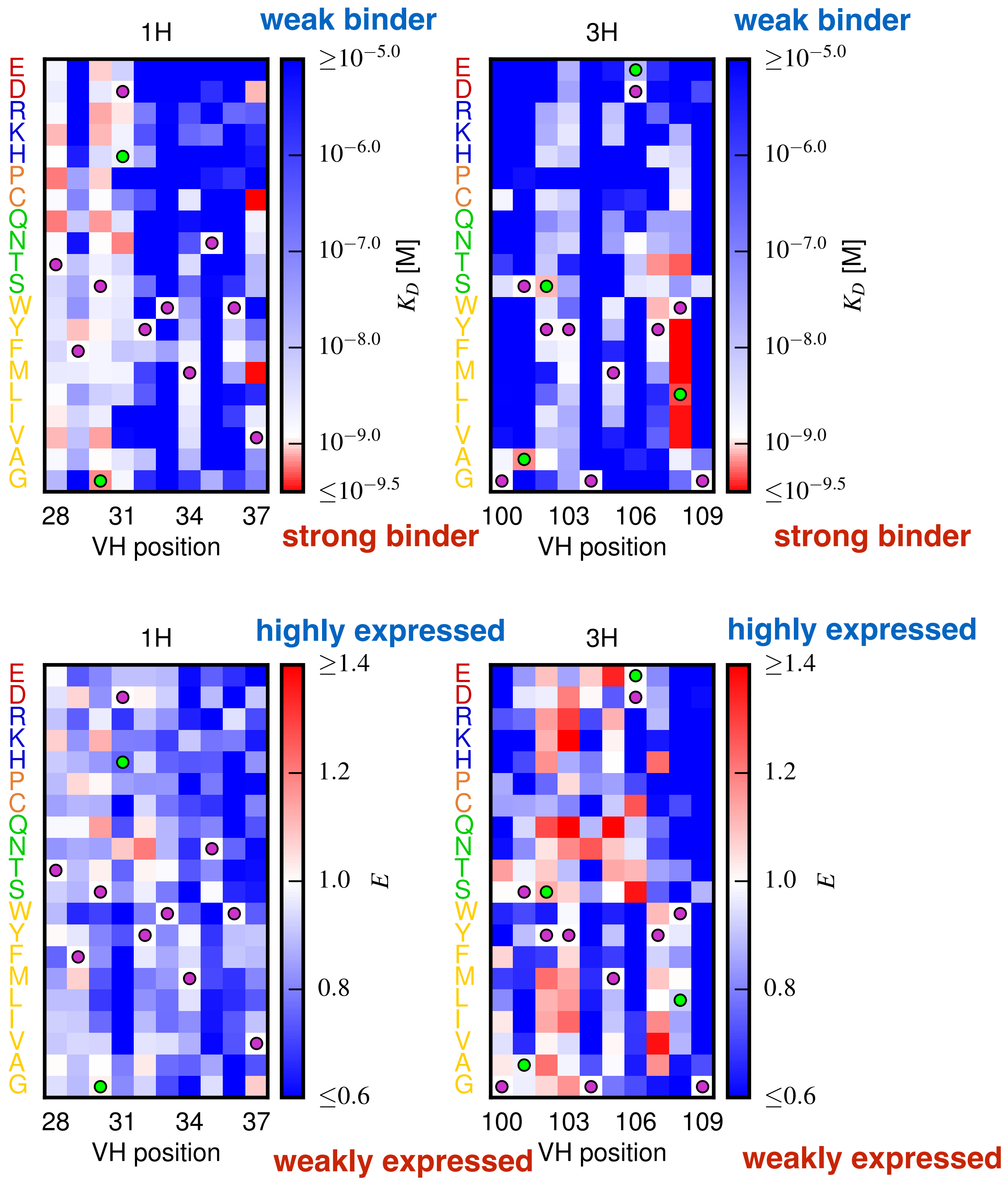




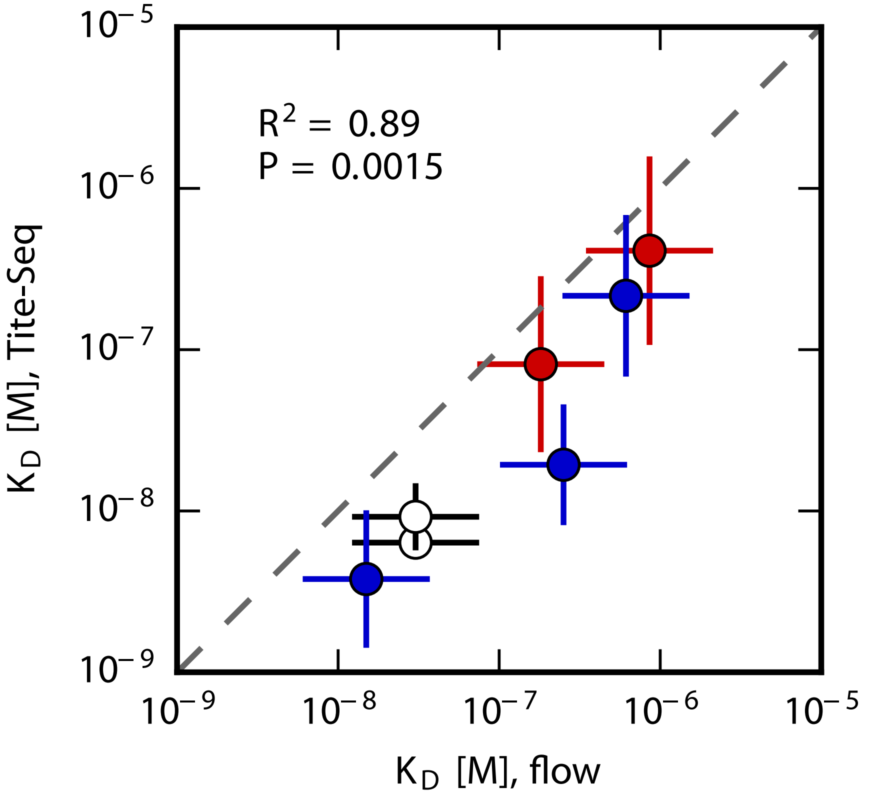
Comparison to fluorescence measurements



- reproducible results from CDR1 and CDR3 mutants



