Randomly made yet ordered: CD4 T cell receptor and functional repertoires

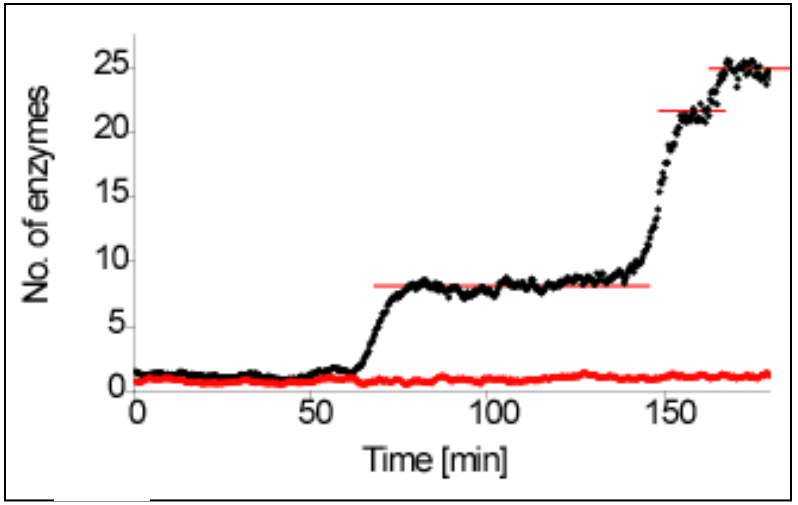
Nir Friedman
Department of Immunology
Weizmann Institute of Science

KITP Miniprogram: Quantitative Immunology - Experiments Meet Modeling. Dec 11th 2012.

Outline

Randomness in the immune system:

- Random generation of lymphocyte receptors (gene rearrangement)
- Stochasticity in gene expression



Cai, Friedman, Xie, Nature 2006

Outline

Randomness in the immune system:

- Random generation of lymphocyte receptors (gene rearrangement)
- Stochasticity in gene expression

Potential advantages for randomness:

- Recognition of a very large set of antigens (unknown, fast evolving)
- Optimal performance in an unpredictable and changing environment
- Harder to evade?

Analysis of the structure of the TCRβ naïve repertoire

Making of the T cell receptor: V-D-J recombination, a biased random process

- T cell receptors and antibodies are made through random DNA rearrangements
- Crucial for recognition of diverse, unknown antigens

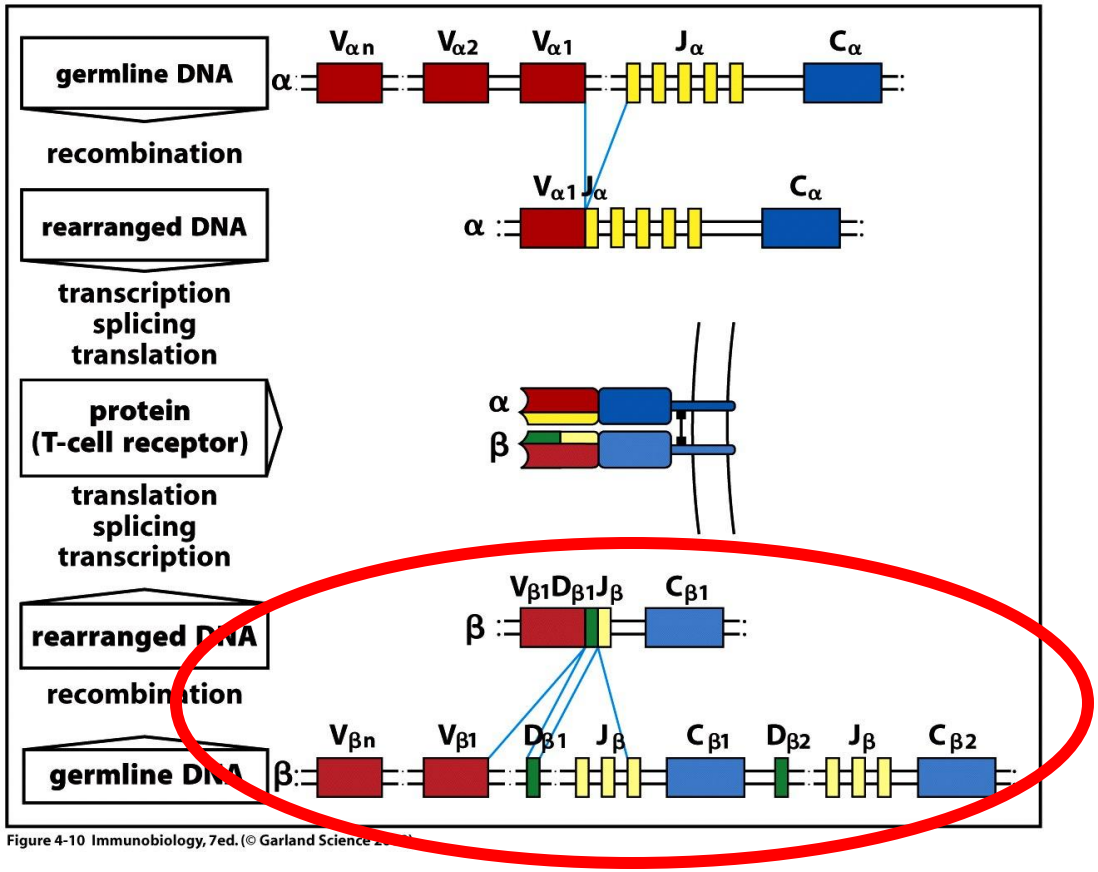


Figure 4-10 Immunobiology, 7ed. (© Garland Science 2008)

