Modeling data on the TCR-frequency distribution of naive T cells

Rob J. de Boer
Theoretical Biology and Bioinformatics
Utrecht University

Two PhD students: Peter de Greef & Bram Gerritsen

Collaboration with group of Benny Chain, UCL.

Utrecht University

The immune system is a distributed complex system composed of circulating random detectors

Naive lymphocytes (detectors) circulate (patrol) via blood and lymph.

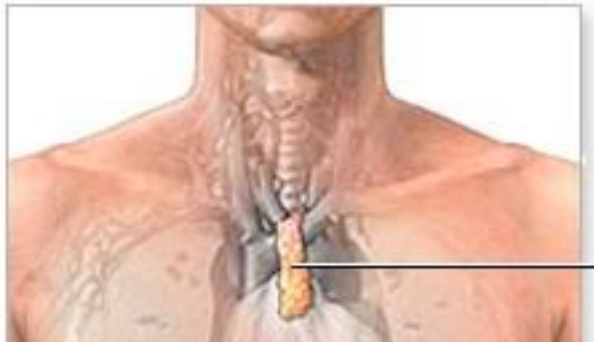
Naive B lymphocytes are born in the bone marrow and can be triggered to produce antibodies.

Naive T lymphocytes are born and selected in the thymus and can differentiate into helper (CD4) or killer (CD8) T cells.

Each lymphocyte express a randomly generated protein (receptor) that by chance binds a very small fraction ($<10^{-5}$) of the proteins (ligands) in our environment.

After binding cognate ligand, naive cells expand and “decide” on their effector function.

Decisions are remembered because a fraction of the cells persist as “memory” cells (immunity).



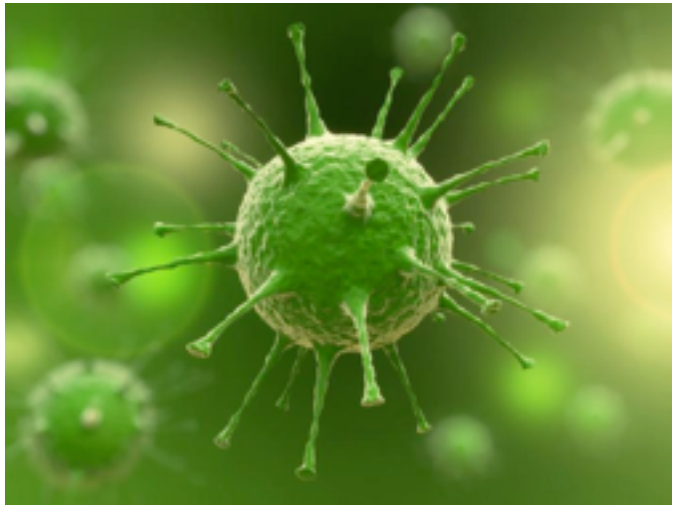
10^{11} naive T cells (CD4 and CD8)

10^8 - 10^{11} different TCRs (colors)

$p < 10^{-5}$

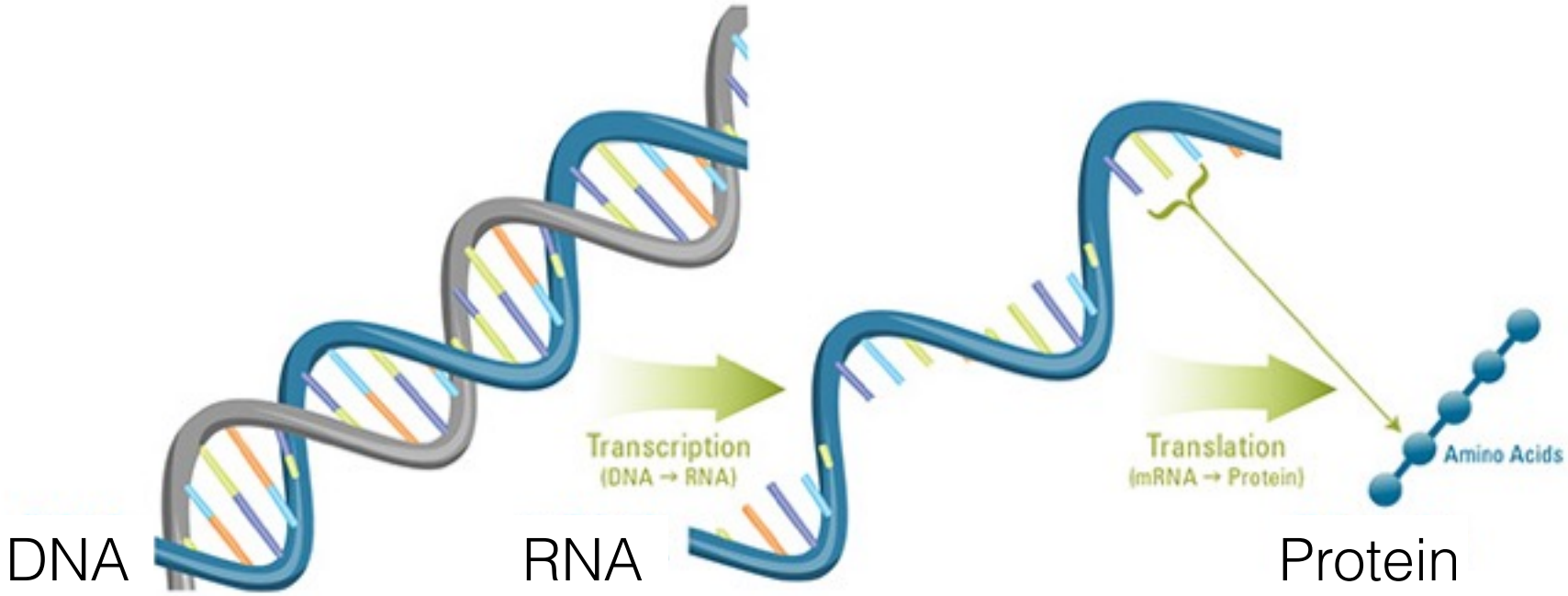
Clonal expansion

$p < 10^{-5}$



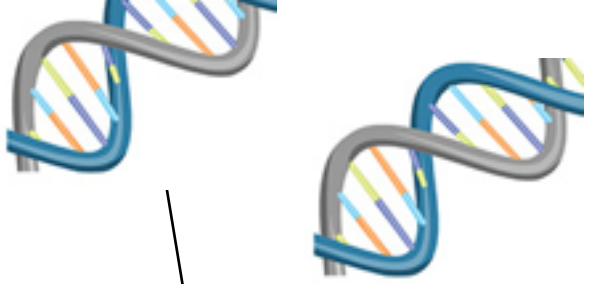
Memory T cells

How can 10^4 genes make 10^8 proteins?

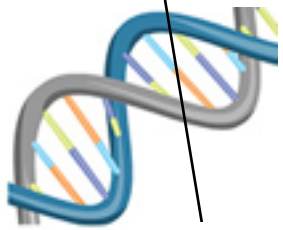


DNA makes RNA makes protein

Variable (V), diversity (D), joining (J) gene regions each containing many variants: diversity by combinatorics, and random insertions and deletions

