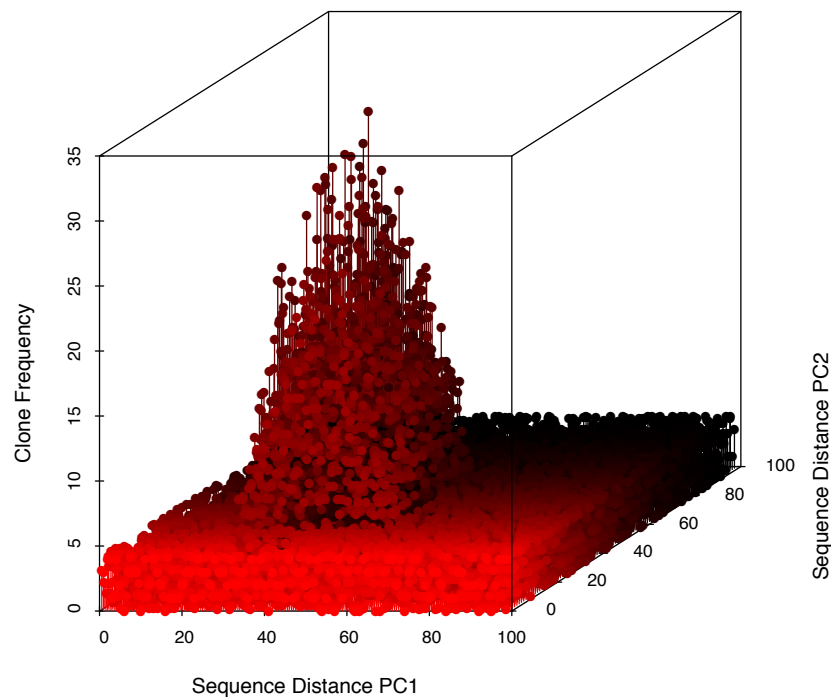


Adaptive repertoires

Jacob Glanville