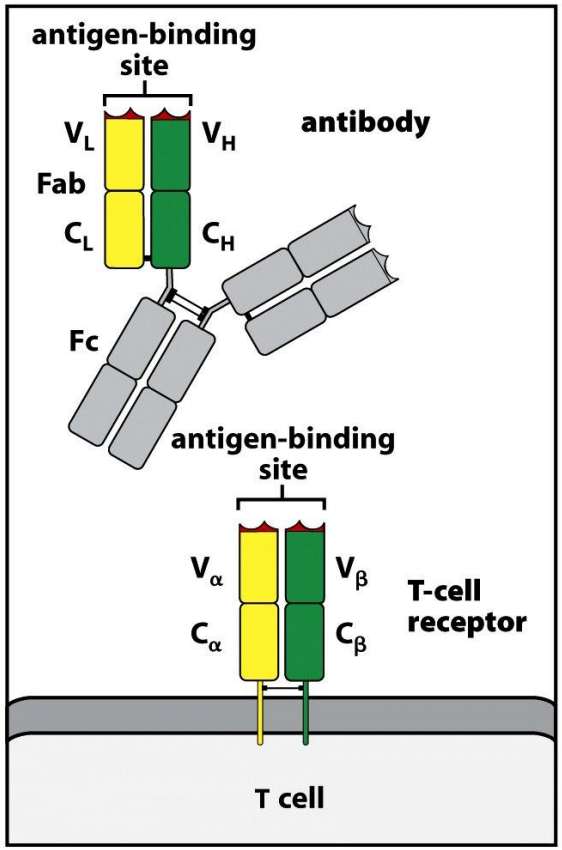
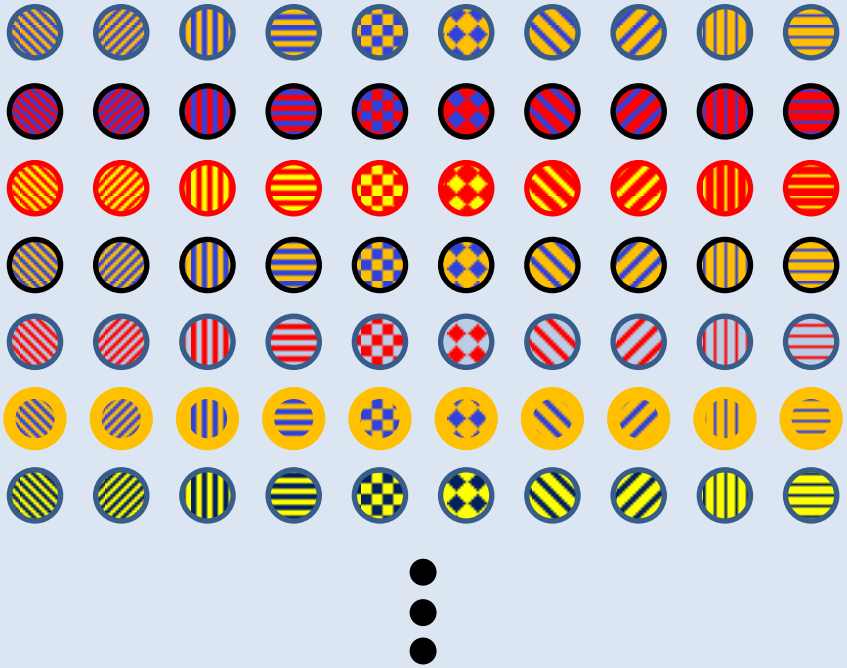


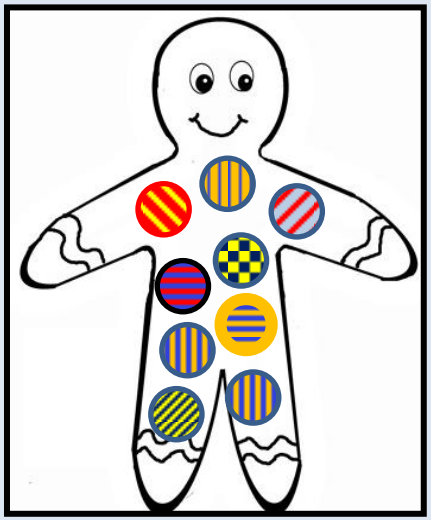
Figure 3-11 Immunobiology, 7ed. (© Garland Science 2008)

The Potential size of the TCR repertoire is huge

Estimated number of possible receptors (TCR αβ, mouse): ~ 10^{15}

Number of T cells: mouse: ~ 10^{8} ; human: ~ 10^{11} << Repertoire size





The TCR repertoire is dynamically changing throughout life

Clonal Selection Theory

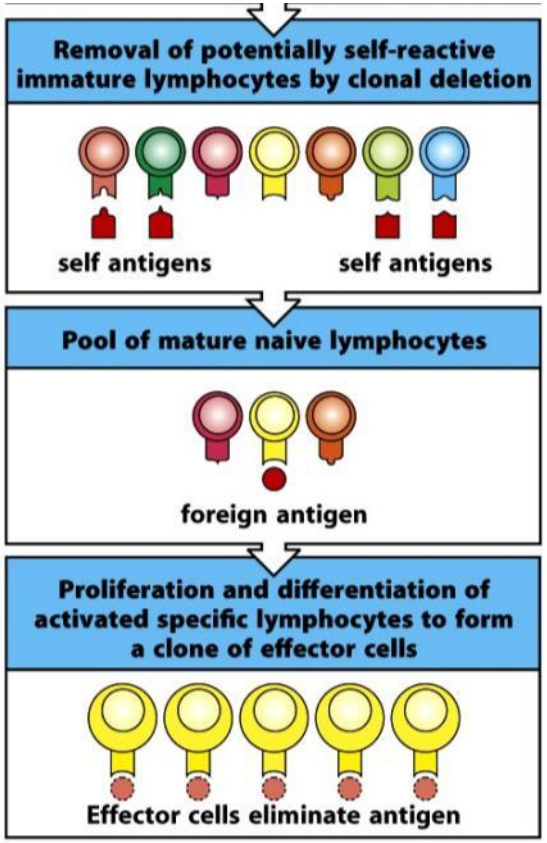


Figure 1-11 Immunobiology, 7ed. (© Garland Science 2008)

The TCR repertoire represents the state of the adaptive immune system and its history

New HTS technologies enable comprehensive repertoire characterization:

- Responses to pathogens
- Vaccinations
- Autoimmunity
- Cancer
- Aging

Mapping TCR repertoires by high-throughput sequencing

Analysis of the structure of the TCRβ naïve repertoire

Are there general organizing principles ?