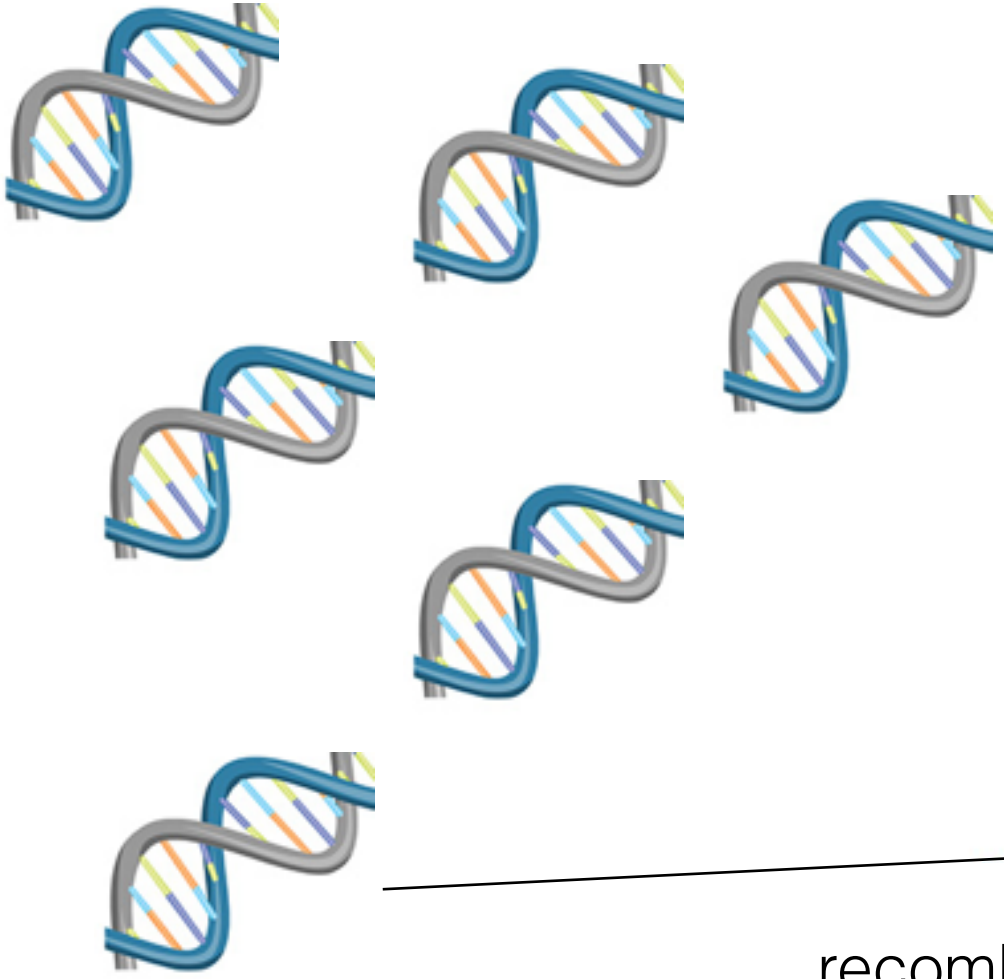


J_i



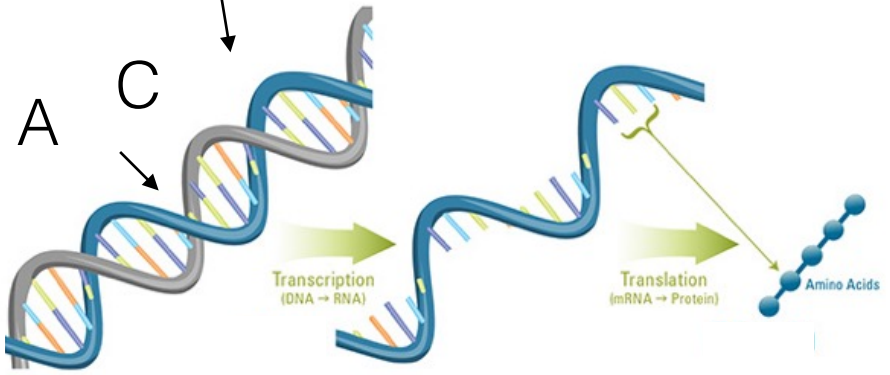
DNA fragments recombine in thymus

V_i



A

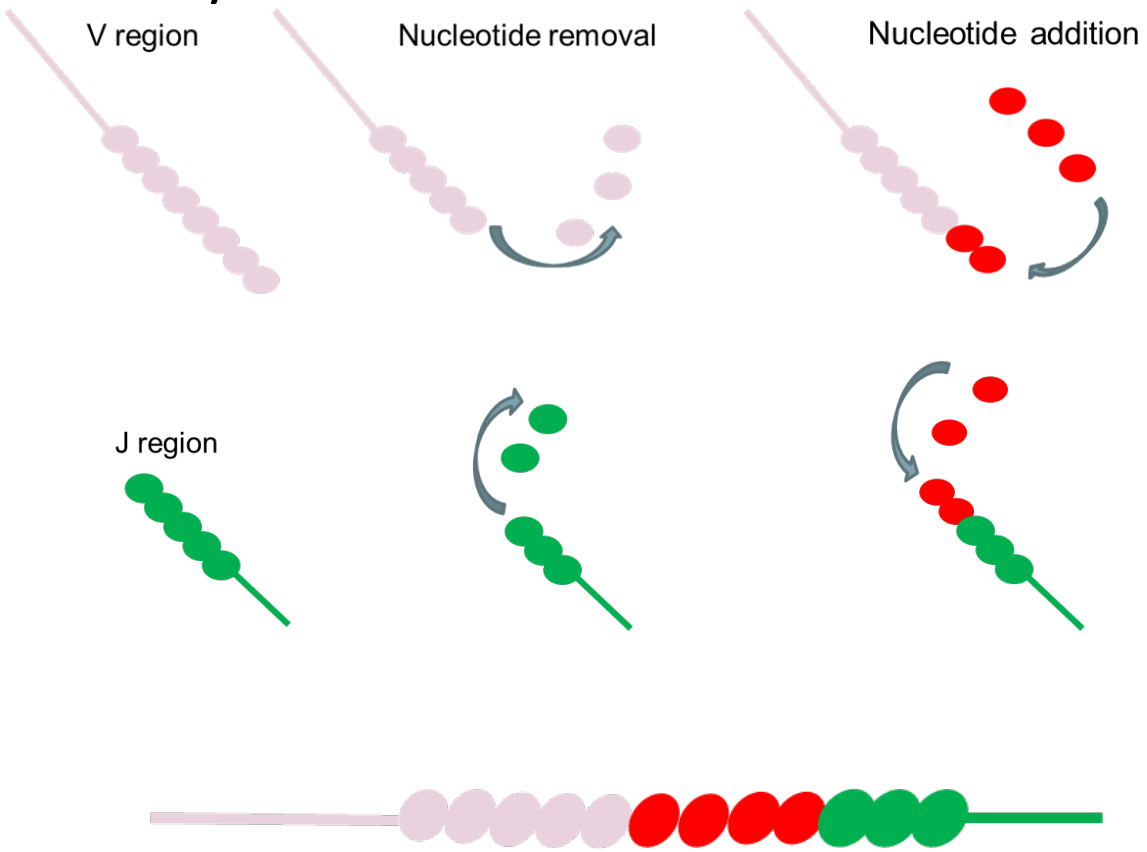
C

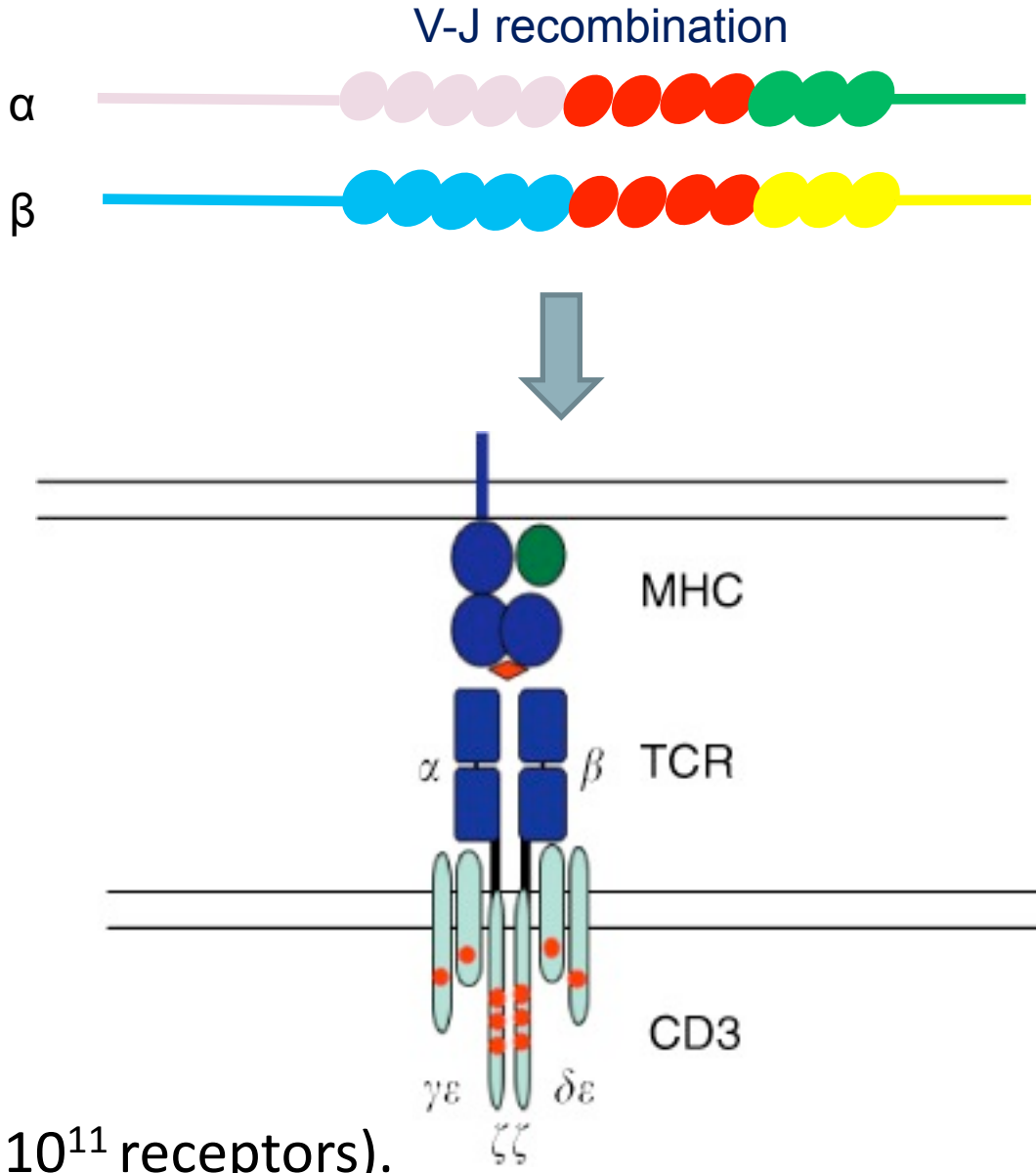


recombined DNA RNA Protein

Formation of T cell receptors: reshuffling of gene segments

one out
of many





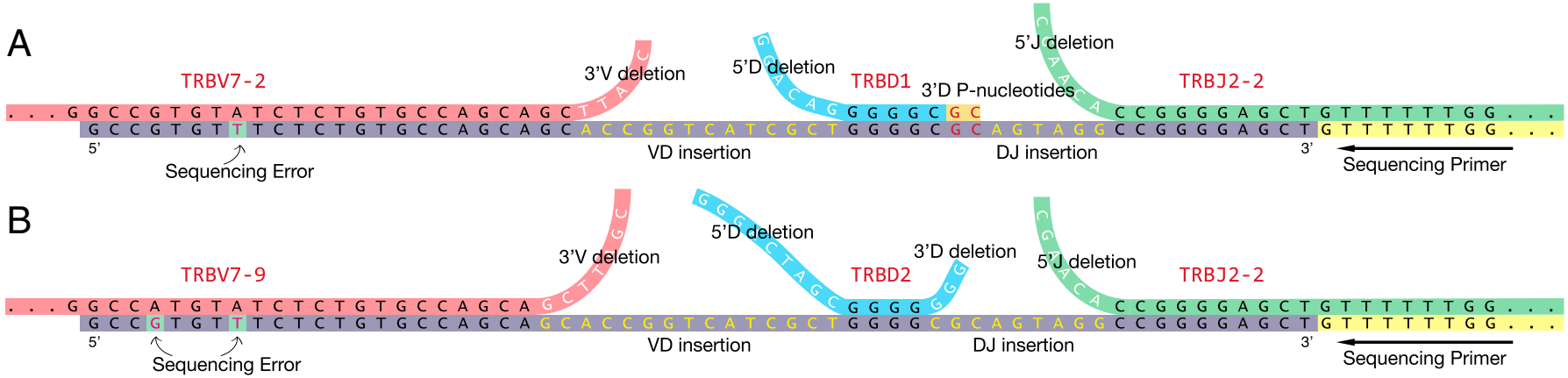
Benny Chain

The naive repertoire is extremely diverse ($10^8 < R < 10^{11}$ receptors).

Statistical inference of the generation probability of T-cell receptors from sequence repertoires

Anand Murugan[a], Thierry Mora[b], Aleksandra M. Walczak[c], and Curtis G. Callan, Jr.[a,d,1]

PNAS | **October 2, 2012** | vol. 109 | no. 40 | **16161–16166**



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

High-throughput immune repertoire analysis with IGoR

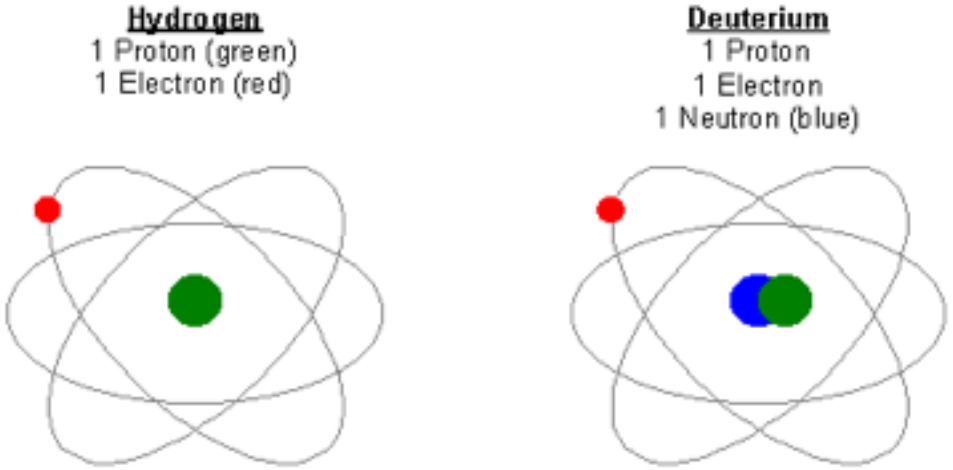
Quentin Marcou[1], Thierry Mora[2] & Aleksandra M. Walczak[1]

nature COMMUNICATIONS | (2018)9:561

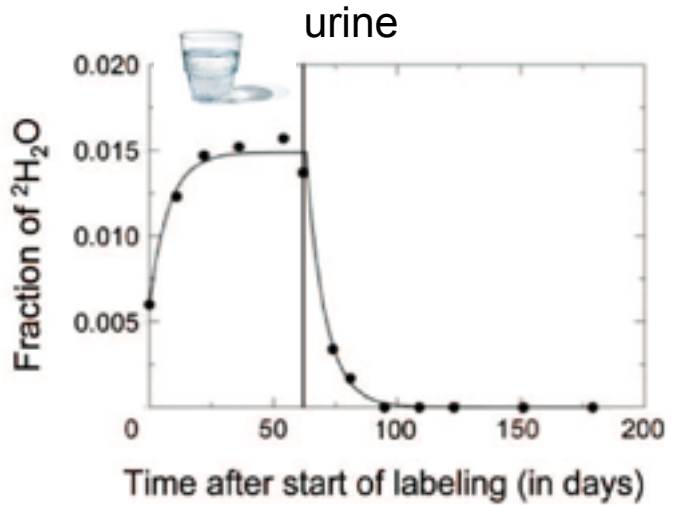
How is this huge diversity maintained in an "ecosystem" of so many competing populations?

- All naive T cells basically compete for a single resource (IL-7) -> exclusion
- Naive T cells require contacts with cognate self-antigen -> niche differentiation
- At young age populations are maintained by immigration (from the thymus) but this source vanishes after puberty -> late exclusion
- Diversity of TCRs in young and elderly people differs "only" 2-fold
- The time scale of the competitive exclusion depends on cellular lifespans
- Naive T cells are long-lived (5-10 y) & memory T cells short-lived (6 mo).
- BTW naive and memory T cells compete for different resources.

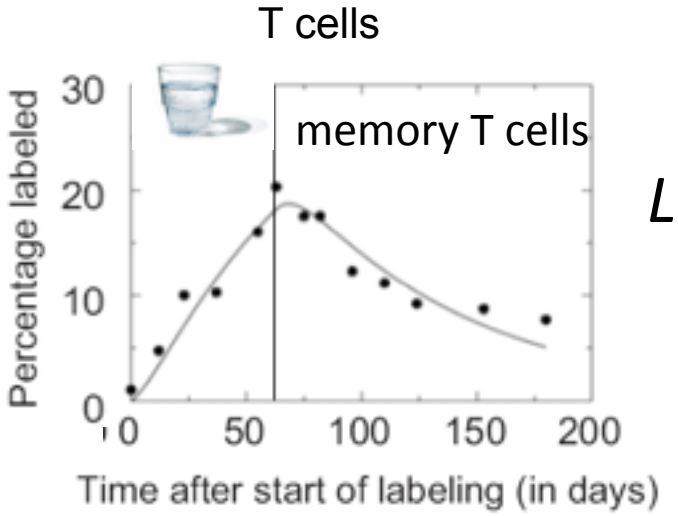
How long do T cells live in humans?



mass-spec



$A(t)$

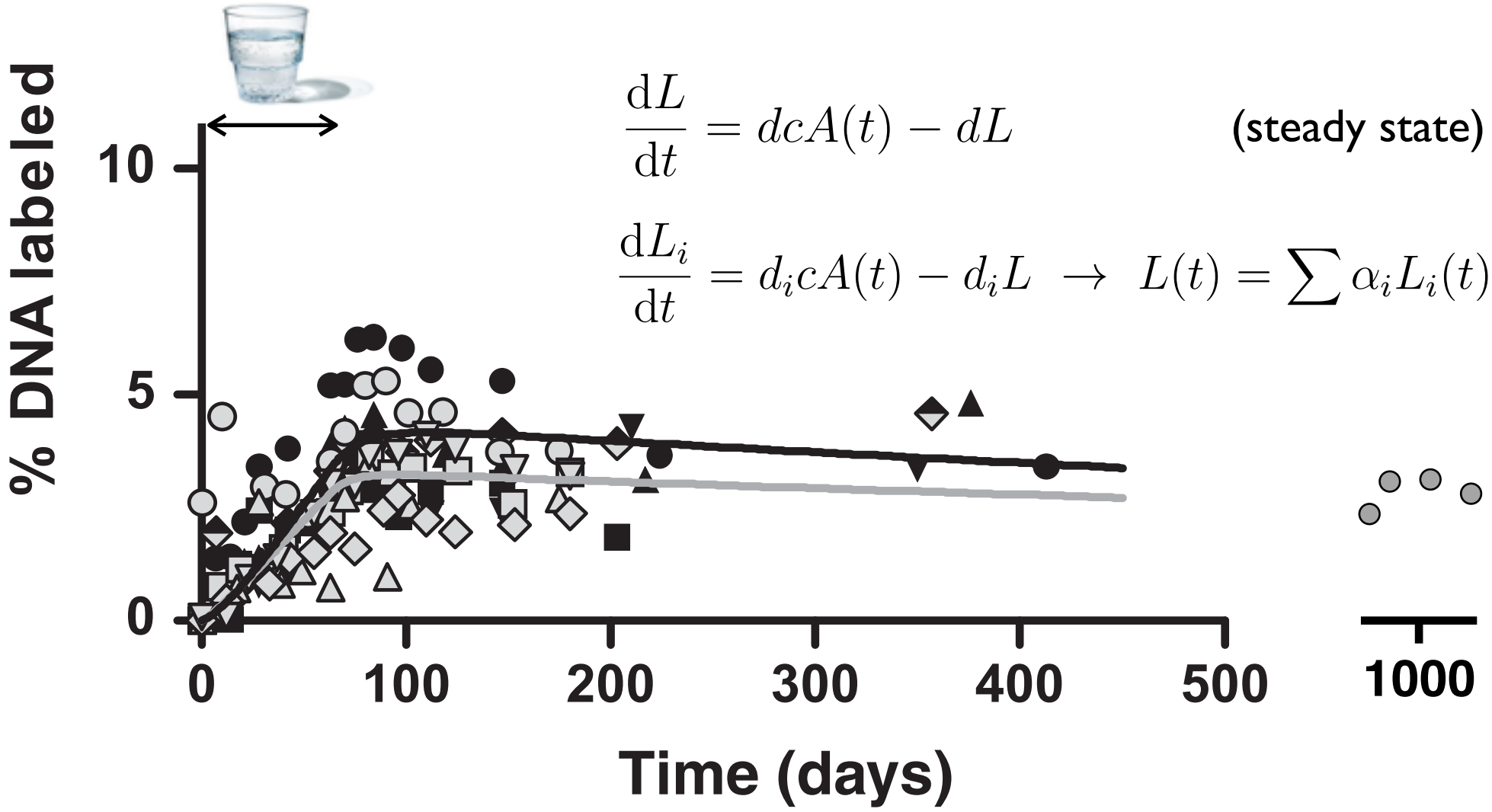


$L(t)$

Deuterium (^2H) and hydrogen are incorporated in DNA upon cell division only

José Borghans, Julia Drylewicz, Kiki Tesselaar, Nienke Vrisekoop, Ineke den Braber, Vera van Hoeven, Liset Westera

Human naive T cells have an average lifespan of 5-10 y



$$\frac{\mathrm{d}L}{\mathrm{d}t} = dcA(t) - dL \qquad \text{(steady state)}$$

$$\frac{\mathrm{d}L_i}{\mathrm{d}t} = d_i cA(t) - d_i L \;\rightarrow\; L(t) = \sum \alpha_i L_i(t)$$

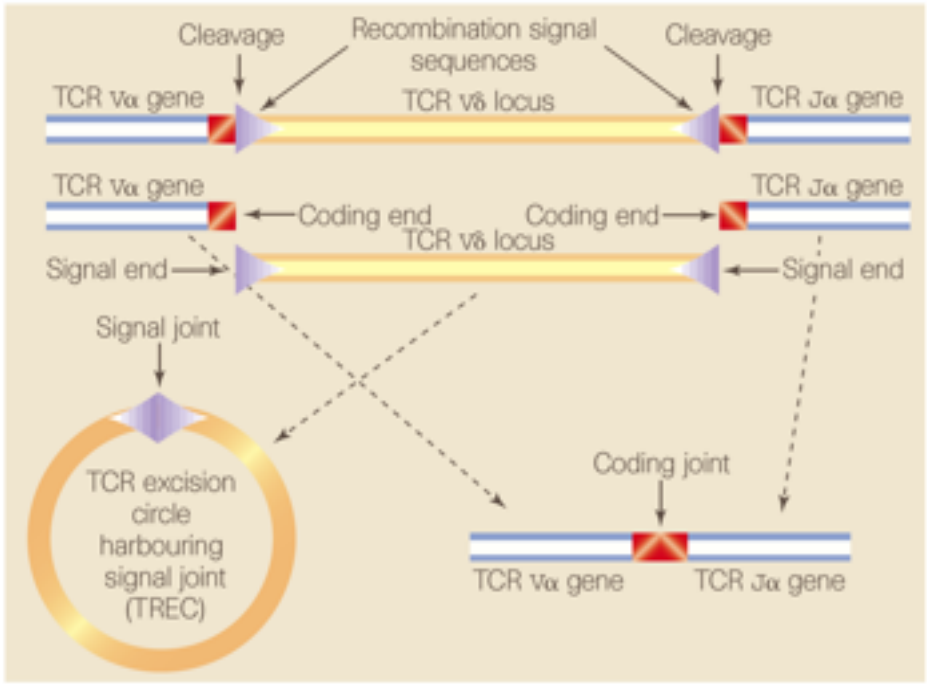
Vrisekoop et al PNAS 2006, AIDS 2015,
Westera, Aging Cell 2015.

Aged (■) and young (■) volunteers drink deuterated water for 8-9 weeks.
By mass spec we track the enrichment in DNA of naive T cells. Modeling translates this into a life span

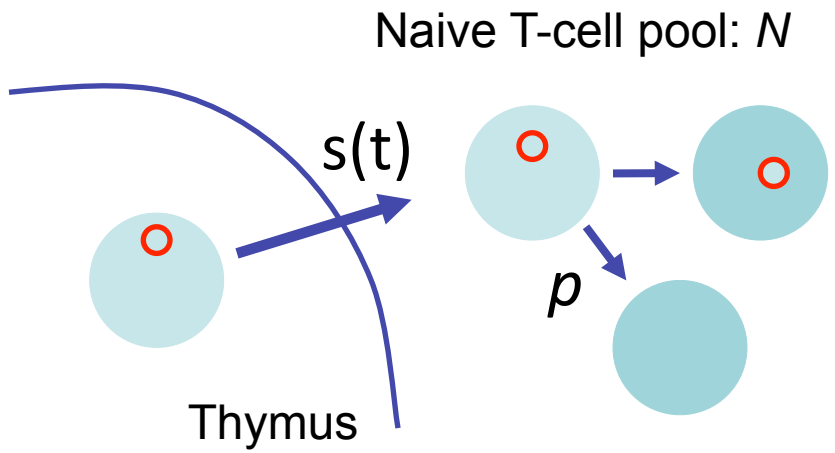
Estimate thymic output by measuring fraction of T cells with a T-cell receptor excision circle (TREC)

TREC is a DNA circle produced when the TCR re-arranges.