Clone Frequency Landscape



2014 immuno 206a

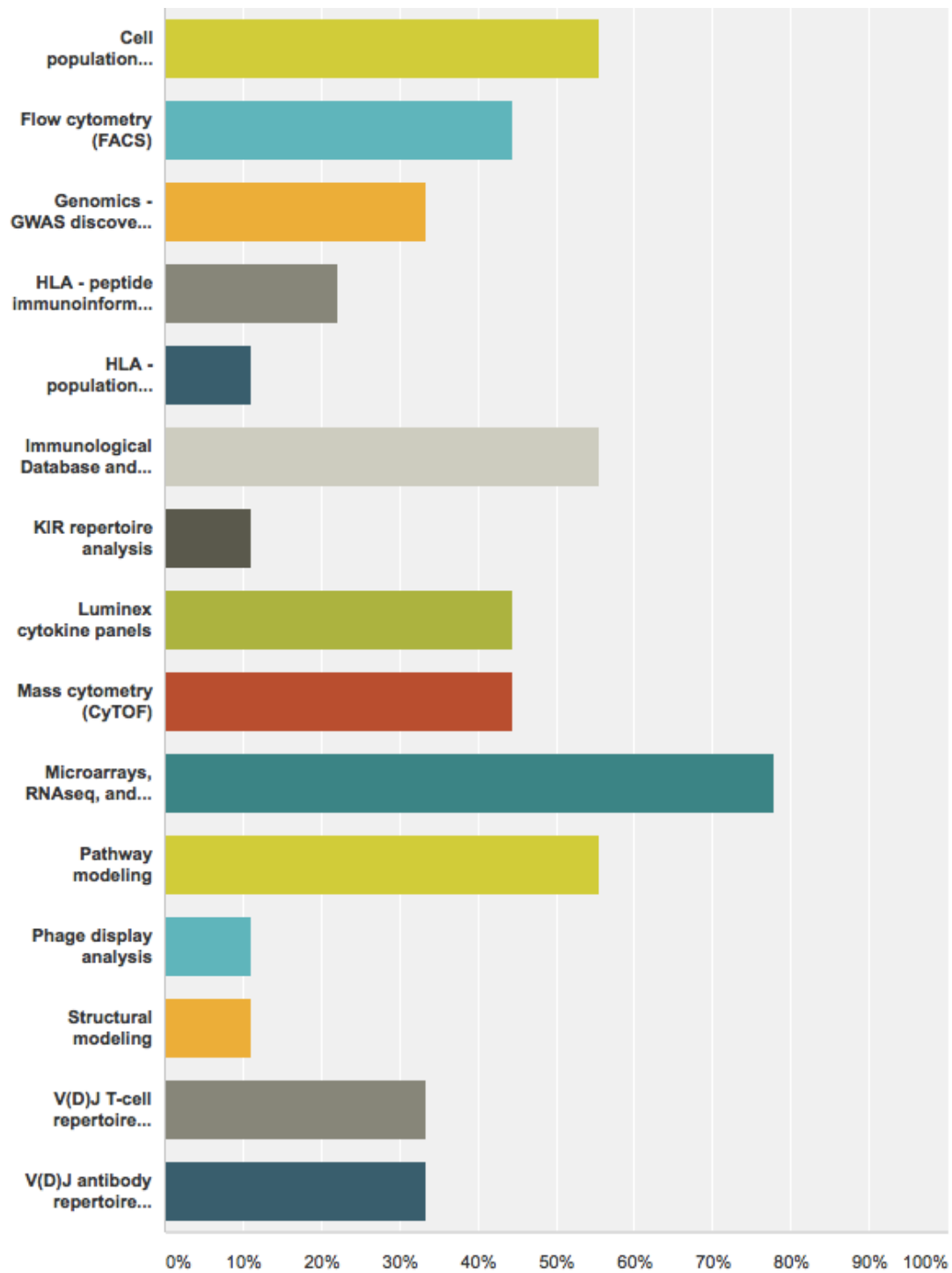