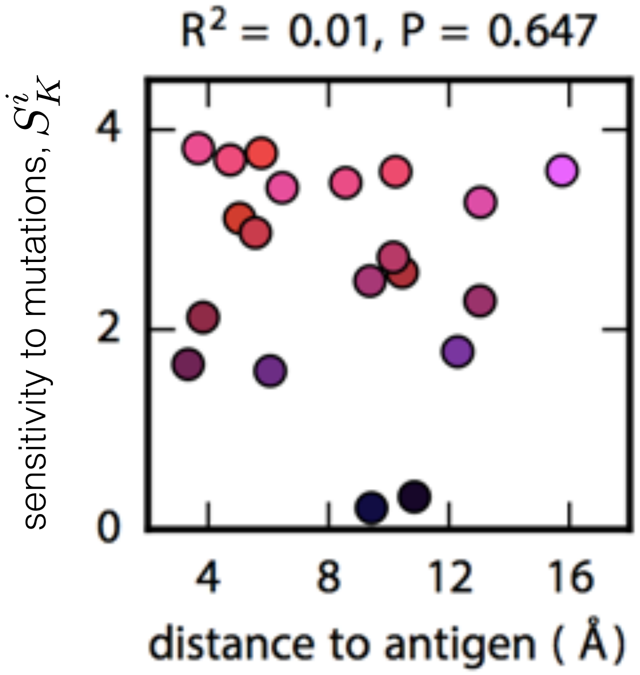
- wide range of affinities



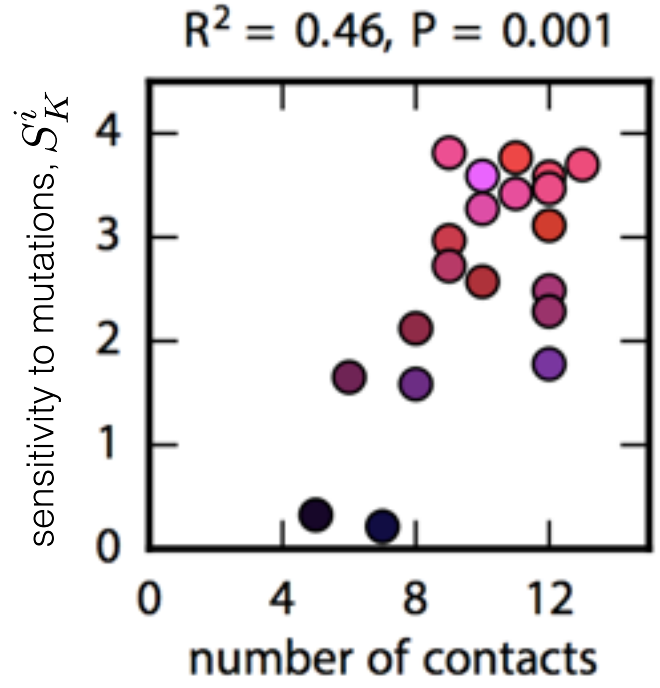
Mutational effects by CDR

affinity:

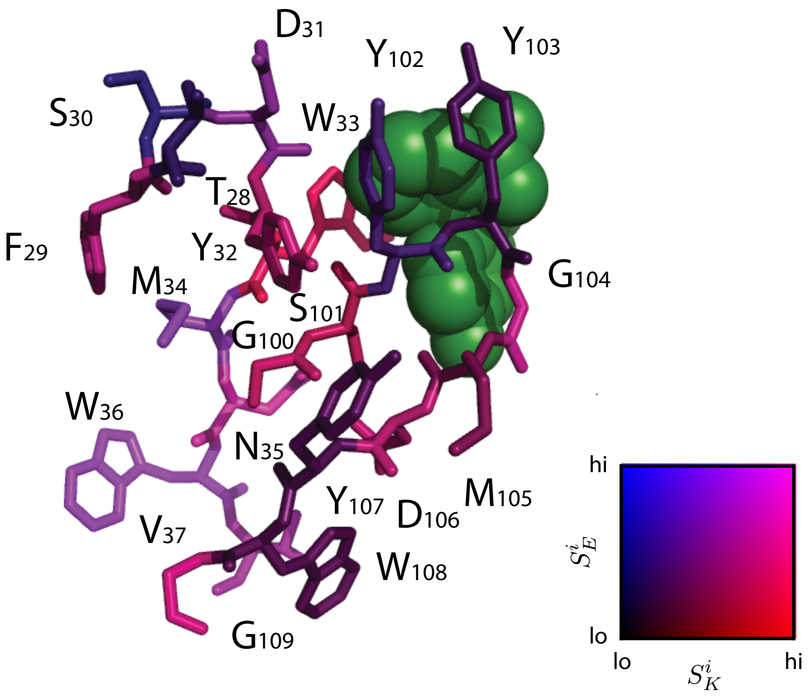
- independent of distance of mutation to antigen



$$S_K^i = \sqrt{\left\langle \left(\log_{10} K_D^{ia} - \log_{10} K_D^{\mathrm{WT}}\right)^2 \right\rangle_a}$$