TRECs not duplicated upon division.



Rodewald Nature 1998



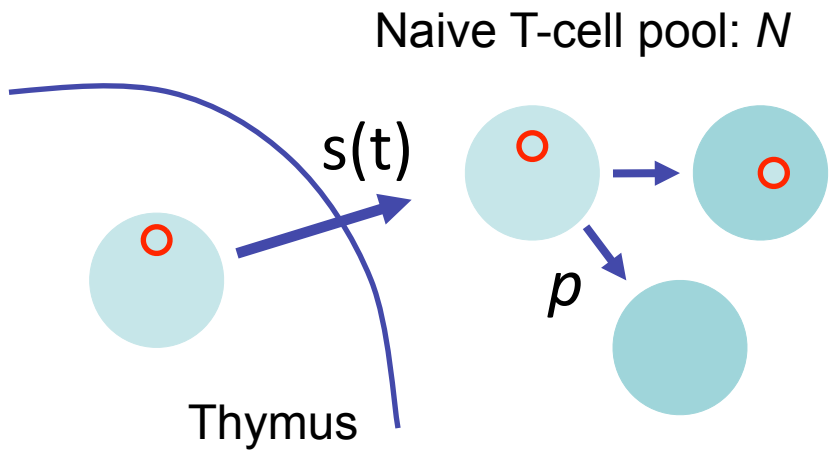
TREC is a marker for a cell originally produced in the thymus

(after normalization)

$$\frac{A}{c} = \frac{s(t)}{s(t) + pN(t)}$$

TRECs are long-lived:
in humans they persist for decades after thymectomy

$$\frac{\mathrm{d}N}{\mathrm{d}t} = s(t) + (p-d)N \ ,$$
$$\frac{\mathrm{d}T}{\mathrm{d}t} = cs(t) - dT \ ,$$



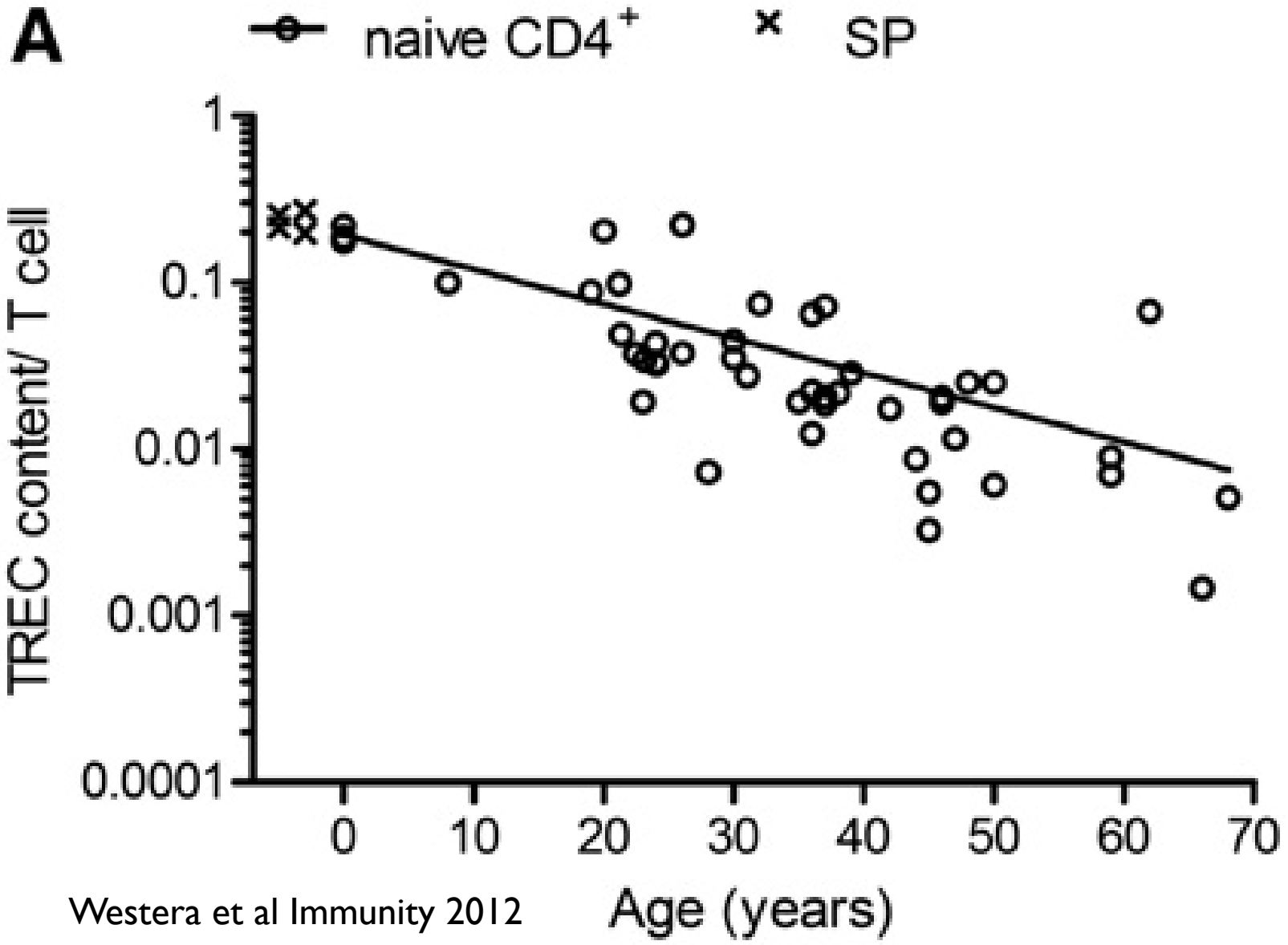
TREC is a marker for a cell originally produced in the thymus

(after normalization)

Define $A \equiv T/N$:

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{s(t)}{N}(c-A) - pA = 0 \ ,$$
$$\frac{A}{c} = \frac{s(t)}{s(t)+pN(t)} \ .$$

Thymus accounts for <20% of the production of naive T cells in young humans adults and for <2% in healthy elderly



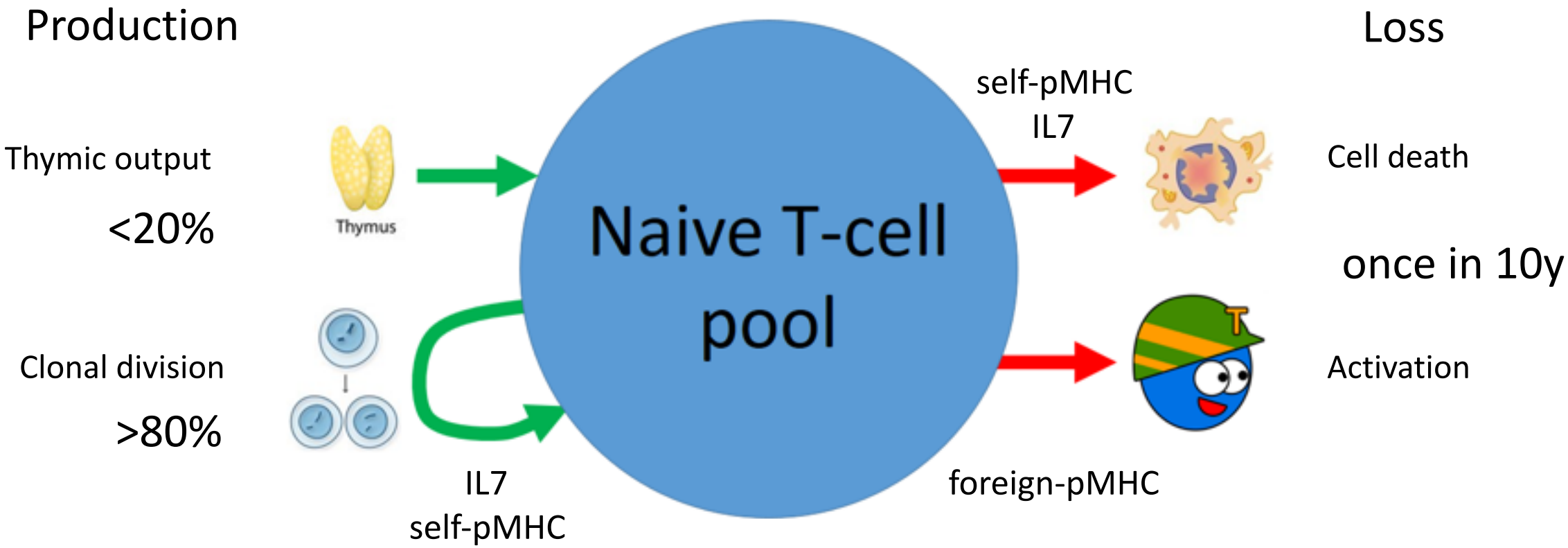
Westera et al Immunity 2012

$$\frac{A}{c} = \frac{s(t)}{s(t) + pN(t)}$$

Donna Farber:
s(t)=0 when *t* > 40y
Review in Immunity
2018

naive T cells divide

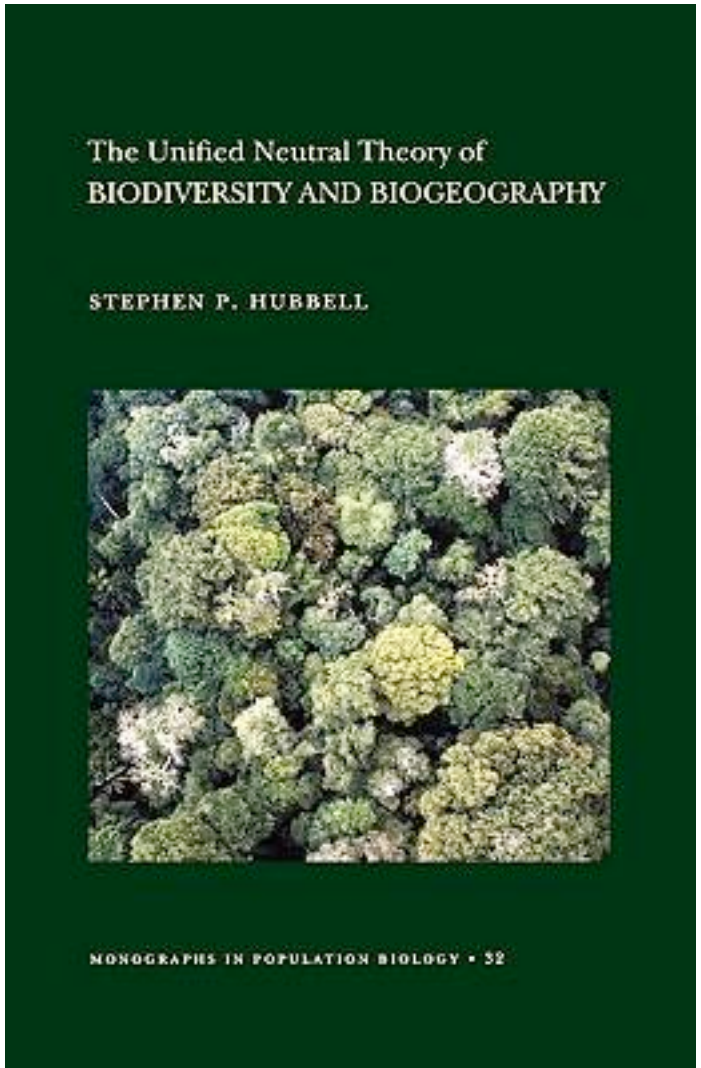
Consider a highly diverse naive T cell pool in which thymic output is the only source of new clonotypes

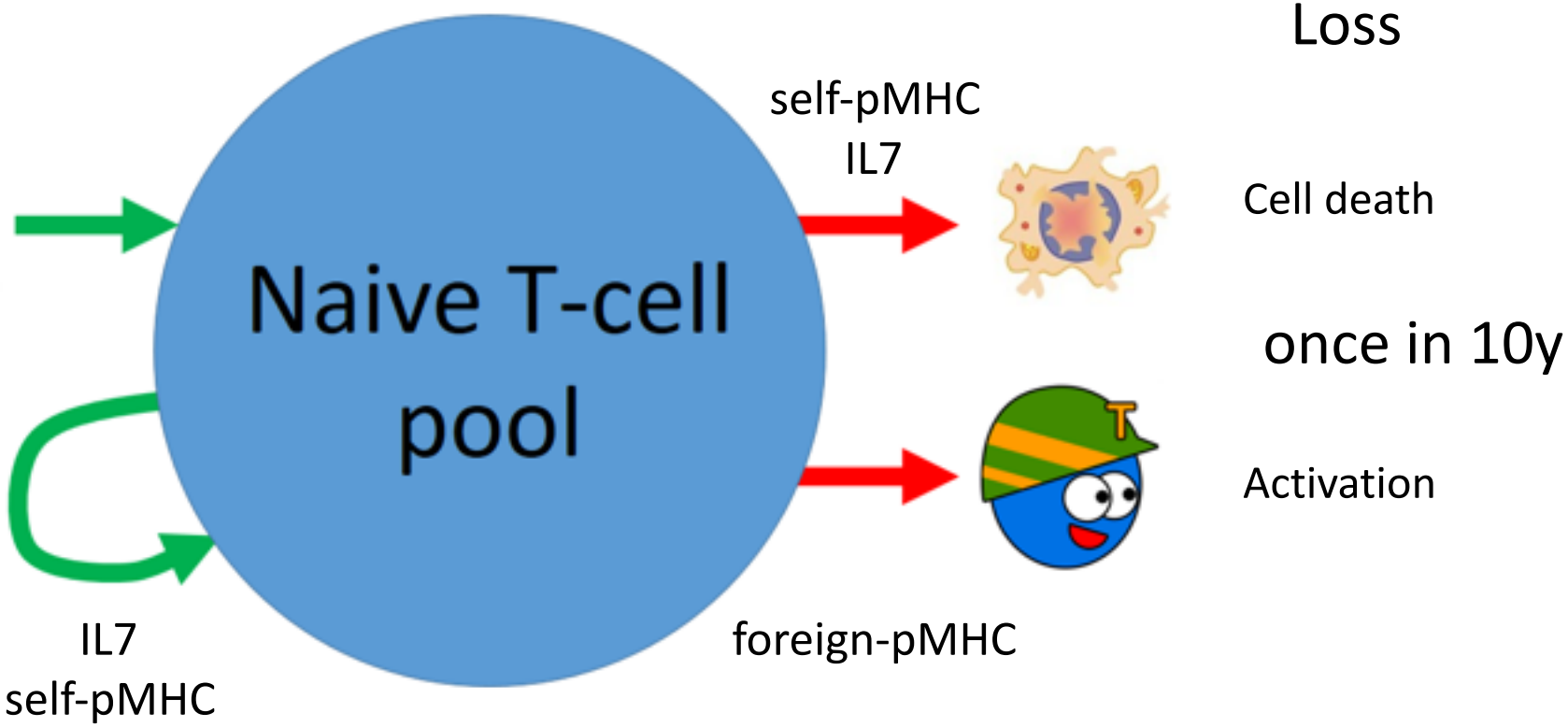


Birth, death & immigration model

wisegeek.org, bioninja.com.au, daviddarling.info, Immunology insight

Let's start with a "neutral" model where all populations have the same division and death rates