Hilah Gal

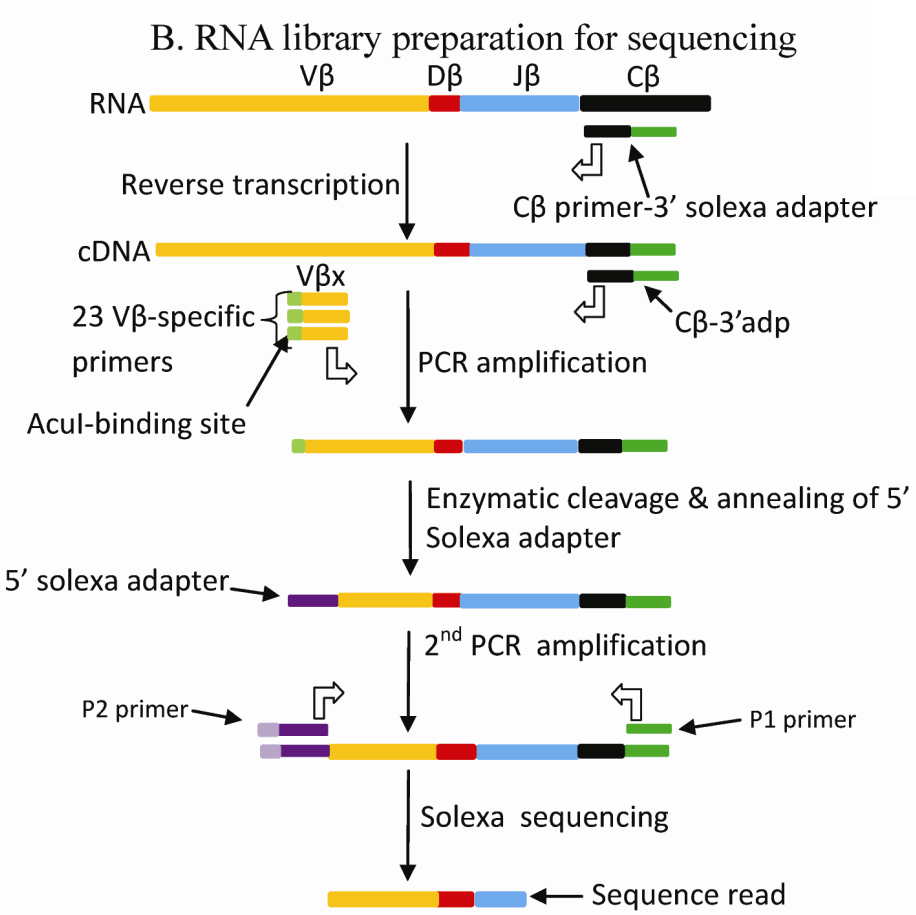
Wilfred Ndifon

Eric Shifrut

Asaf Madi

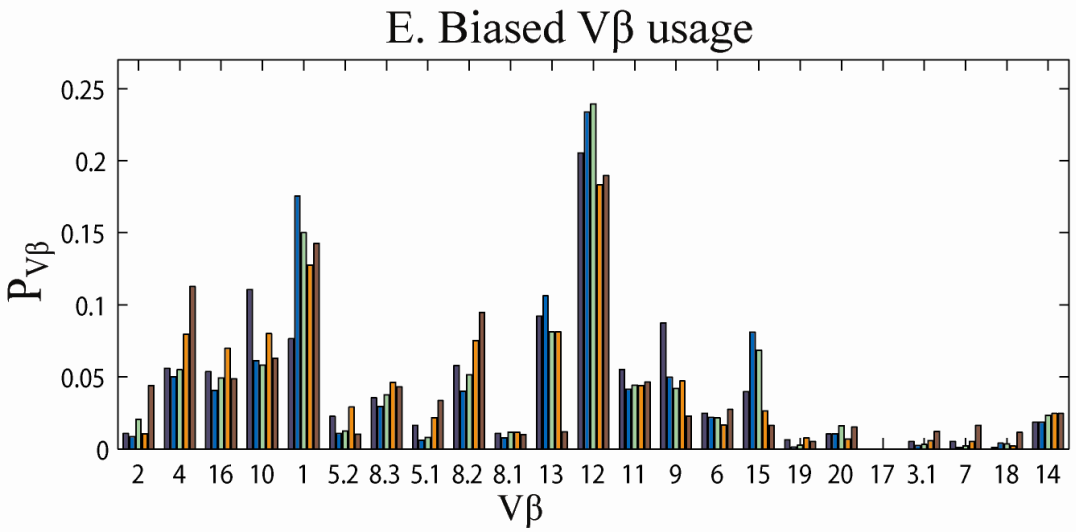
A protocol for quantitative multiplexed high throughput sequencing of the TCRβ repertoire

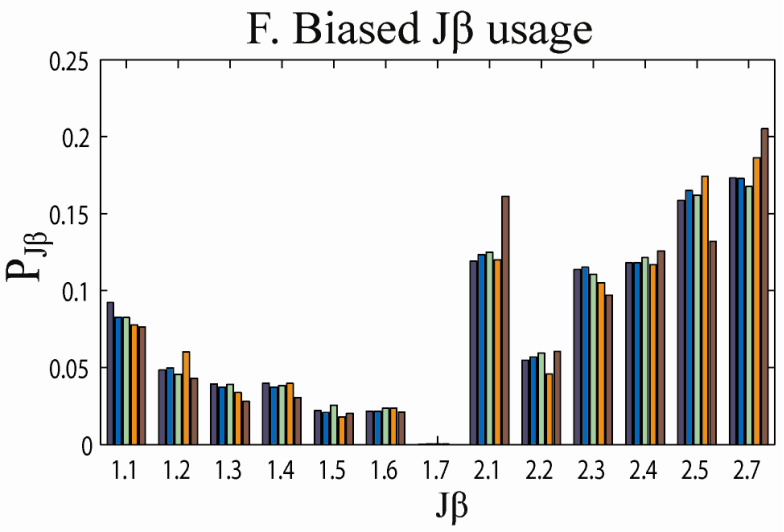
- Illumina sequencing
- Compensating PCR biases: Control plasmids library
- Challenges: resolving sequencing errors from real biological variance: Clustering, strict thresholds.



High throughput sequencing (TCR-seq) reveals common biases in the TCR repertoire

The TCRβ repertoire has a well defined structure, which is similar among individual mice