- depends on the number of contacts a residue makes in the receptor

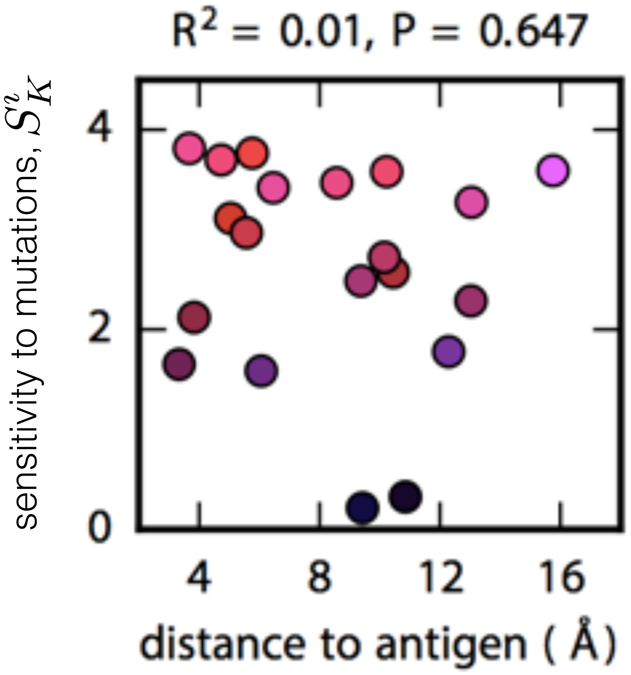


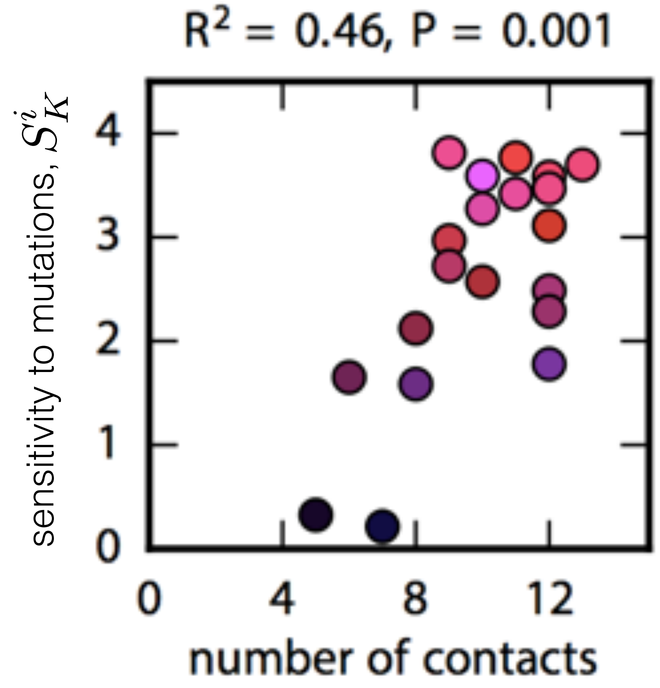
Mutational effects by CDR

$$S_K^i = \sqrt{\left\langle \left(\log_{10} K_D^{ia} - \log_{10} K_D^{\mathrm{WT}}\right)^2 \right\rangle_a}$$

affinity:

- independent of distance of mutation to antigen

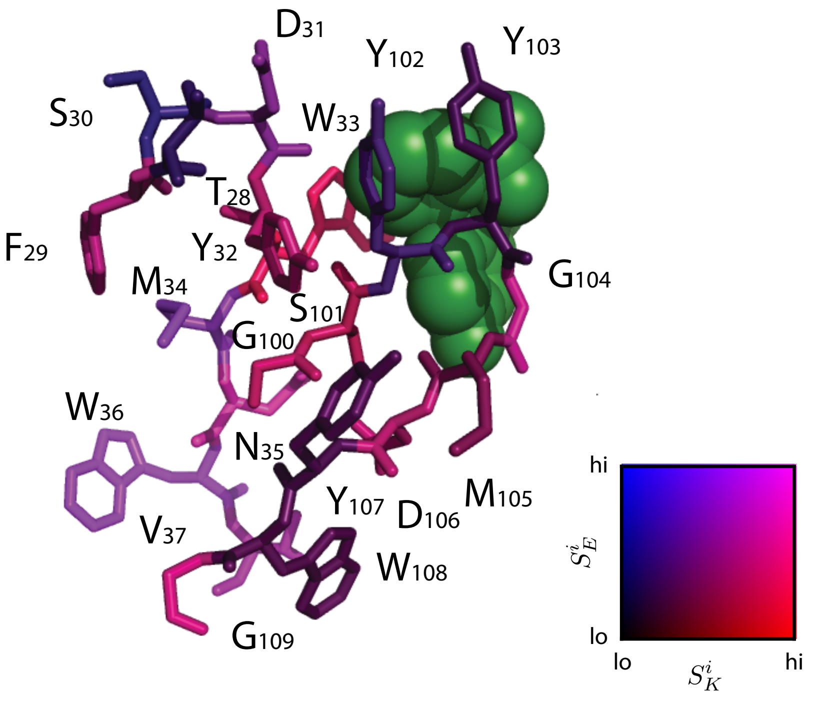




- depends on the number of contacts a residue makes in the receptor

- non-local effects

→ effect of interactions between receptor residues

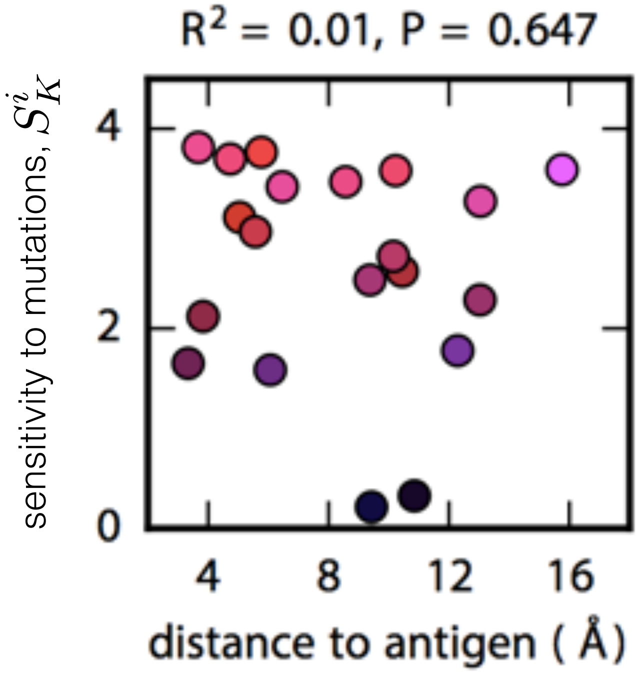


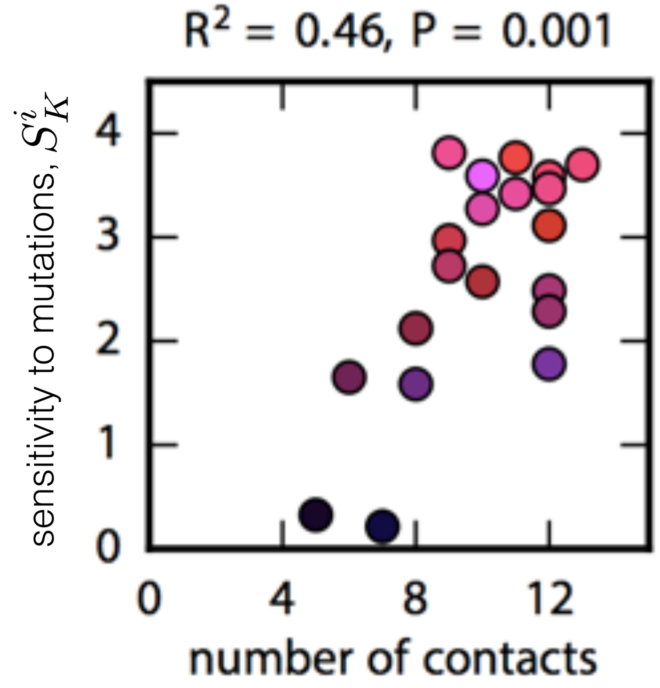
Mutational effects by CDR

$$S_K^i = \sqrt{\left\langle \left(\log_{10} K_D^{ia} - \log_{10} K_D^{\mathrm{WT}}\right)^2 \right\rangle_a}$$