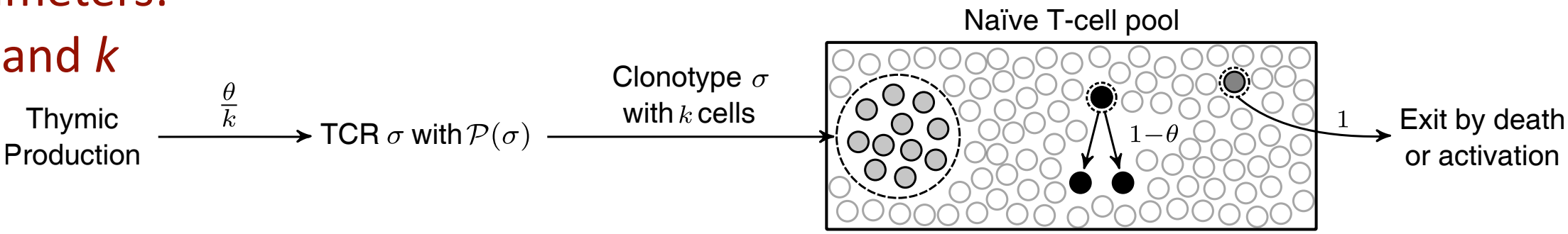
$N = 10^{11}$ cells

wisegeek.org, bioninja.com.au, daviddarling.info, Immunology insight

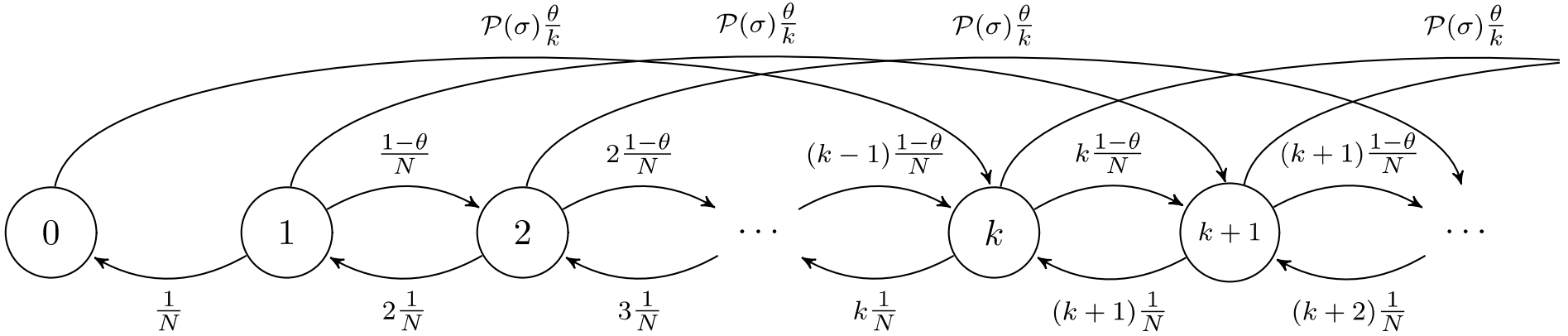
A **BDI** model for naive T-cell dynamics

Two (known) parameters: θ and *k*

Event driven dynamics on the level of the full pool: remove a single cell and replace with a new one coming from the thymus (θ) or a division event (1-θ)



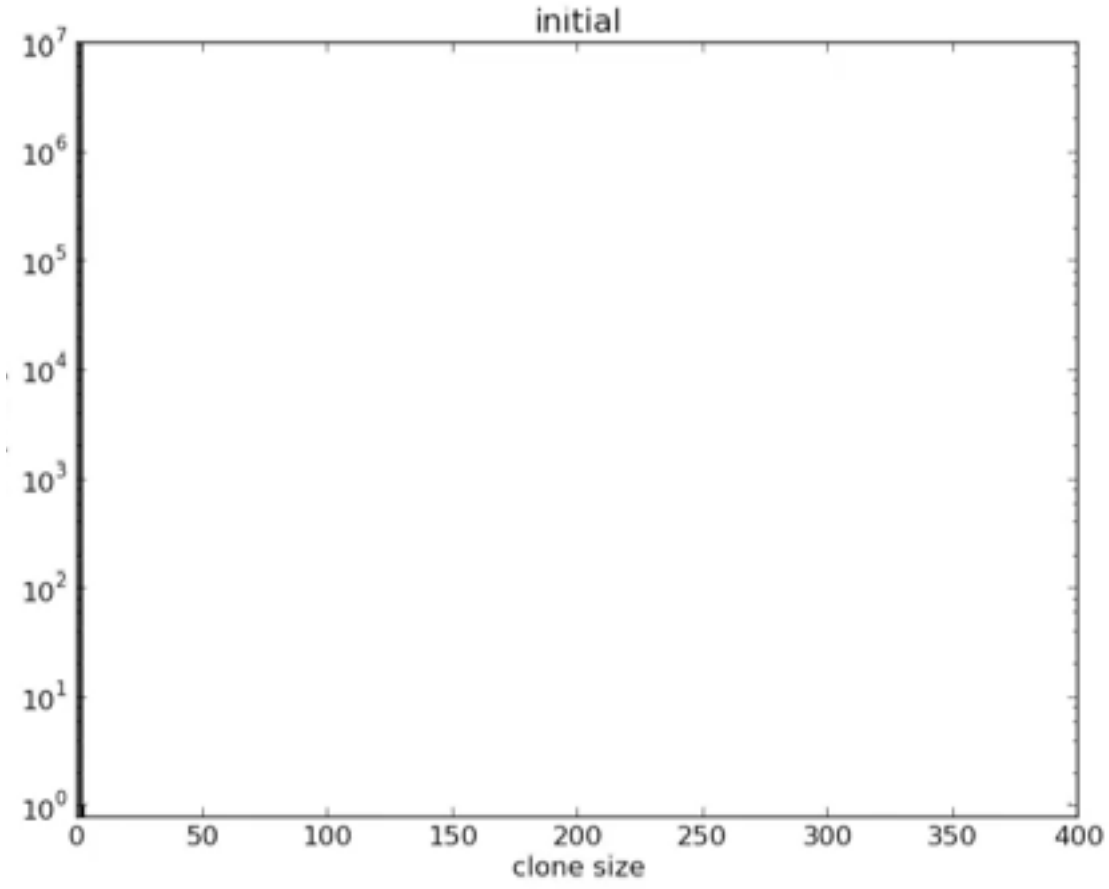
Markov-chain of a single clonotype



Simulate a whole mouse of 10^7 naive T cells: Clone size distribution approaches steady state

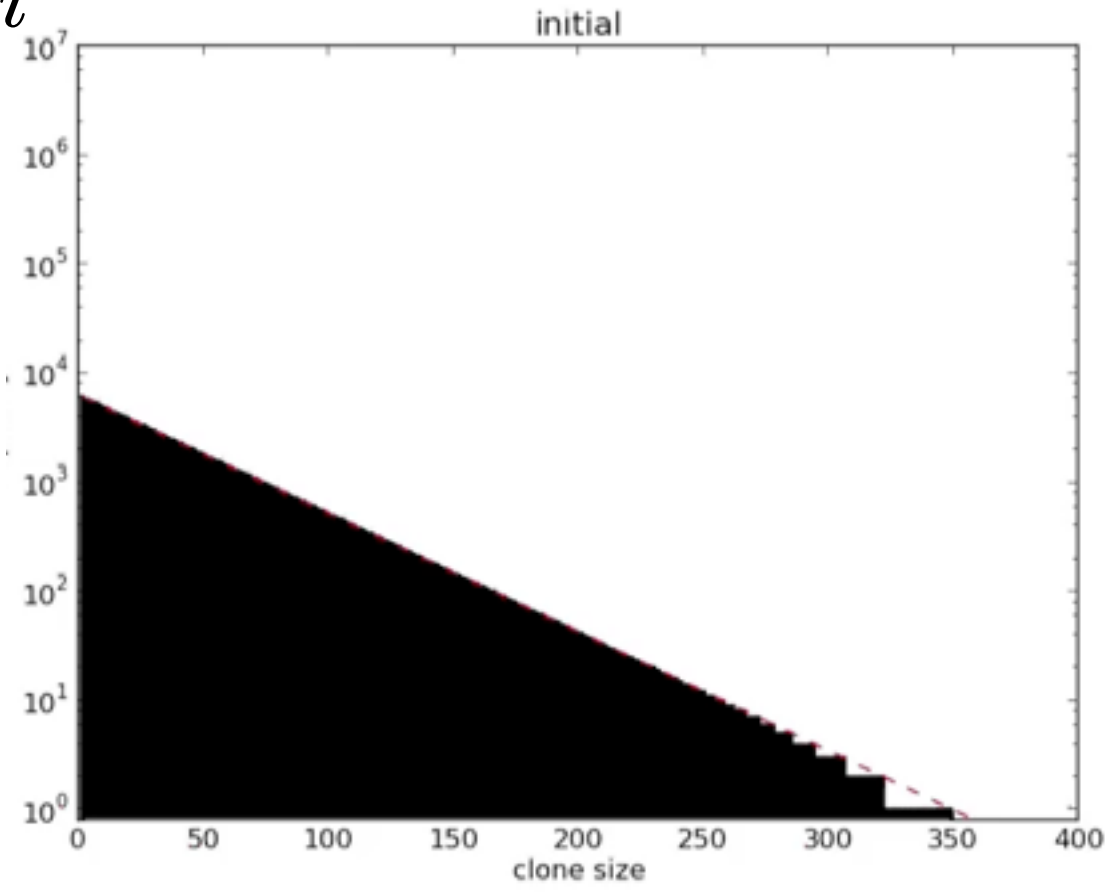
F_i

Initial condition:all cells of clone size one



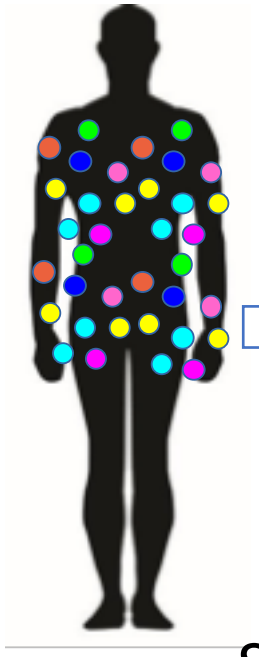
F_i

Initial condition: exponential distribution

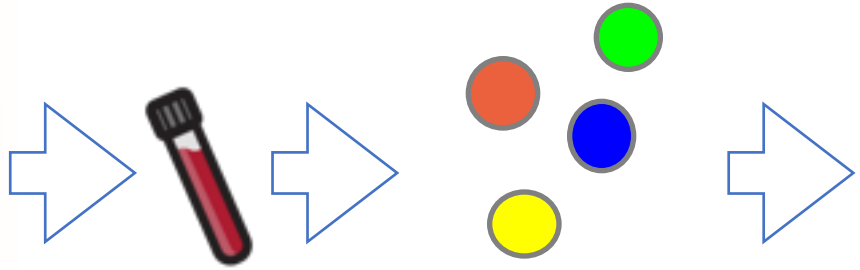


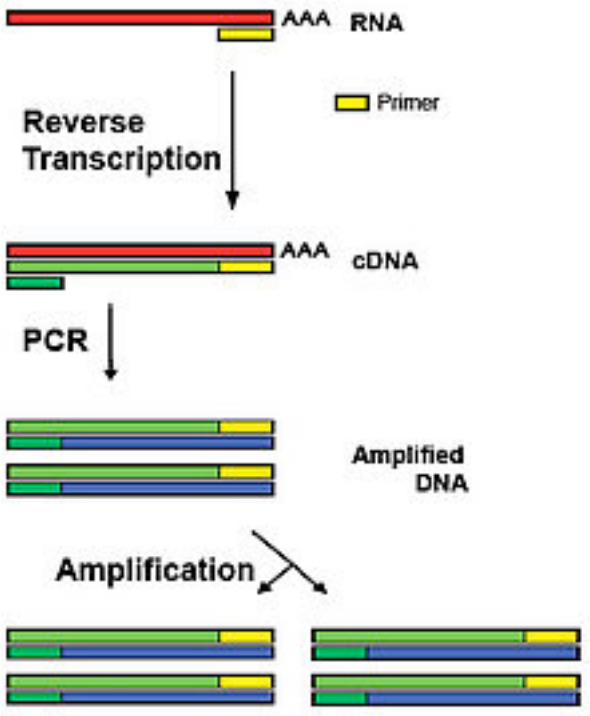
N = 10^7 cells; θ = 0.1;k=1; 10^9 events (θ is a humanized choice here)

Now the data



10^{-5}





Small sample: sort for T cells

Amplify TCR mRNA

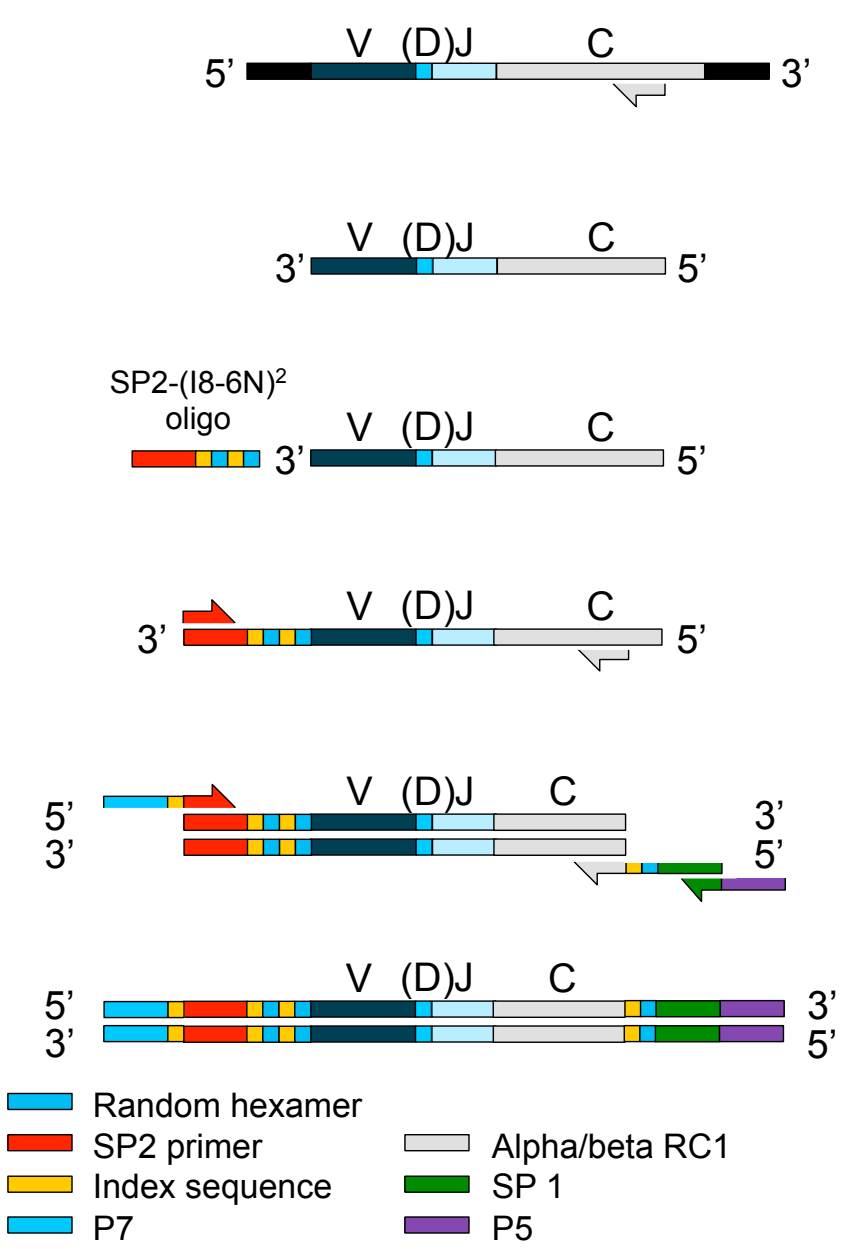
Sequence

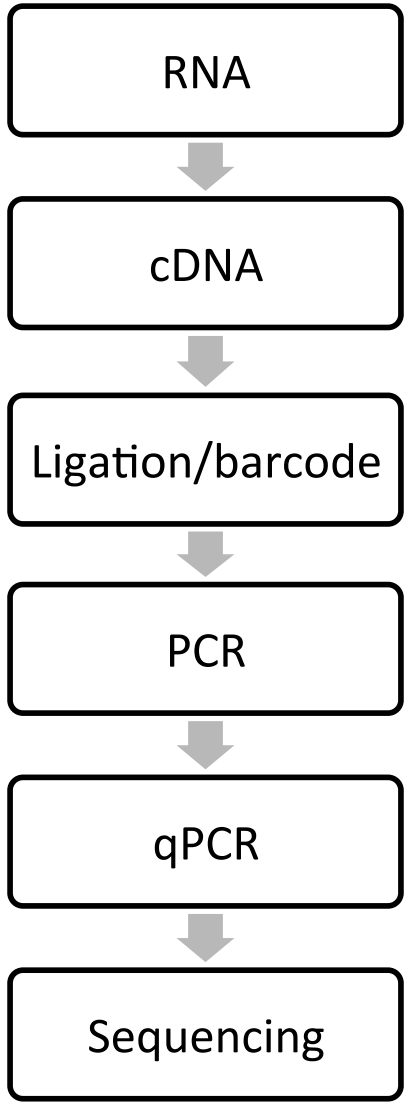
Error correction algorithms

Interpretation
by modeling

	A	B	C	D	E	F	G
1	tag	idx01	idx02	idx03	idx04	idx05	idx06
2	TRCN0000000604:CSNK1E	2901	1538	4868	353	2038	1093
3	TRCN0000000606:CSNK2A1	5784	14299	12627	2693	1039	2662
4	TRCN0000000610:CSNK2A1	6528	24228	3848	15010	5418	7627
5	TRCN0000000686:CDK10	4843	15585	5482	6319	8741	16349
6	TRCN0000000688:CDK10	97	2149	3619	736	6326	5253
7	TRCN0000001720:DAPK2	25228	9064	21827	7774	34162	27729
8	TRCN0000001721:DAPK2	0	0	0	0	0	1593
9	TRCN0000001722:DAPK2	1	0	0	0	0	1533
10	TRCN0000001723:NEK6	0	0	135	0	0	358
11	TRCN0000001724:NEK6	2761	0	1757	0	4047	744
12	TRCN0000001725:NEK6	3368	338	0	5334	1018	3194
13	TRCN0000001726:NEK6	0	3476	0	291	756	0
14	TRCN0000001727:NEK6	10132	0	4149	1679	0	0
15	TRCN0000001734:MAST2	8920	9254	2	0	7290	2021
16	TRCN0000001735:MAST2	0	0	3994	6689	10814	2115
17	TRCN0000001736:MAST2	0	3468	9026	5277	15668	15223
18	TRCN0000001742:STYK1	0	0	0	9308	7	0
19	TRCN0000001744:STYK1	0	1	4	3	4	3
20	TRCN0000001745:STYK1	0	7	1	3	6	3
21	TRCN0000001746:STYK1	14413	1878	554	3387	0	2338
22	TRCN0000001747:PAK6	8560	4091	0	0	12879	10761
23	TRCN0000001751:PAK6	0	0	1966	372	0	213