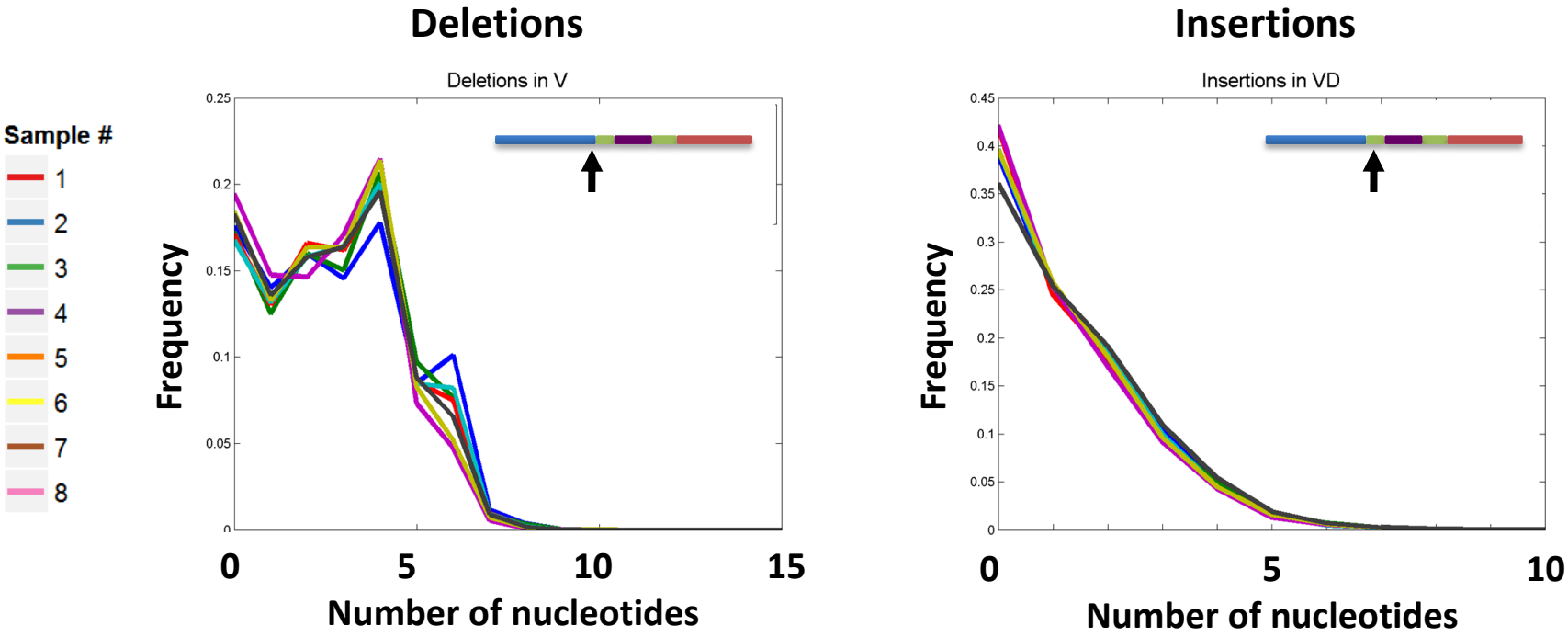




Ndifon, Gal, et. al., PNAS, 2012

High throughput sequencing (TCR-seq) reveals common biases in the TCR repertoire

The TCRβ repertoire has a well defined structure, which is similar among individual mice



Eric Shifrut, unpublished

Conclusions I:

The naive TCRβ repertoire:

a. **Is highly biased**

b. **Has very similar properties among individual mice**

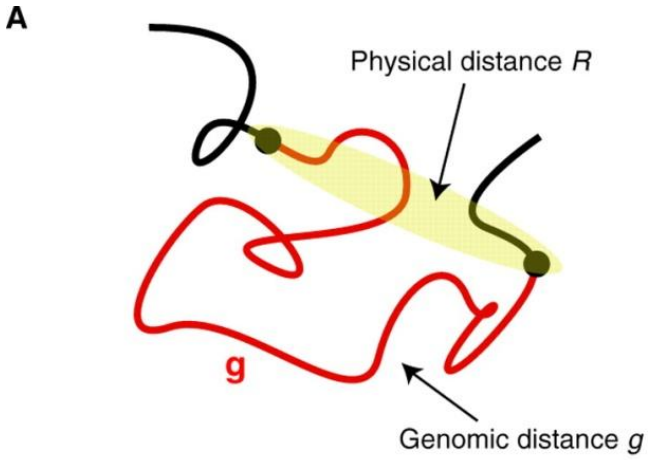
   genetically identical, including MHC, young, clean environment,…

c. **While randomly made, it has a well defined structure**

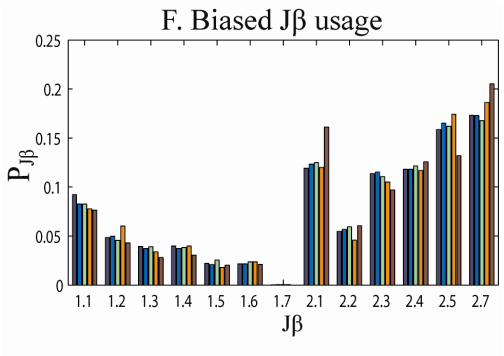
Similarity suggests common underlying principles

Mechanistic explanations for biases ?

A biophysical model can explain bias in J usage

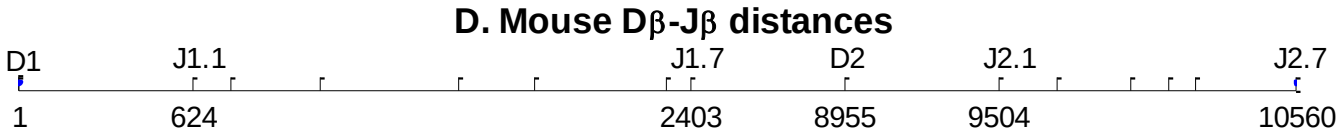


Recombination frequency depends on the physical distance between the segments, which in turn depends on their genomic distance and chromatin conformation

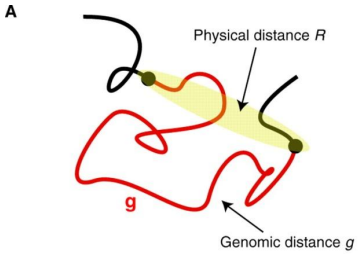


Tark-Dame M et al. J Cell Sci 2011;124:839-845

A biophysical model can explain bias in J usage



+

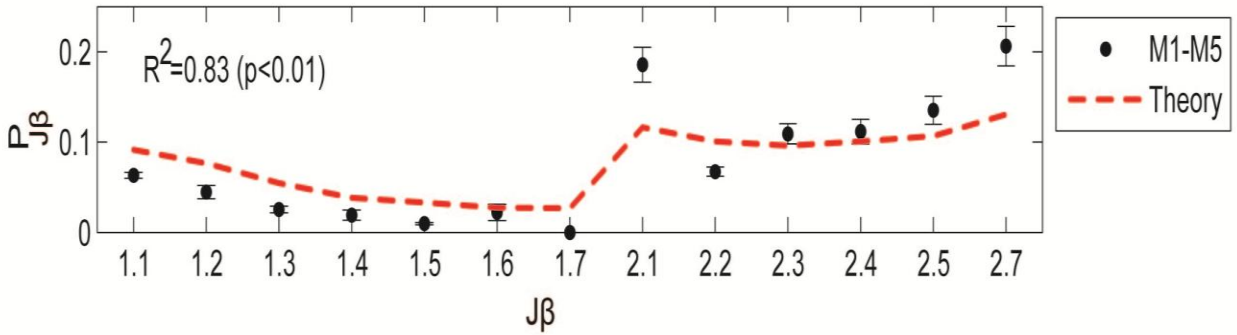


$$P(J\beta_i)=K[\alpha_1^{-3/2}\exp(-2\alpha_1^{-2})+\alpha_2^{-3/2}\exp(-2\alpha_2^{-2})]$$