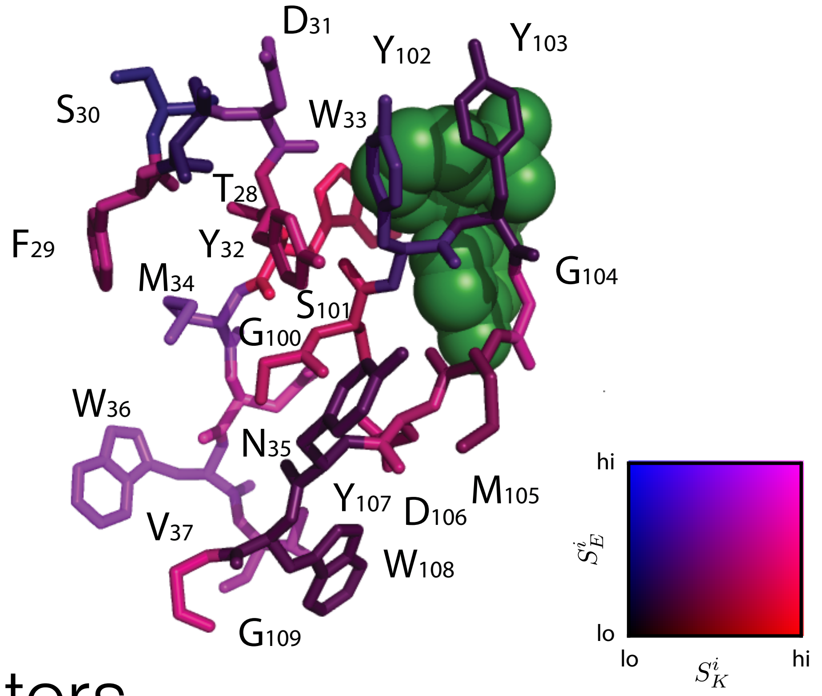
affinity:

- independent of distance of mutation to antigen





- depends on the number of contacts a residue makes in the receptor



- non-local effects

→ effect of interactions between receptor residues

- CDR3 greater effect on affinity

→more likely to be mutated in functional receptors

Conclusions

generating diversity and selection:

- random overlap between (most) individuals
- very long lived clones

response:

- identify responding clonotypes
- observe selection on standing variation

Tite-Seq:

- high-throughput affinity and expression assay
- interactions within receptors

T Mora, AM Walczak, W Bialek, CG Callan, PNAS (2010)
A Murugan, T Mora, AM Walczak, CG Callan, PNAS (2012)
Y Elhanati, A Murugan, CG Callan, T Mora, AM Walczak, PNAS (2014)
A Mayer, V Balasubramanian, T Mora, AM Walczak, PNAS (2015)
Y Elhanati, Z Sethna, Q Marcou, CG Callan, T Mora, AM Walczak, Phil. Trans. B (2015)
J. Desponds, T. Mora, AM Walczak, PNAS (2015)
A Mayer, T Mora, O Rivoire, AM Walczak, PNAS (2016)
Y Elhanati, Q Marcou, T Mora, AM Walczak, Bioinformatics (2016)
RM Adams, JB Kinney, T Mora, AM Walczak, eLife (2017)
M. Pogorelyy et al, PloS CB (2016)
T Mora, AM Walczak, qbio/bioarxiv (2016)
Z Sethna, Y Elhanati, CG Callan, T Mora, AM Walczak, PNAS (2017)
Q. Marcou. T. Mora, AM Walczak, qbio/bioarxiv (2017)