PCR amplification protocol

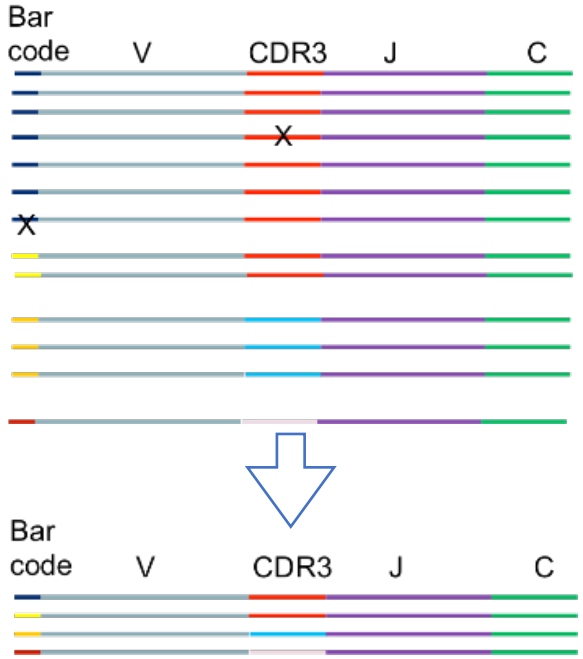




Benny Chain

Bioinformatics

Each cDNA is identified with a “barcode”: UMI

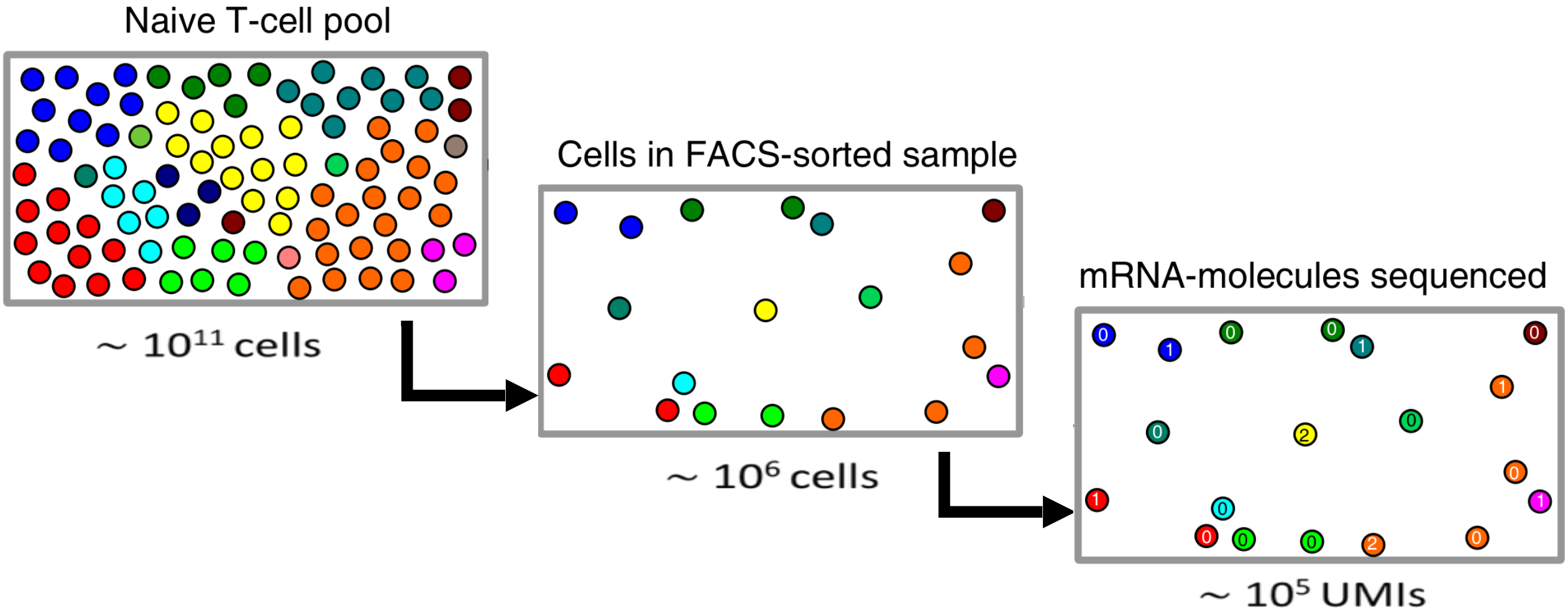


RTCR: a pipeline for complete and accurate recovery of T cell repertoires from high throughput sequencing data

Bram Gerritsen[1,*], Aridaman Pandit[1], Arno C. Andeweg[2] and Rob J. de Boer[1]

Bioinformatics, 32(20), 2016, 3098–3106
doi: 10.1093/bioinformatics/btw339

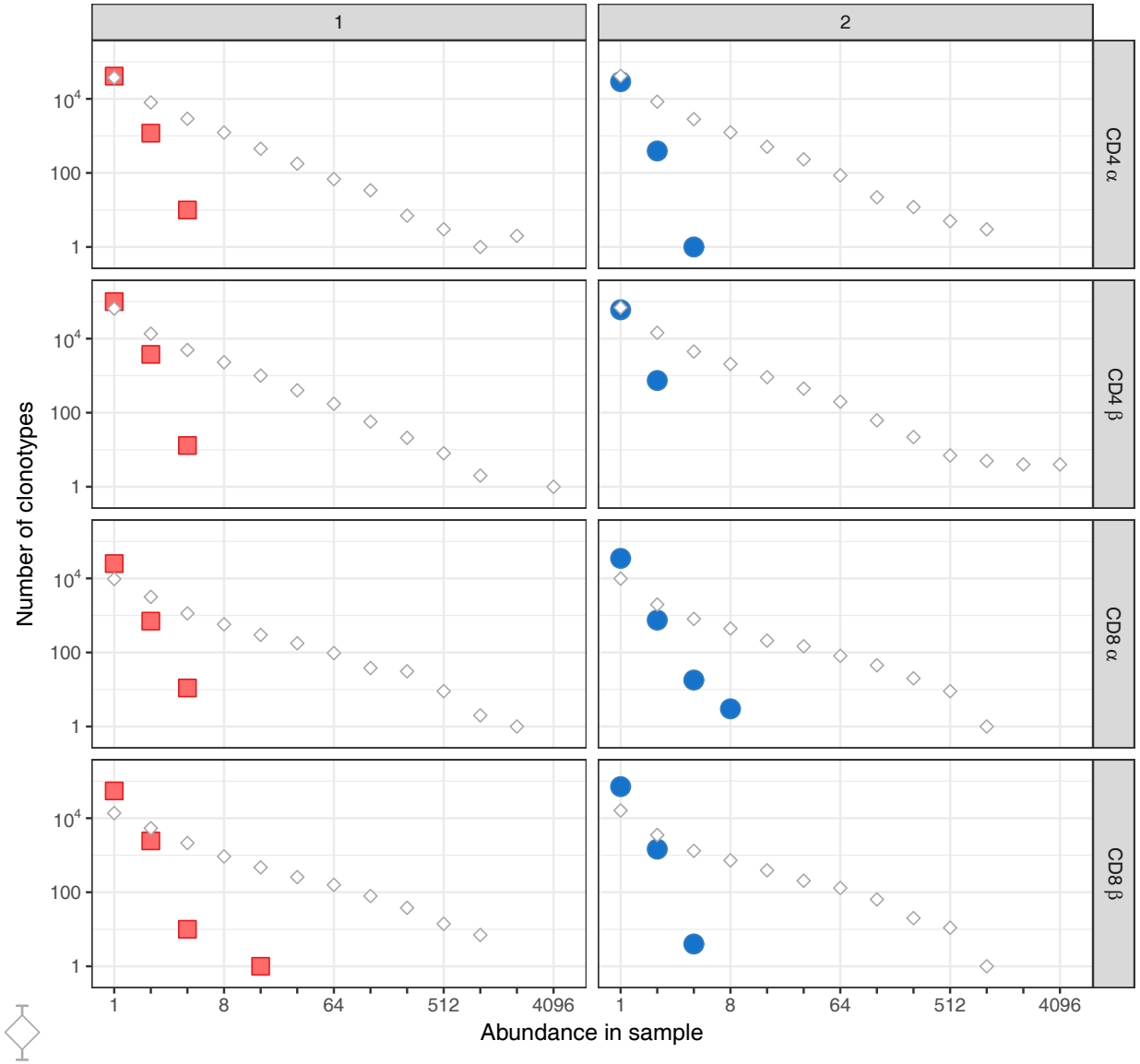
Sequencing of TCR mRNA involves two sampling steps



A few cells may contribute several mRNAs and then seem to represent large clonotypes

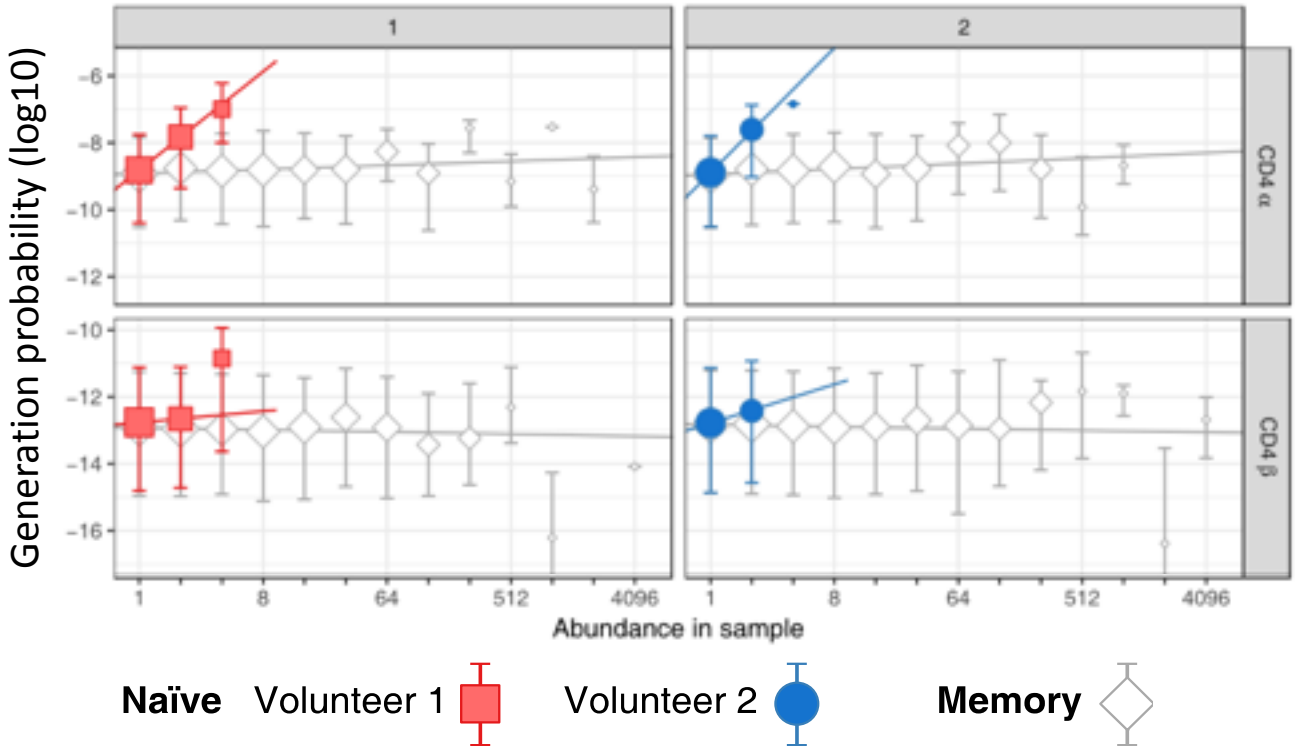
TCRA and *TCRB* mRNA from naive CD4+ and CD8+ T cells sampled from blood in two healthy adult volunteers

- 4 data sets (A & B chain x CD4 and CD8) of about 10^5 reads (Illumina MiSeq)
- correct for impurities by removing reads overlapping with memory cells
- reads are error corrected by UMIs and by correcting UMIs



Large naive clonotypes have high production probabilities

- Some TCRs are made much more easily than others
- Generation probabilities of all TCRA and TCRB sequences determined with IGoR (Marcou, Nat. Comm. 2018)



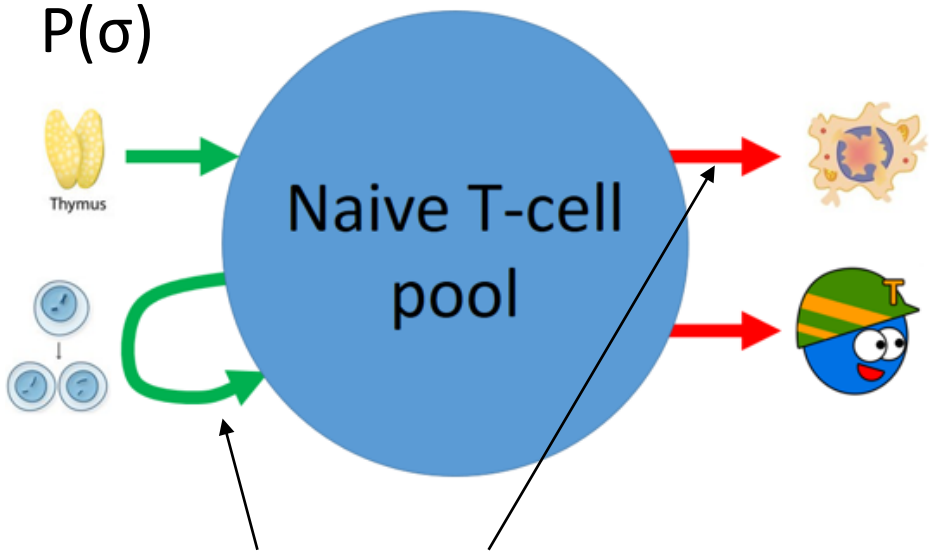
High-throughput immune repertoire analysis with IGoR

Quentin Marcou[1], Thierry Mora[2] & Aleksandra M. Walczak[1]