Model: Dekker J. et al., Science 2002

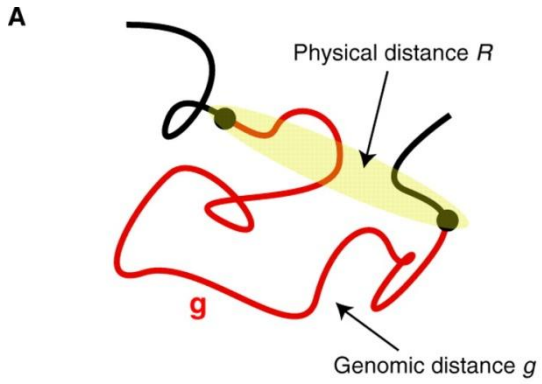
=

Mouse:



Ndifon, Gal, et. al., PNAS, 2012

A biophysical model can explain bias in J usage



Recombination frequency depends on the physical distance between the segments, which in turn depends on their genomic distance and chromatin conformation

Tark-Dame M et al. J Cell Sci 2011;124:839-845

$$P(J_i)=K[\underbrace{\alpha_1^{-3/2}\exp(-2\alpha_1^{-2})}_{\text{J-D1}}+\underbrace{\alpha_2^{-3/2}\exp(-2\alpha_2^{-2})}_{\text{J-D2}}]$$

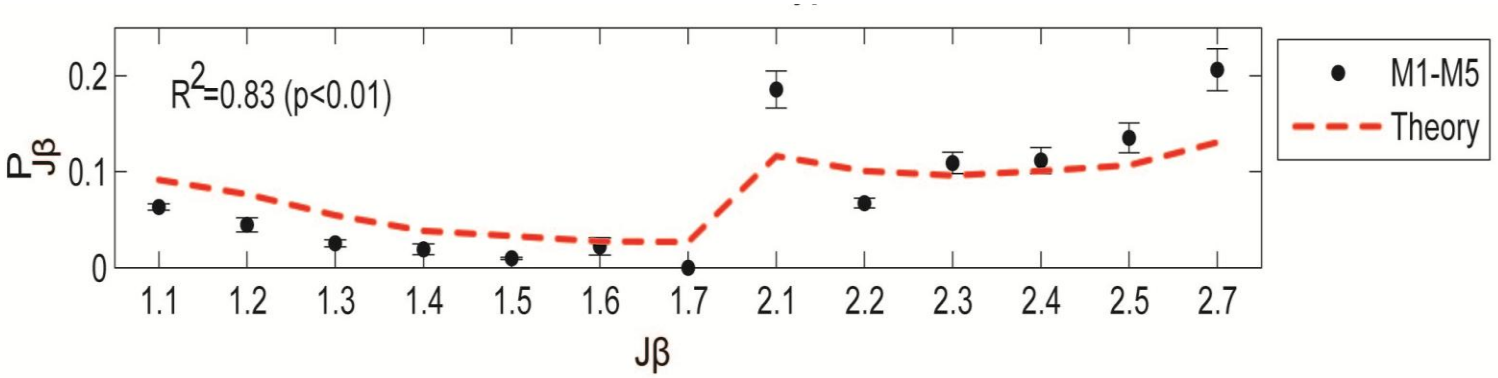
$\alpha_j=(d_j/b)(1-d_j/c)$

$d_{i,j}$ is the genomic distance between $J\beta_i$ and $D\beta_j$, K is a normalization constant.

b and c are free parameters: chromatin flexibility and curvature, respectively.

Dekker J. et al., Science 2002

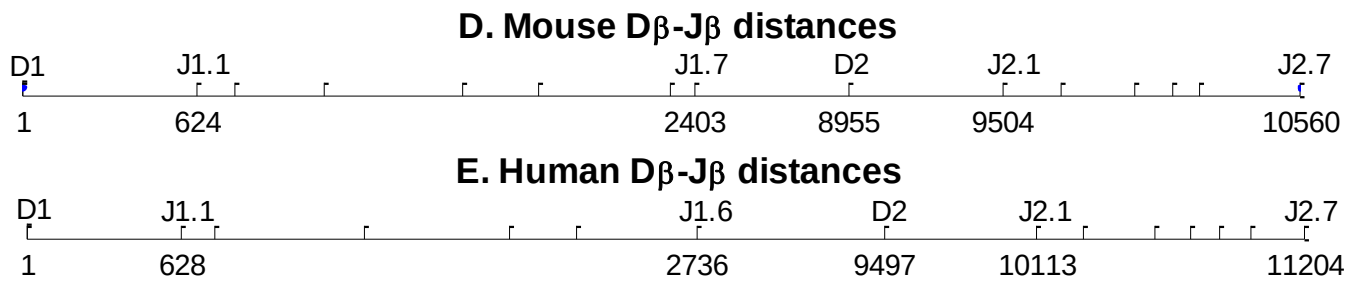
A biophysical model can explain bias in J usage



Model fit predicts a highly flexible chromatin during D-J rearrangement
Persistence length ~20nm
(in accordance with existing data on recruitment of chromatin modifiers).

The biophysical model correctly predicts biases in human TCR repertoire

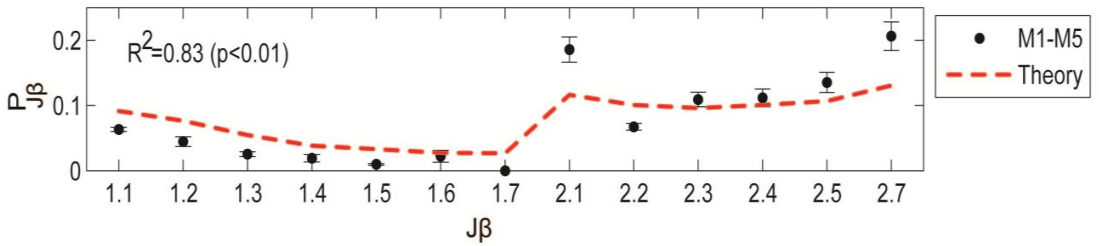
Jβ-Dβ genomic distances are different between species:



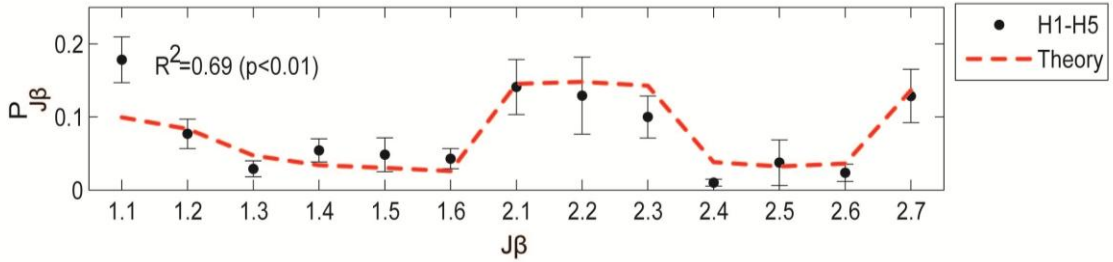
We use the model to calculate Jβ frequencies in human

using fitting parameters obtained with mouse data:

Mouse:



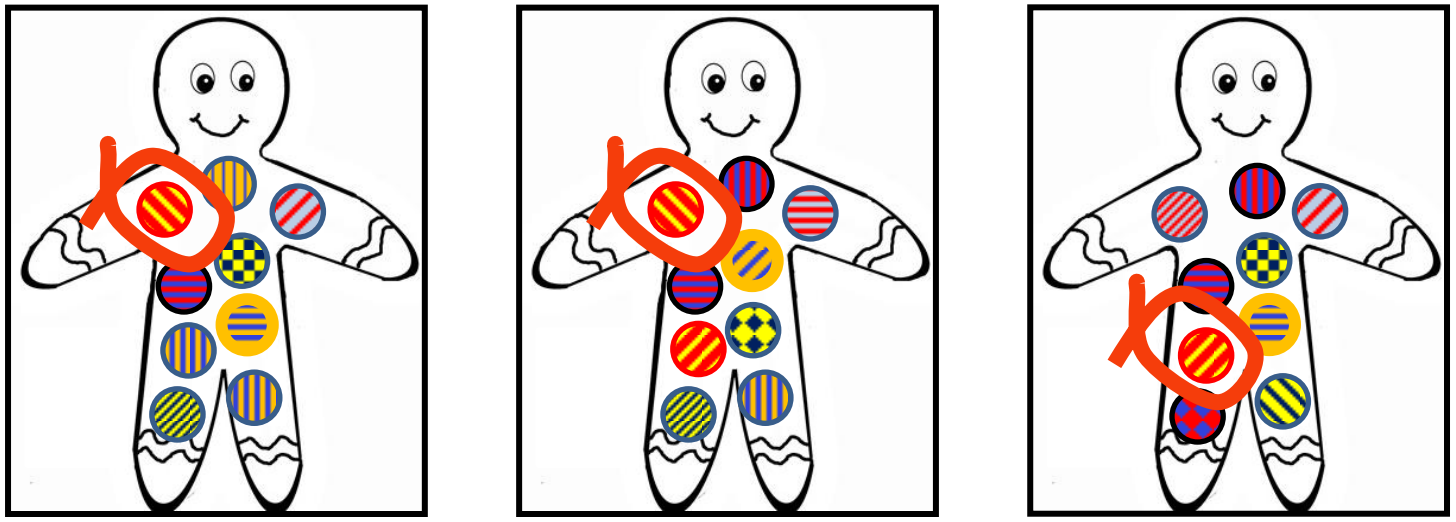
Human:



Human TCR-seq data from: Robins et al., Blood 2009, Freeman et al., Genome Res. 2009

Public TCR clones are frequently observed

Shared (“public”) clones were found between individuals that suffer from a similar pathology (viral infection, autoimmune disease, cancer, etc.), and share an HLA allele



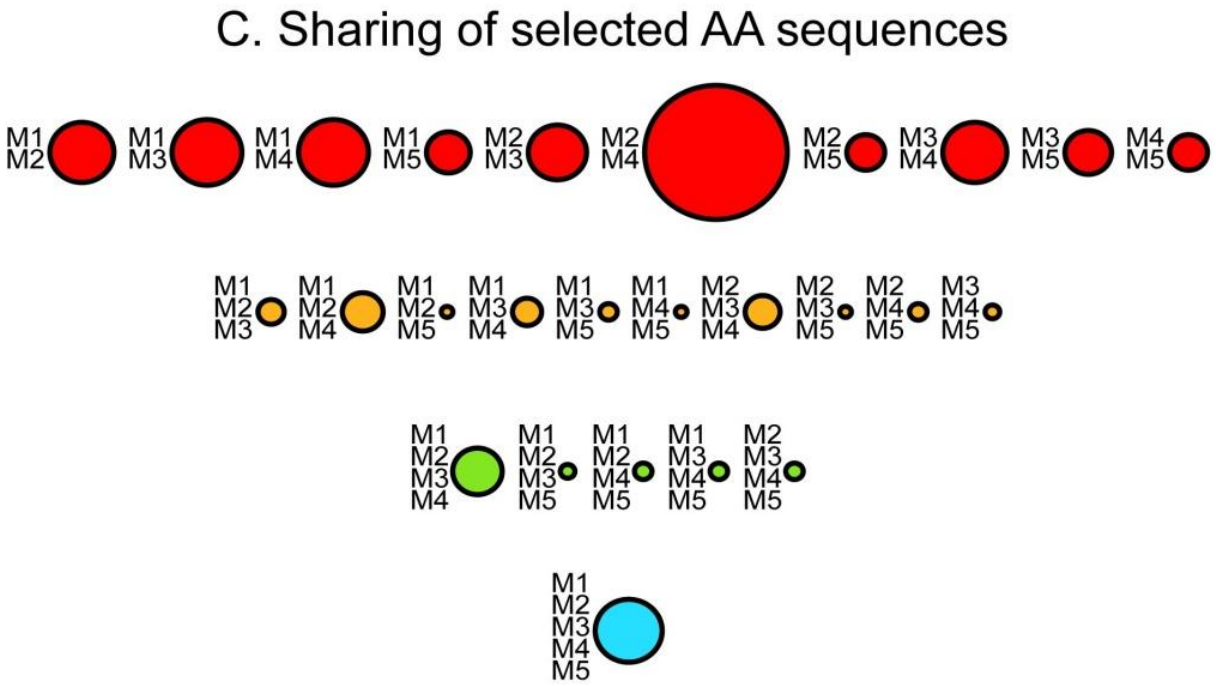
Public TCR clones are frequently observed:

Viruses, cancer, autoimmunity

Disease/pathogen	*Bias class*	*Target antigen*	*MHC restriction*	*TRBV*[a]	*TRBJ*[a]
Humans					
Influenza A	IV	MP_{58-66}	A*0201	19	2-7
Epstein–Barr virus	III	EBNA $3A_{339-347}$	B*0801	7-6	2-7
Epstein–Barr virus	III	EBNA $3B_{399-408}$	A*1101	29	2-2
Epstein–Barr virus	III and IV	$BZLF1_{54-64}$	B*3501	10-3	1-5
Epstein–Barr virus	III and IV	$EBNA1_{407-417}$	B*3501	9	2-2
Epstein–Barr virus	III and IV	$BZLF1_{52-64}$	B*3508	6-1	2-7
Epstein–Barr virus	IV	$BRLF1_{109-117}$	A*0201	19	Unknown
Epstein–Barr virus	III and IV	$BMLF1_{259-267}$	A*0201	20-1	1-2
Cytomegalovirus	III and IV	$IE1_{316-324}$	A*0201	5-1	1-3
Cytomegalovirus	III and IV	$pp65_{495-503}$	A*0201	12	1-2
Human T-cell leukemia virus type 1	III and IV	Tax_{11-19}	A*0201	6-5	2-7
Hepatitis B virus	IV	Unknown	Unknown	5-6	2-1
Hepatitis C virus	IV	Unknown	Unknown	10	2-7
Human immunodeficiency virus	III and IV	$Gag_{162-172}$	B*5701	19	1-2
Clostridium tetani	III and IV	Tetanus toxin	DRB1*0301	5-4	2-3
Herpes simplex virus	III	Virion protein 22_{49-57}	B*0702	10	2-1
Melanoma	III and IV	Melan-A_{26-35}	A*0201	27	2-1