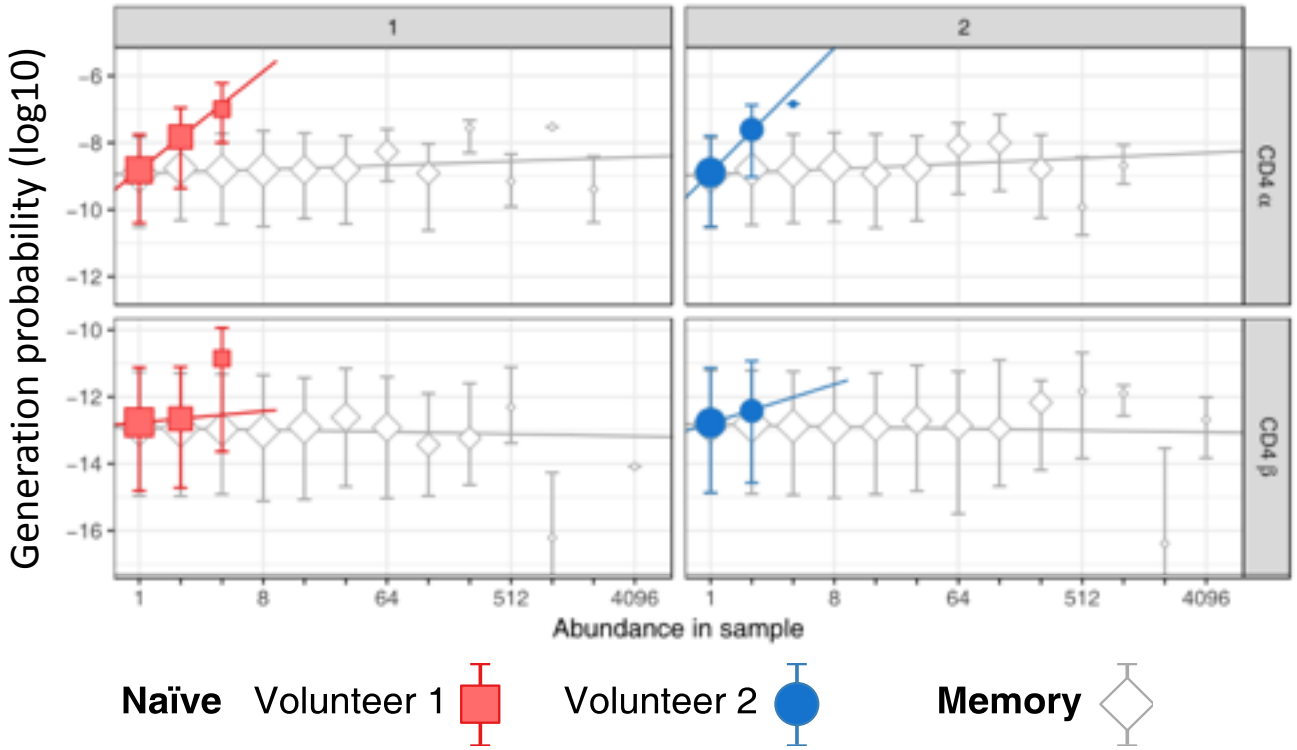
nature COMMUNICATIONS (2018)9:561

Supports neutrality: if TCR-based competition (niches) would dominate naive T-cell dynamics, one would not expect this.

To what extent can generation probabilities explain clone-sizes of naïve T-cells?



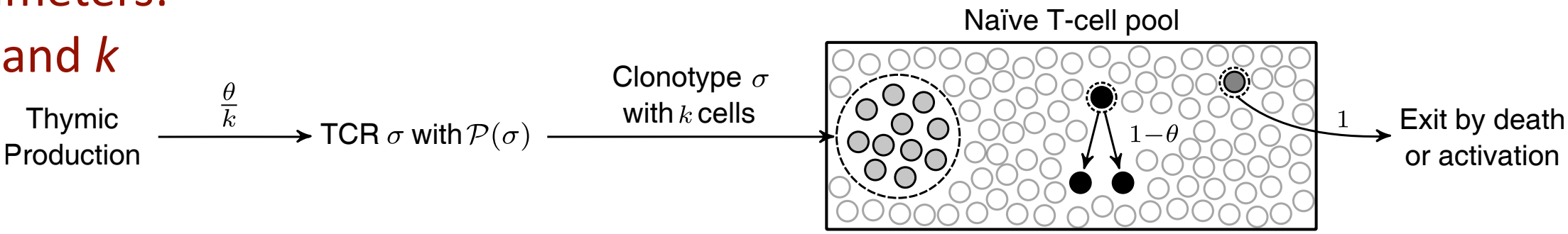
Competition for IL7 modulated by self-pMHC?



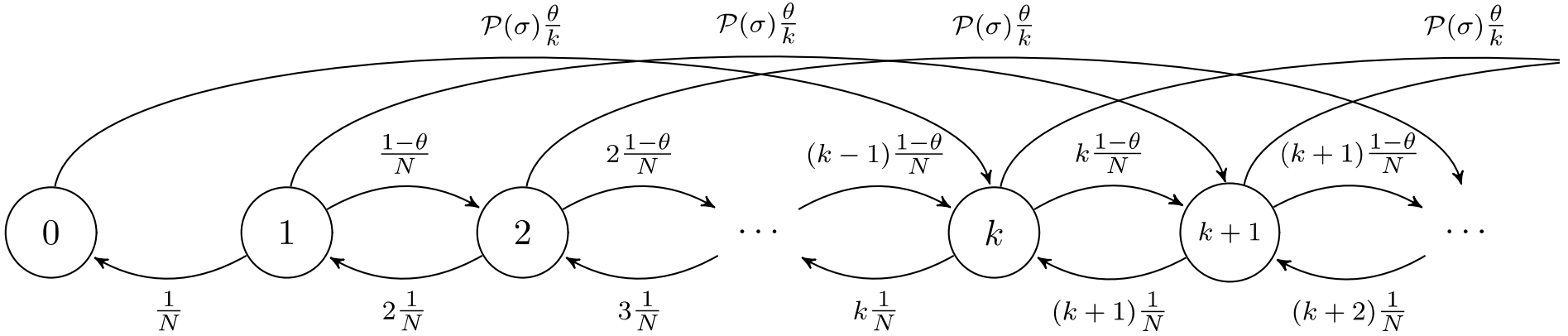
A **BDI** model for naive T-cell dynamics

Two (known) parameters: θ and *k*

Event driven dynamics on the level of the full pool: remove a single cell and replace with a new one coming from the thymus (θ) or a division event (1-θ)

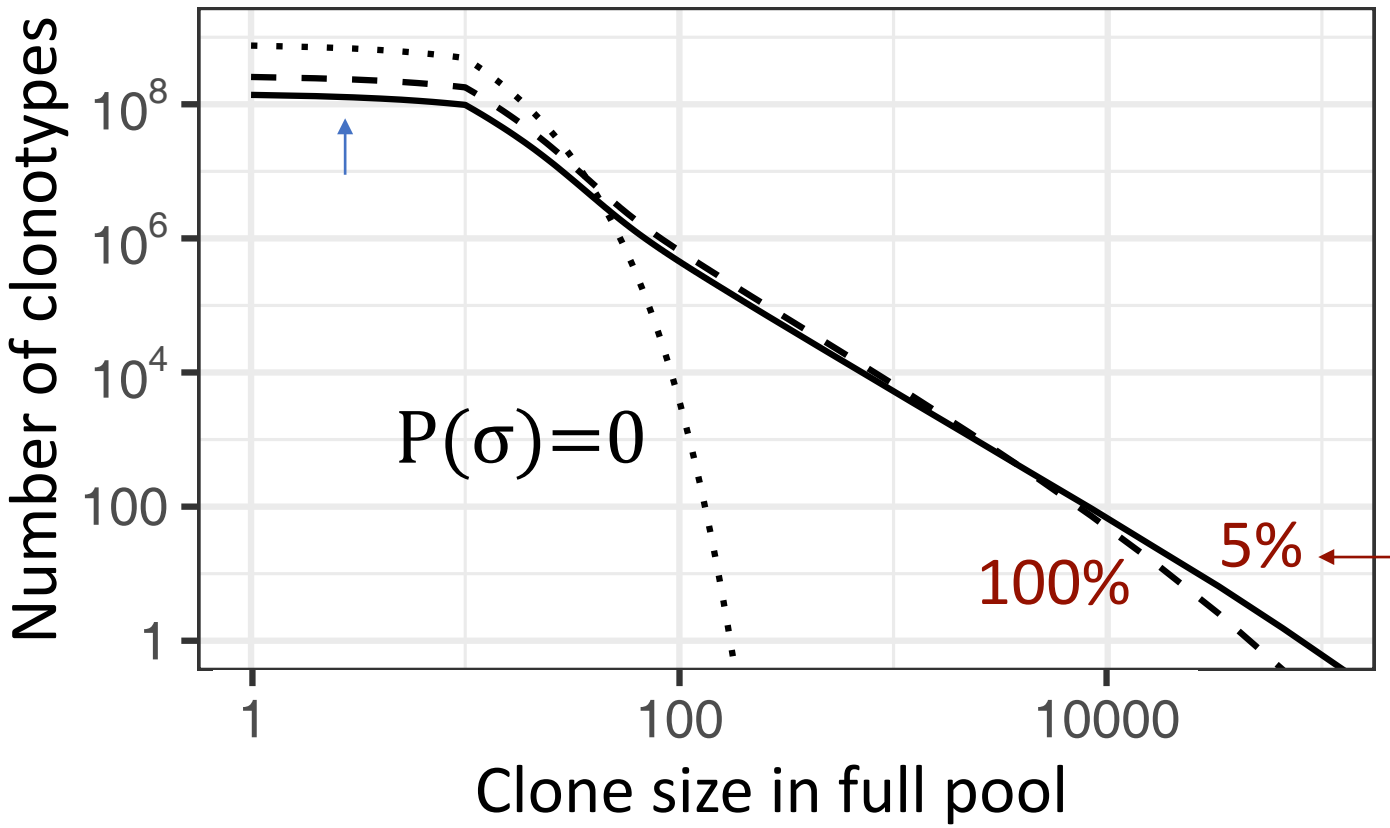


Markov-chain of a single clonotype



Steady-state solution of the model allows us to predict the full clone-size distribution

$$S_i i \frac{1}{N} = S_{i-1}(i-1)\frac{1-\theta}{N} + \mathcal{P}(\sigma)\frac{\theta}{k} \sum_{j=\max(i-k,0)}^{i-1} S_j, \quad \text{for } 1 \le i \le N$$

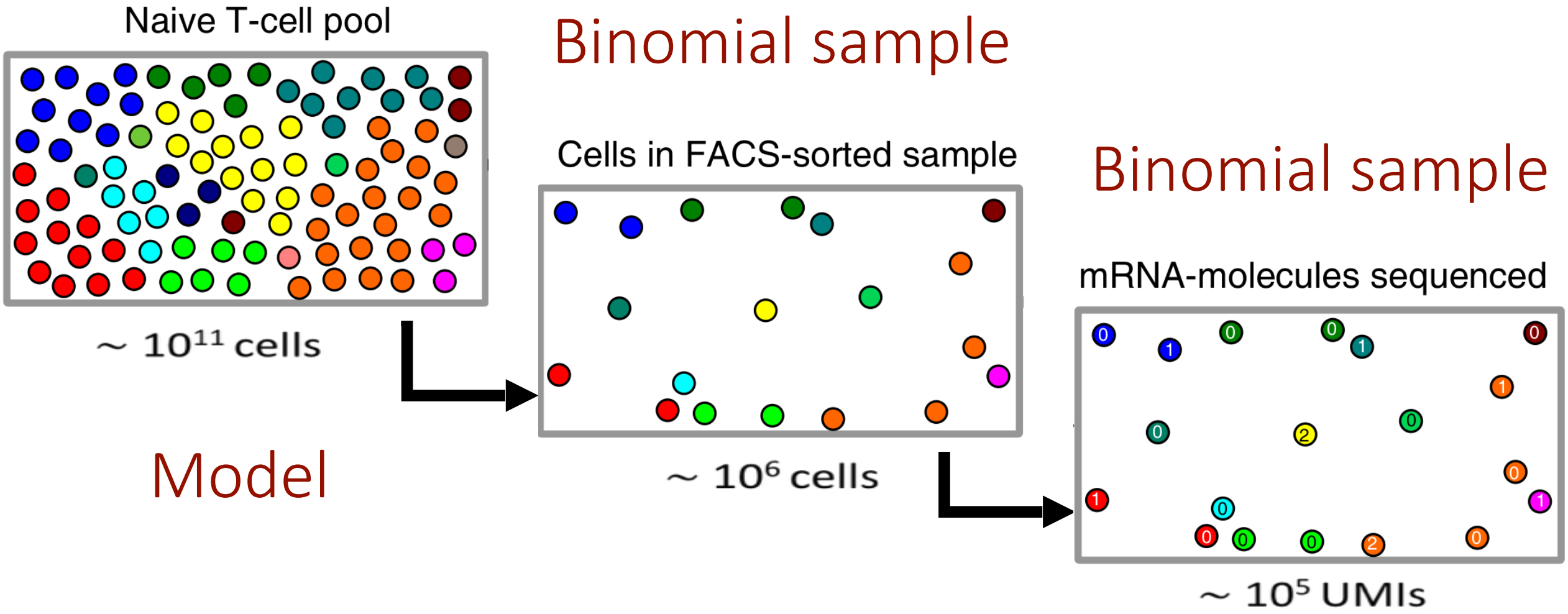


$N = 7.5 \times 10^{10}$
$\theta = 0.1$
$k = 10$

5% ← Thymic selection

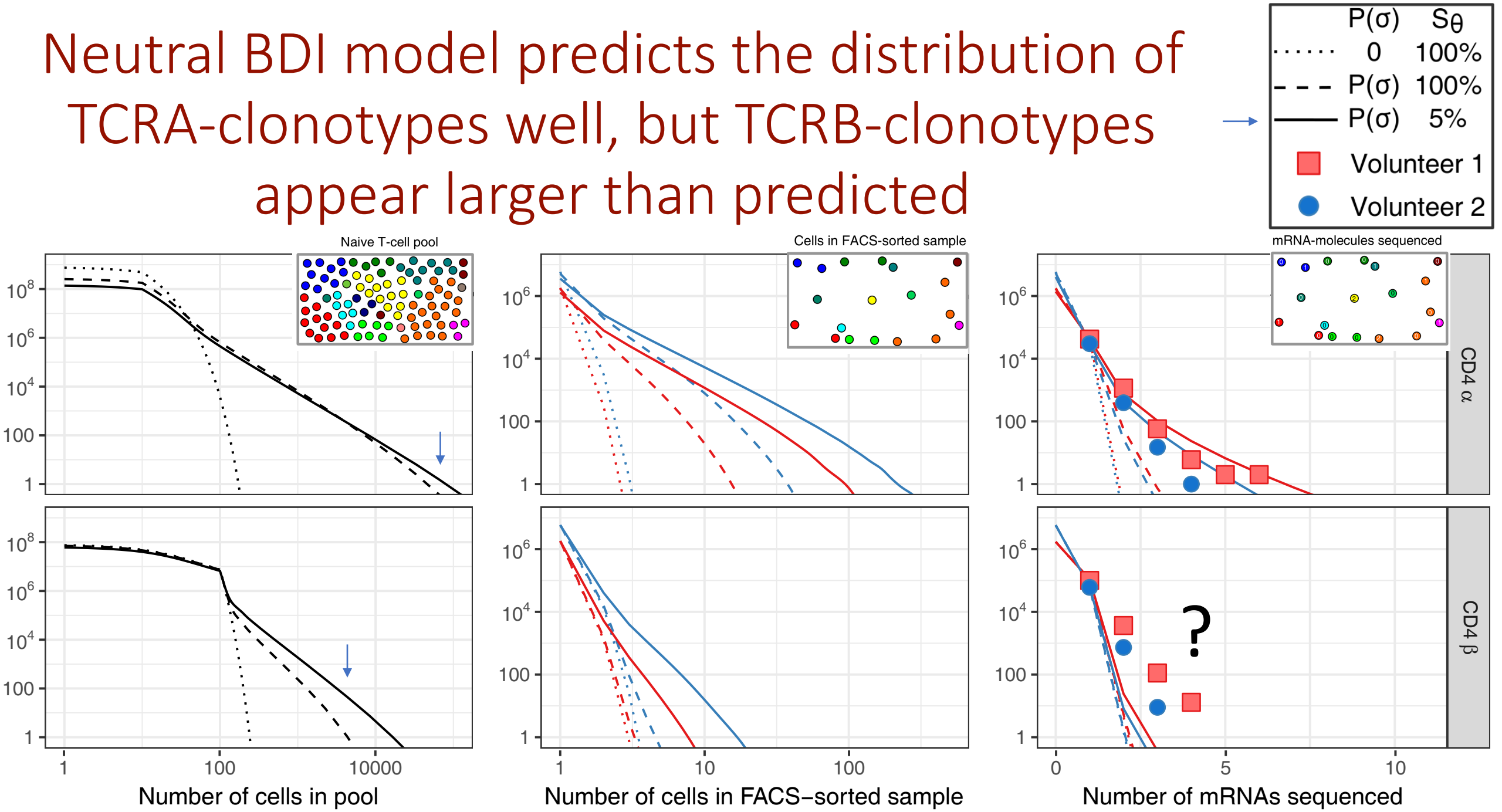
100%

Sequencing of TCR mRNA involves two sampling steps



A few cells may contribute several mRNAs and then seem to represent large clonotypes

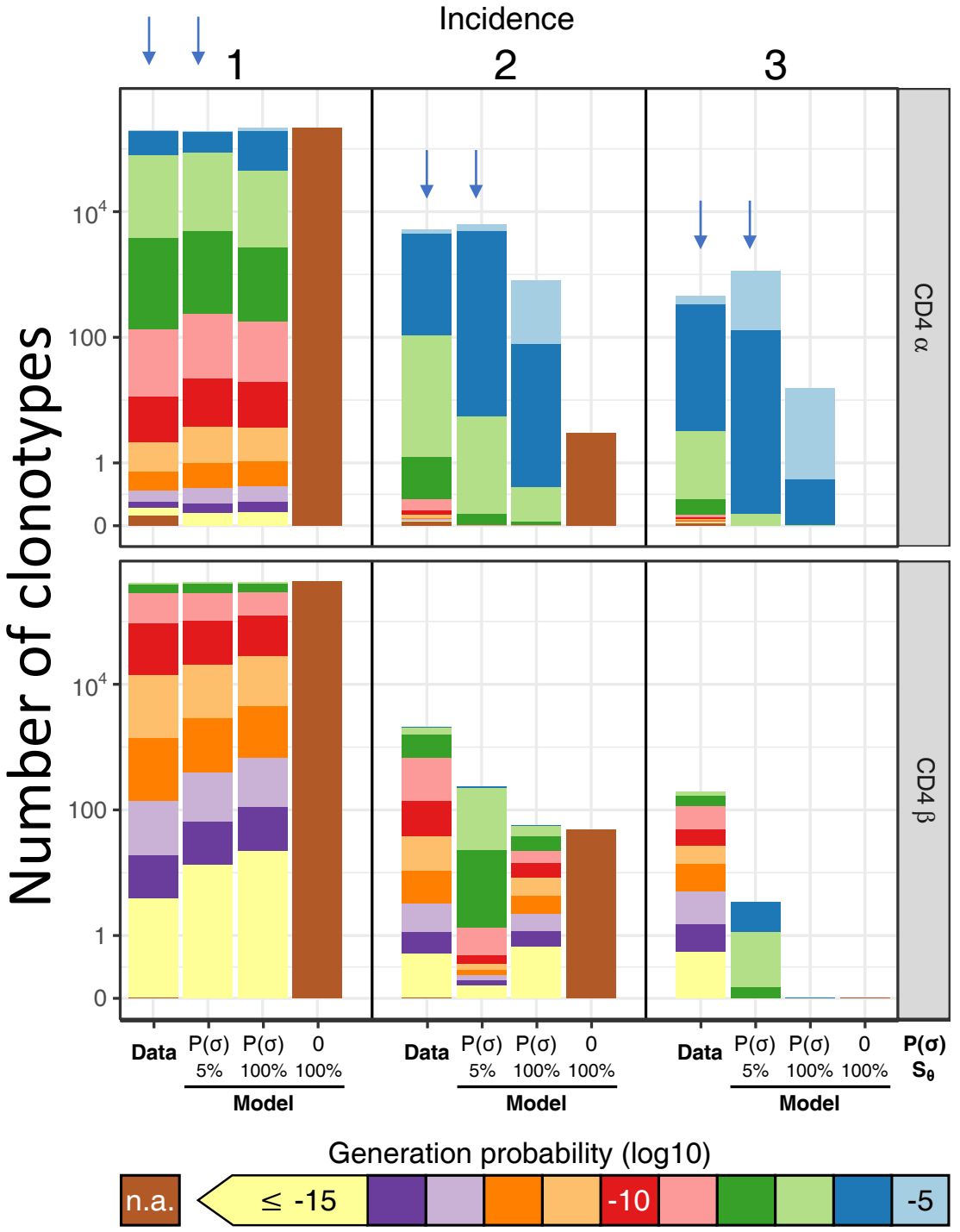
Neutral BDI model predicts the distribution of TCRA-clonotypes well, but TCRB-clonotypes appear larger than predicted



So a few clones are very large.

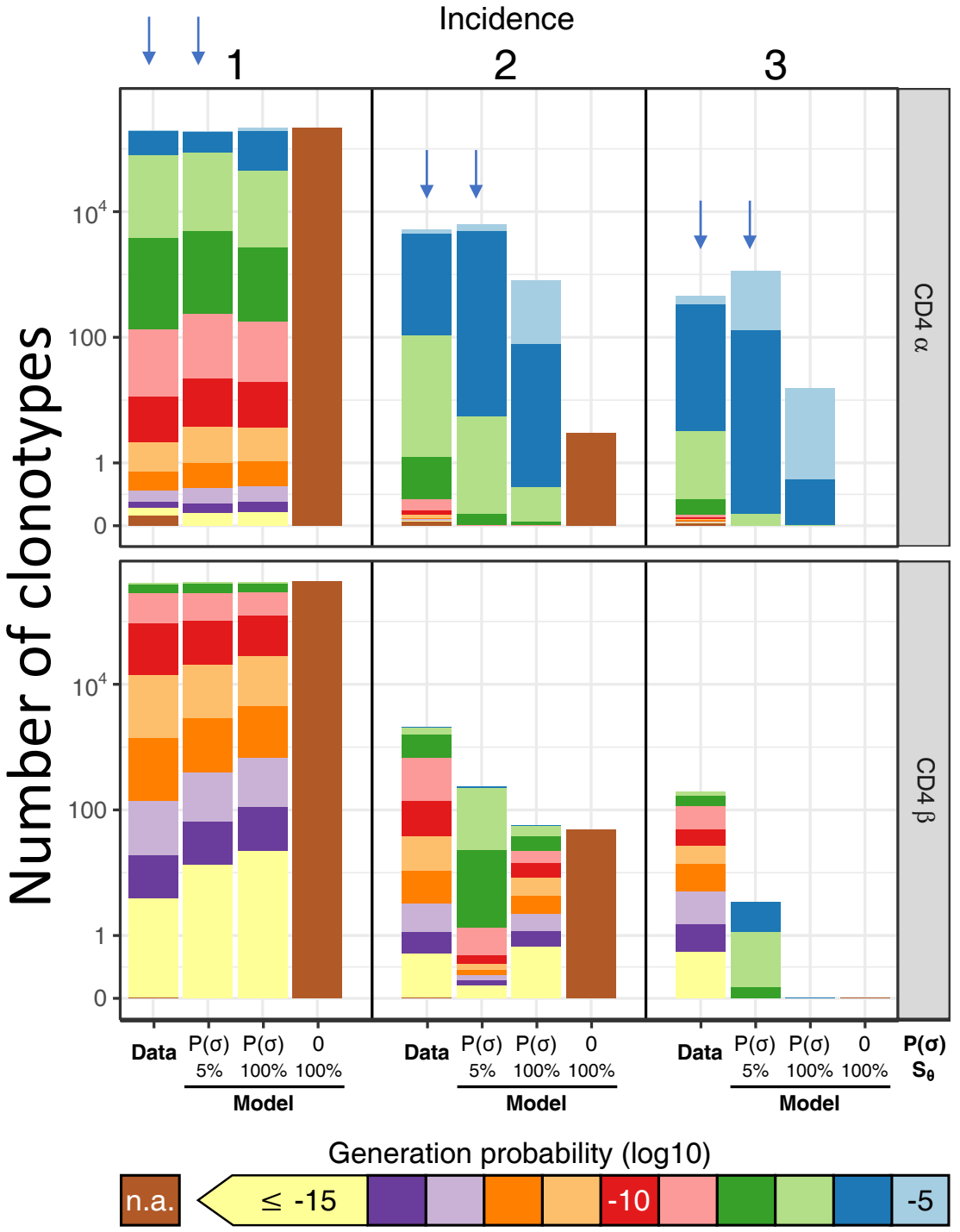
Is this true? Circumvent the mRNA problem by taking 3 sub-samples before RNA-extraction

- Use the number of sub-samples a clonotype appears in **(incidence)** to estimate its clone-size
- Single cells can only contribute mRNA to a single subsample

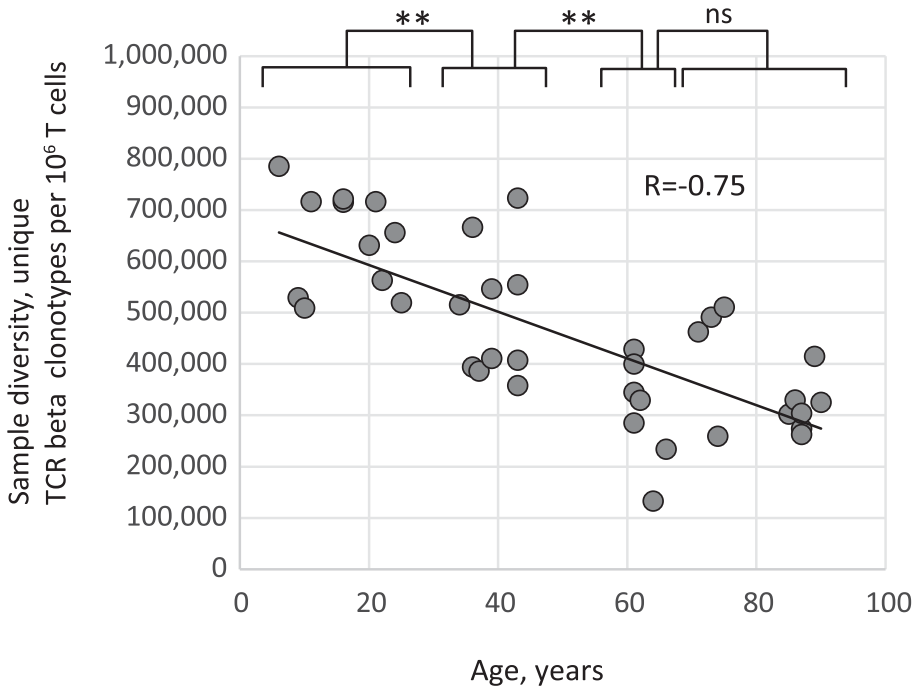


Solve the mRNA problem by making 3 sub-samples before RNA-extraction

- TCRA-clonotypes appear in multiple subsamples as a result of their high generation probability (blue colors)
- TCRB-clonotypes are larger than predicted, but for another reason (not blue)



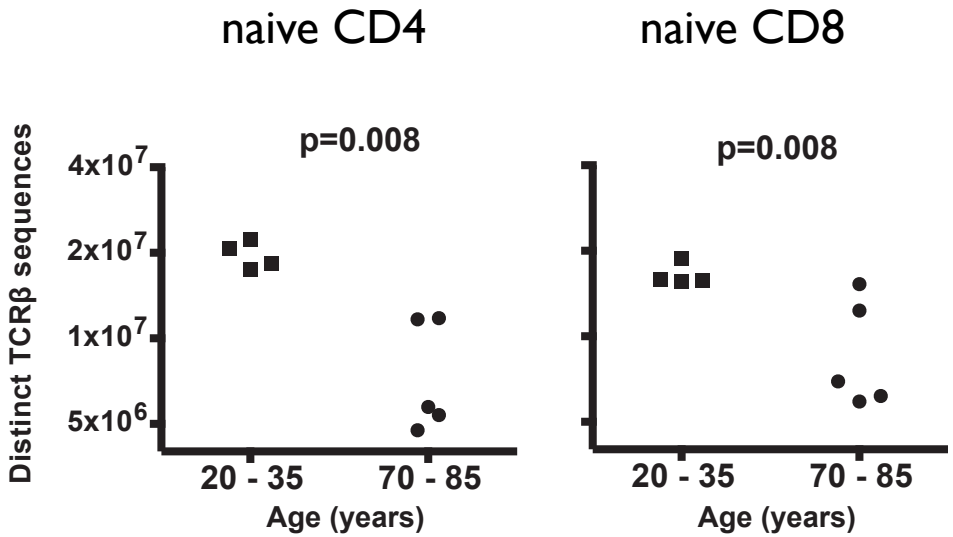
Study aging by NGS sequencing of TCR repertoire



%Naive: 54 35 17 27

Loss of diversity by loss of naive T cells

Britanova et al J. Immunology 2014



5 samples of 10^6 cells: Chao2 estimator
young adults 60-120x10^6 different TCRBs
70-85y-old adults 8-57x10^6 different TCRBs

Qi et al PNAS 2014

Richness during aging not at steady state

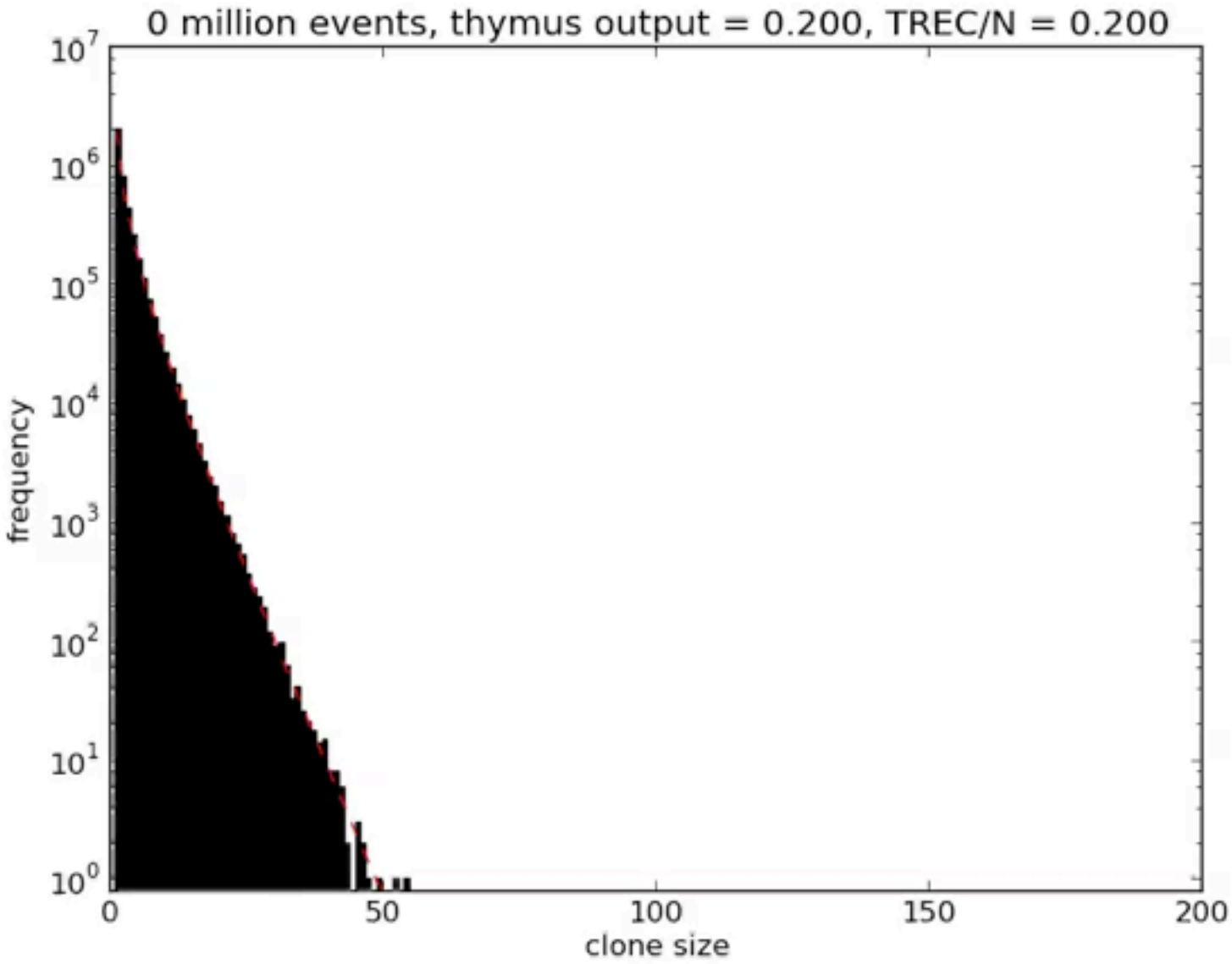
Humanized mouse simulation:

N = 10^7 cells, k=1, red: solution.