Miles, Douek, Price, Immun. Cell Biol. 2011;
Convergent recombination: Venturi, Price, Douek, Davenport, Nat. Rev. Immun. 2008

Bias affects sequence sharing



Bias affects sequence sharing: a statistical model

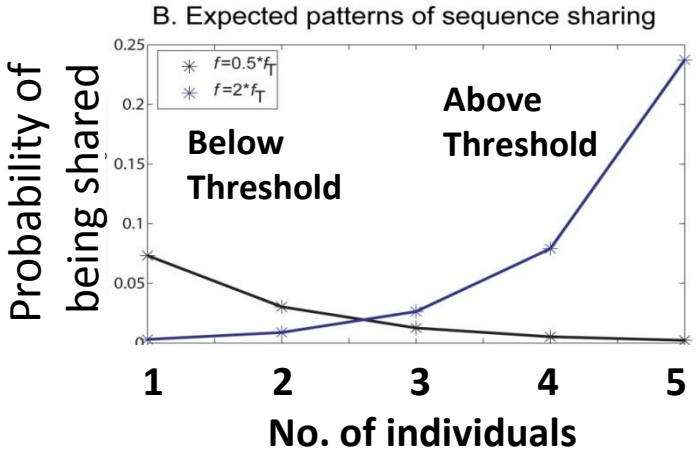
Assumptions:

- Each sequence has an a-priori probability of being made (*f*).
- Each individual has N sequences (T cell clones),

  which are randomly drawn from all possible sequences.

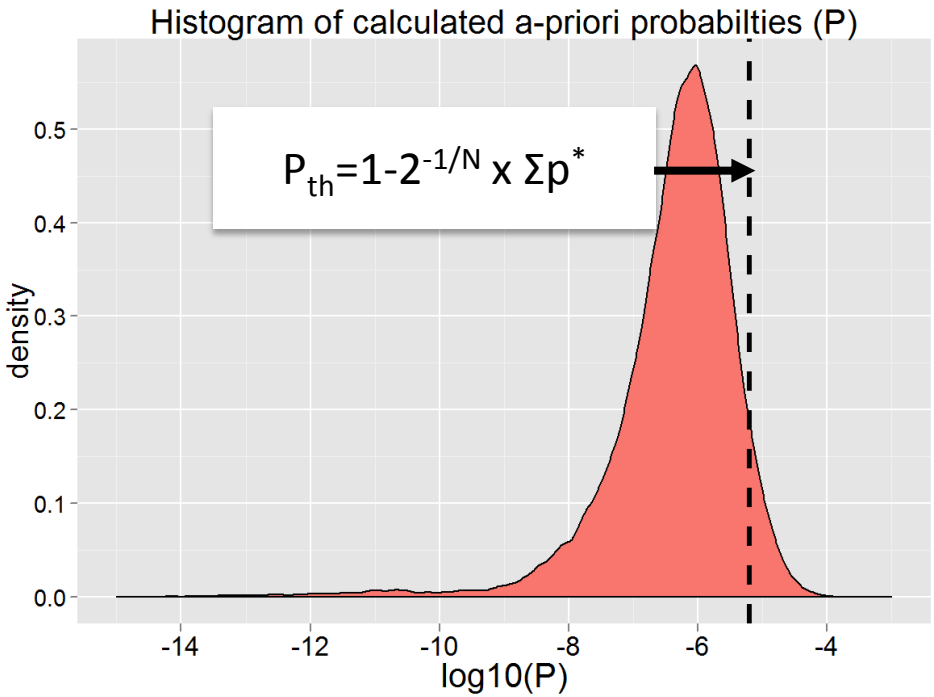
We find that there is a threshold frequency, f_T, above which clones have a higher chance of being public:

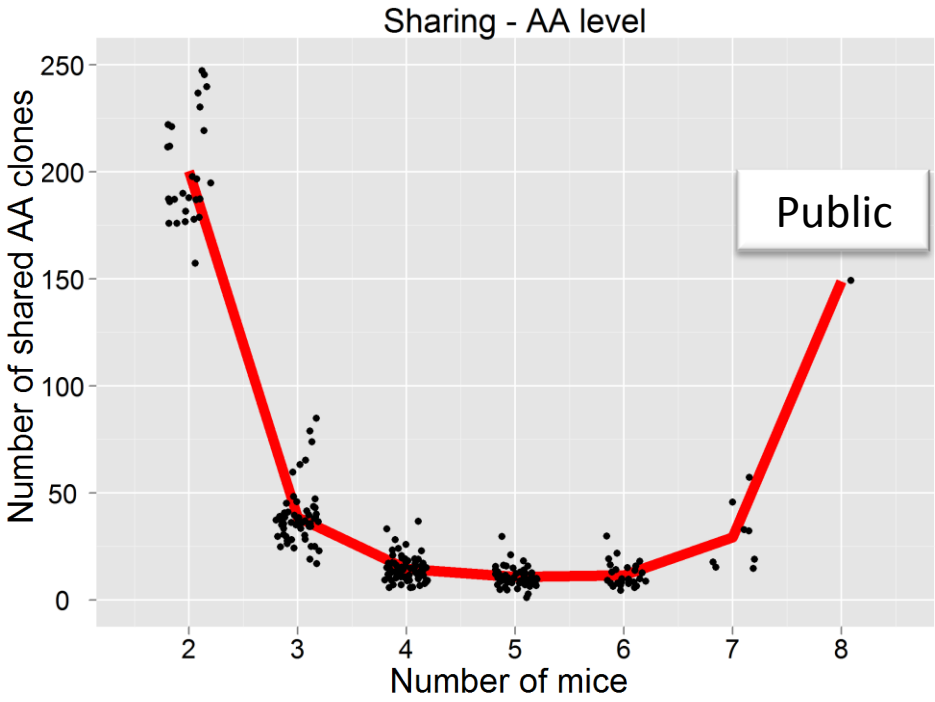
$$f_T = 1 - 2^{-1/N} \approx \ln 2/N$$



Estimating a-priori probability of clones to be generated

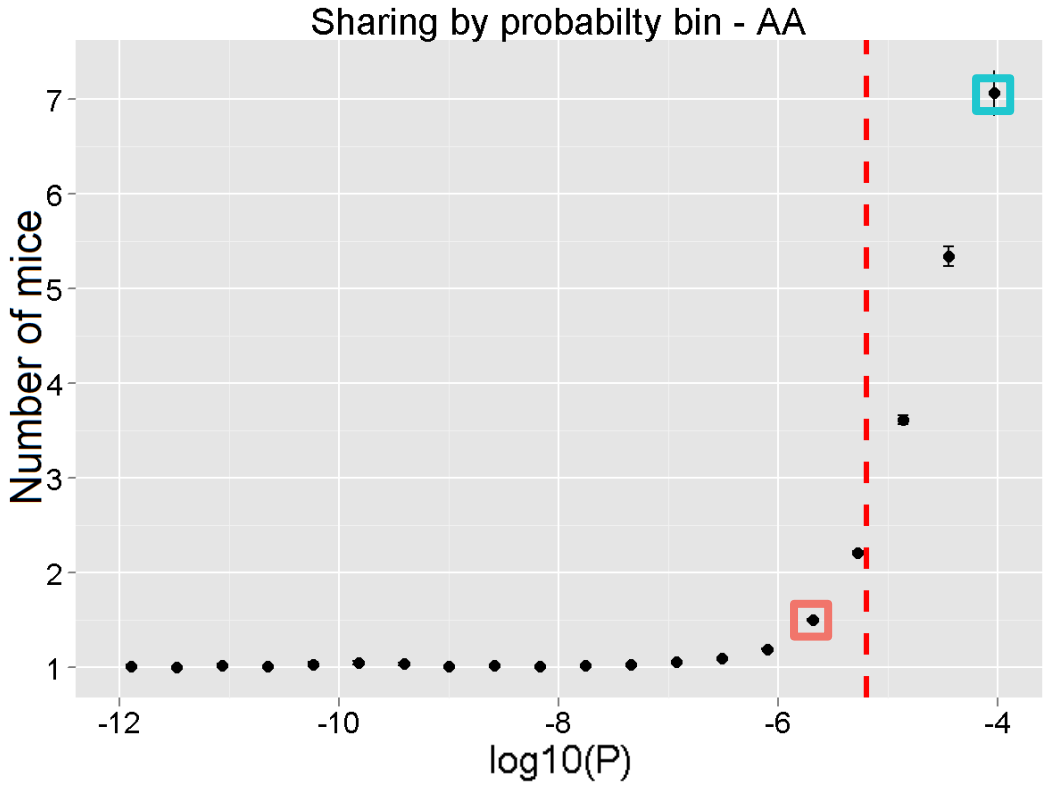
- Randomly sampled 10,000 clones from each mouse – to reduce size bias
- TCR sharing is higher than expected by uniform distribution

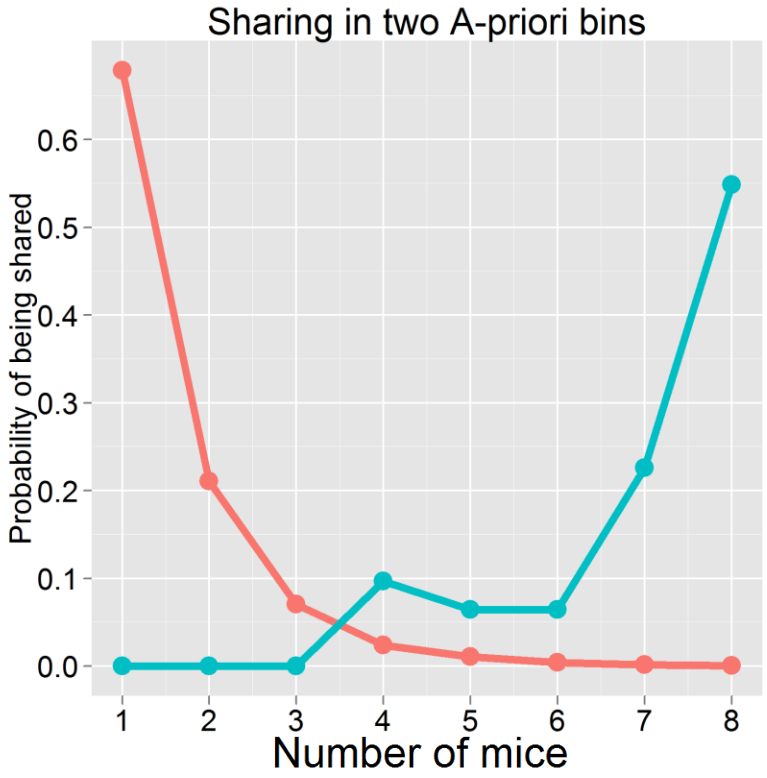




*Ndifon *et al. PNAS* (2012)

Predicting clone publicity using the a-priori probability



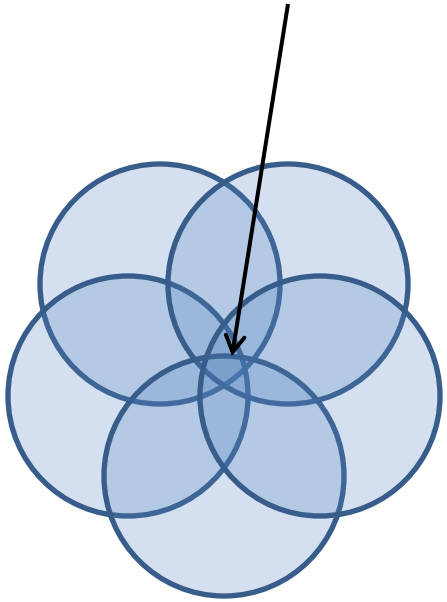


Conclusions II:

- Bias in primary repertoire allows for seemingly contradicting properties:

Huge diversity (against unknown pathogens) together with

a **predictable public "core" set of TCRs** (against frequent pathogens? Self?)



Immunological homunculus? (I. Cohen)

Thanks:

Shlomit Reich-Zeliger

Hilah Gal

Wilfred Ndifon

Asaf Madi

Yaron Antebi

Ira Zaretsky

Tamar Arieli

Eric Shifrut

Jacob Rimer

Michal Polonsky

Inbal Eizenberg

Keren Levinstein

Michal Mark

Collaborations:

Ruth Arnon, Weizmann

Rina Aharoni

Michal Schwartz, Weizmann

Kuti Baruch

Irun Cohen

Benjamin Chain, UCL

Dana Pe'er, Columbia

Uri Alon, Weizmann

Avi Mayo

Yuval Hart

Funding: ISF, HFSPO

Thank You