Thymic output θ initially 0.2

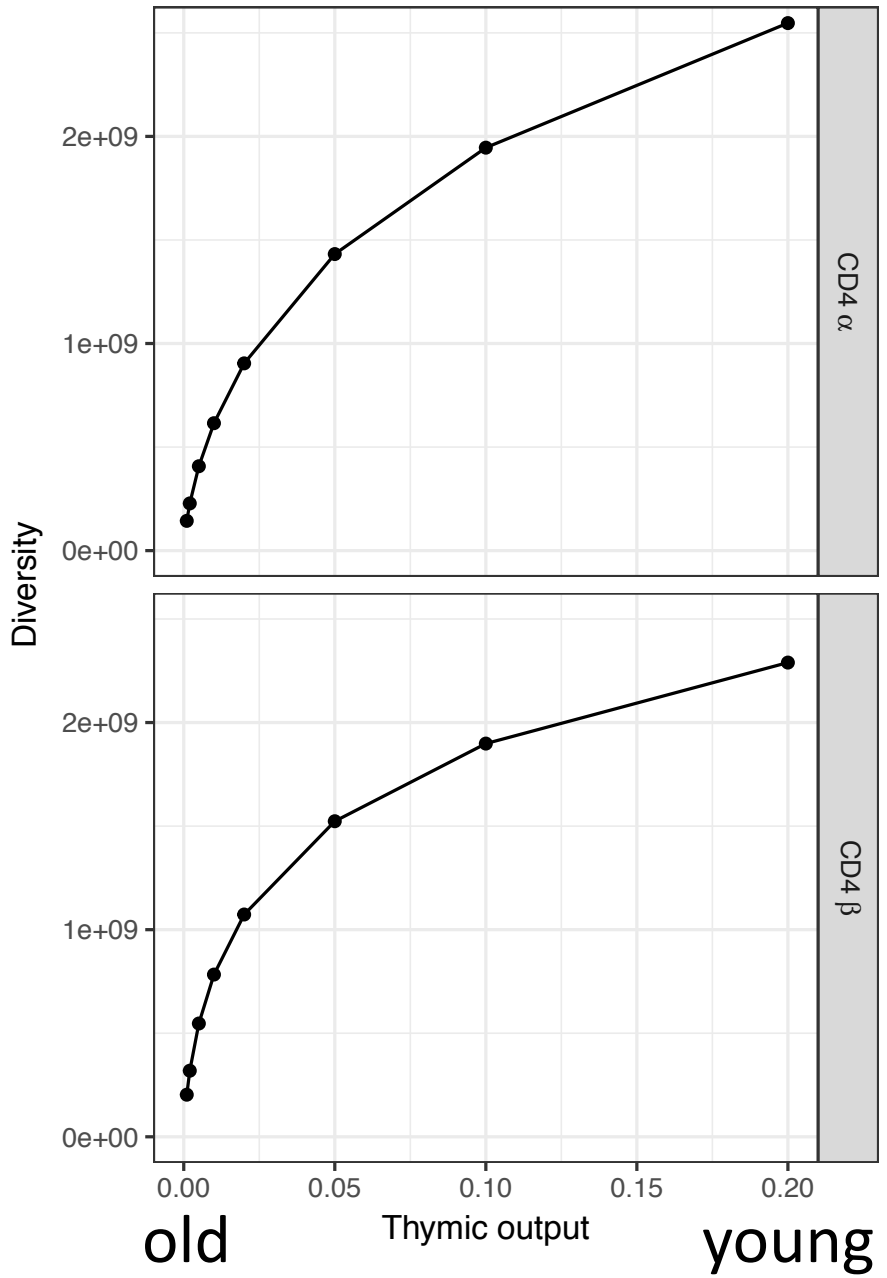
5% decrease thymic output/year

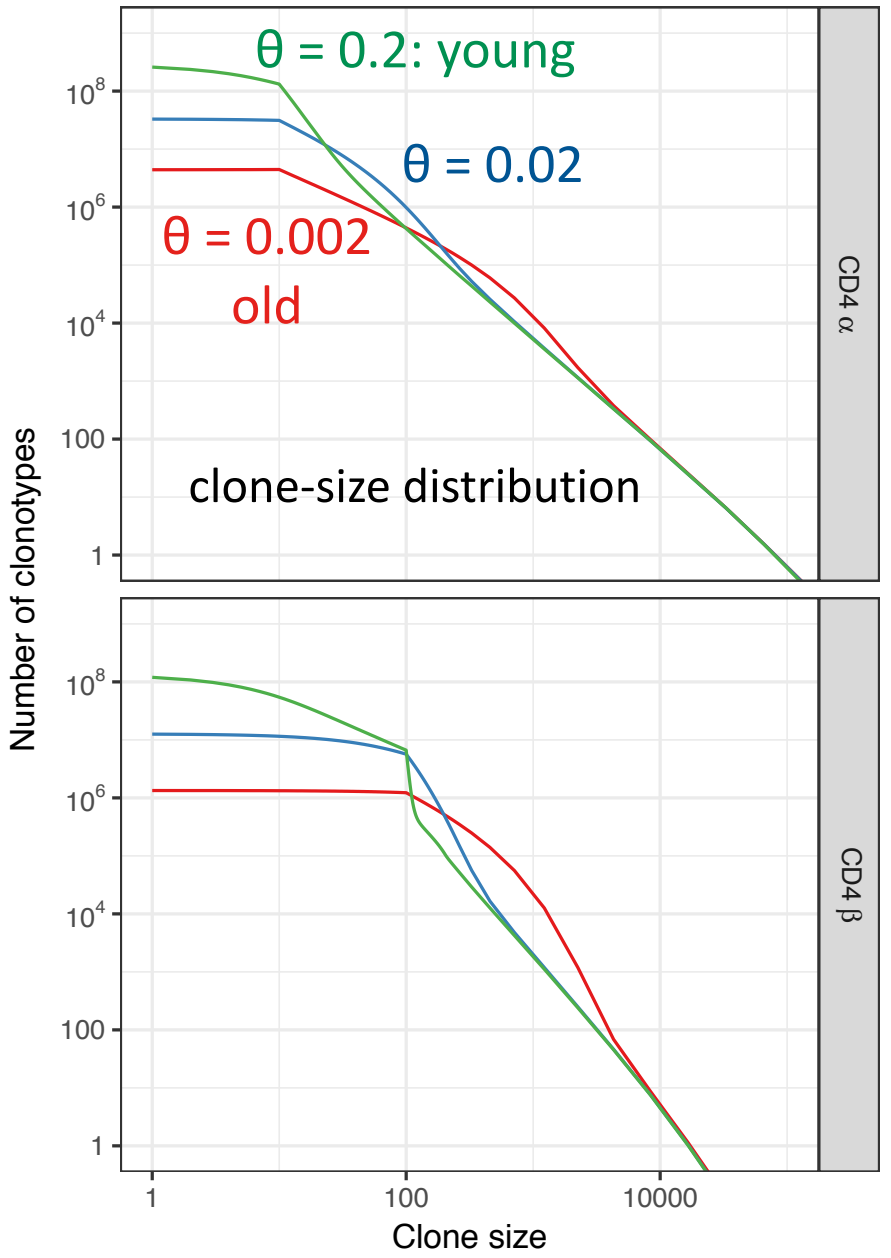
cells live 10y: a year is 10^6 events



Steady state repertoire is aging

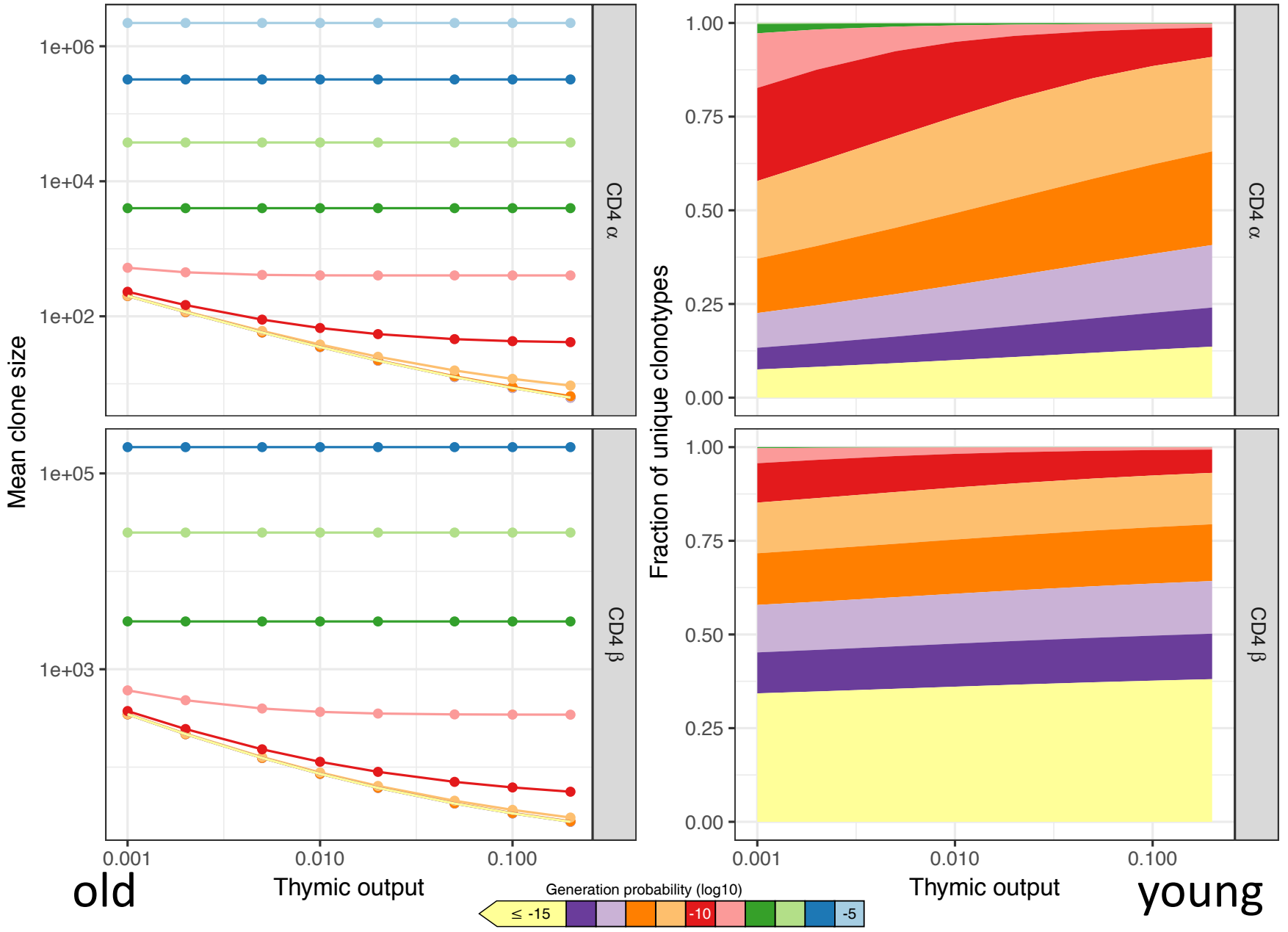
- steady state diversity declines
- largely because small clones tend to go extinct
- middle-sized clones fill in and become larger
- large clones remain the same





Steady state repertoire is aging

- Blue clones (high P(σ)) remain large
- The yellow-red clones (low P(σ)) that survive become larger
- immune responses in elderly biased toward high P(σ)?
- prediction to be tested with Igor
- back to germline?



Conclusions

Most naive clones are expected to be very small, but a few are very large.

Large clonotypes tend to have high generation probabilities

A neutral BDI model is sufficient to explain the TCRA data and most of the TCRB data

VDJ-recombination probabilities dominate over TCR-dependent fitness differences in shaping the naive T-cell pool. Tonic signaling is neutral.

Repertoire diversity erodes by aging, but very slowly.

Aging enriches for easy-to-make clones (testable prediction)

V(D)J recombination shapes the distribution of TCR chains in the naive T-cell repertoire

Peter C. de Greef[a,1], Theres Oakes[b,1], Bram Gerritsen[a,c,1], Mazlina Ismail[b], James M. Heather[b], Rutger Hermsen[a], Benjamin Chain[b,2], and Rob. J. de Boer[a,2]

Acknowledgements

Utrecht University

University Medical Center Utrecht

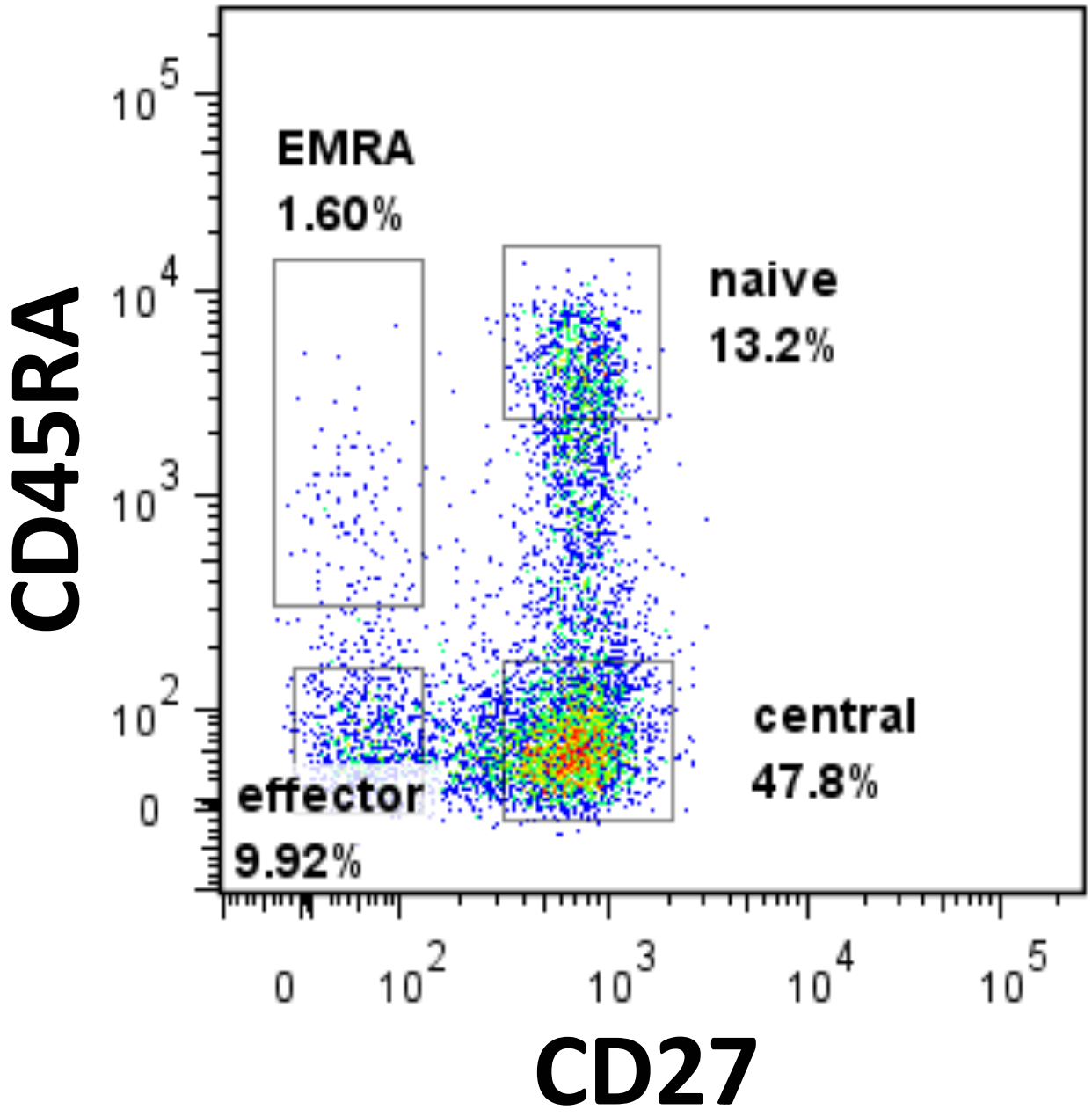
UCL

Peter de Greef
Bram Gerritsen
Rutger Hermsen
Laurens Krah

José Borghans
Julia Drylewicz
Kiki Tesselaar
Nienke Vrisekoop
Ineke den Braber
Vera van Hoeven
Liset Westera

Benny Chain
Theres Oakes
James Heather

CD4



Enrich for naive clones by removing all naive TCRs also occurring in any of the non-naive subsets.

Correct for mutations in barcodes.