04/02/2014

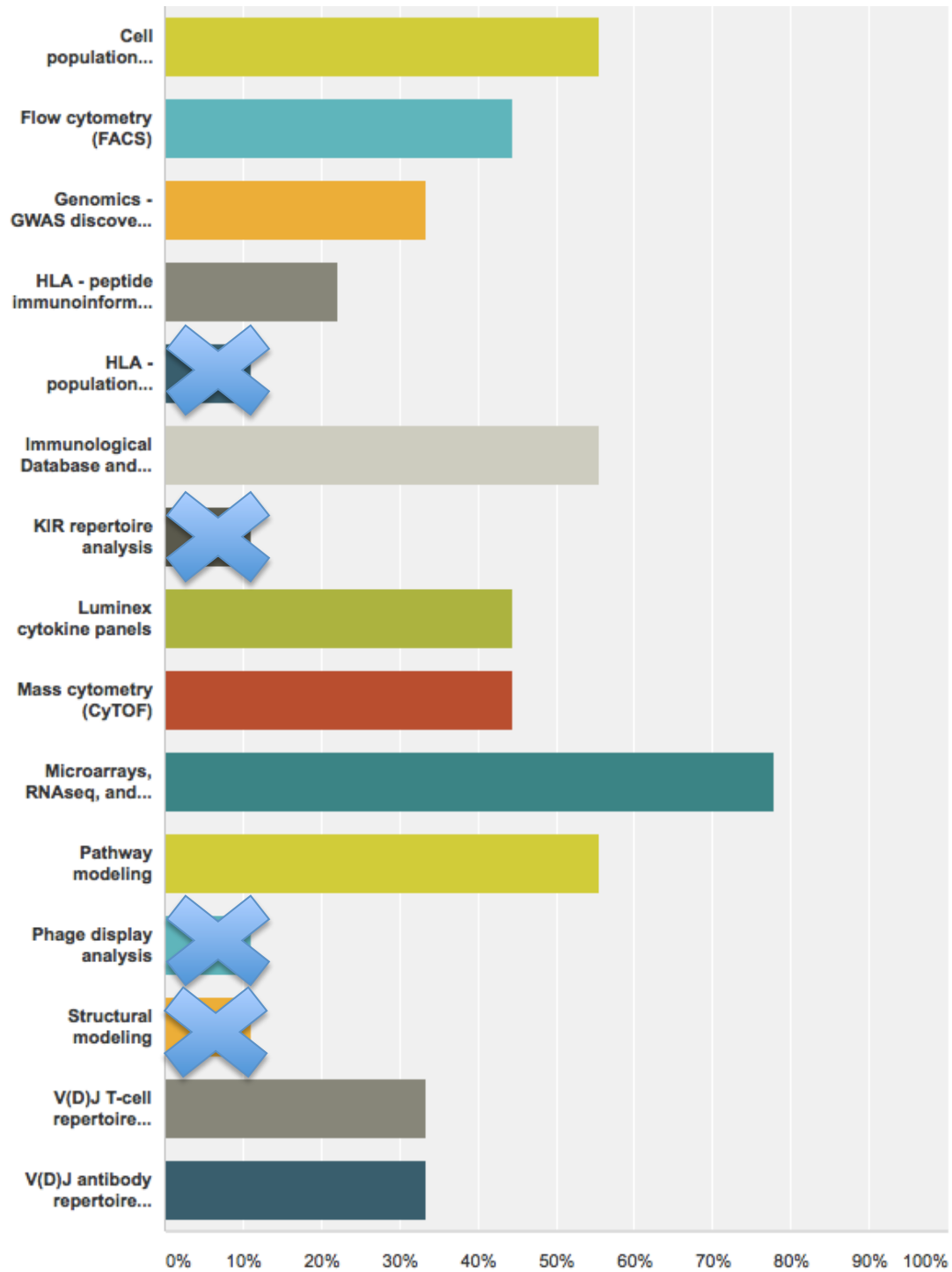
Wednesday format:

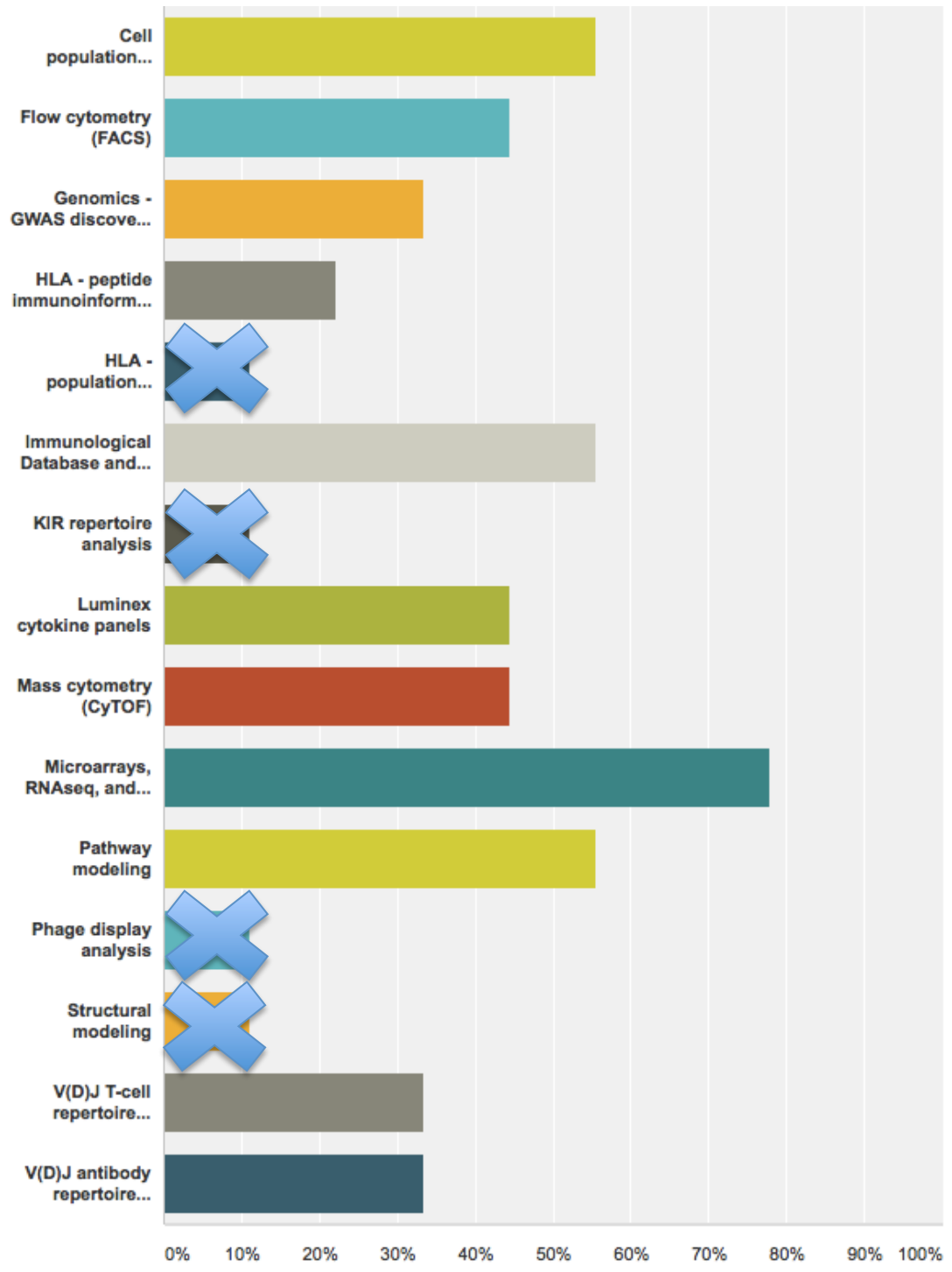
Applied Methods Lecture (30–60 minutes)

Code review

Round-table discussion







UC Berkeley

Molecular and Cell Biology BA program

Genetic Genomics Development

Thomson Population Genetics Lab

Sjolander Berkeley Phylogenomics Lab

Pfizer Inc.

Principal Scientist

Algorithm Development

Antibody engineering

Repertoire sequencing

Antibody library design

Precision medicine biomarker discovery

Distributed Bio

Chief Science Officer

Cloud-based repertoire analysis platforms

High throughput sequencing algorithms

Bioengineering applications

Stanford

Immunology PhD program

Computational and Systems Immunology

Mark Davis Lab

Scott Boyd Lab



Title / Author

Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire

J Glanville, W Zhai, J Berka, D Telman, G Huerta, GR Mehta, I Ni, L Mei, ...
Proceedings of the National Academy of Sciences 106 (48), 20216-20221

Berkeley Phylogenomics Group web servers: resources for structural phylogenomic analysis

JG Glanville, D Kirshner, N Krishnamurthy, K Sjölander
Nucleic acids research 35 (suppl 2), W27-W32

Synthetic antibodies designed on natural sequence landscapes

W Zhai, J Glanville, M Fuhrmann, L Mei, I Ni, PD Sundar, T Van Blarcom, ...
Journal of molecular biology 412 (1), 55-71



Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation

J Glanville, TC Kuo, HC von Büdingen, L Guey, J Berka, PD Sundar, ...
Proceedings of the National Academy of Sciences 108 (50), 20066-20071



B cell exchange across the blood-brain barrier in multiple sclerosis

HC von Büdingen, TC Kuo, M Sirota, CJ van Belle, L Apeltsin, J Glanville, ...
The Journal of clinical investigation 122 (12), 4533

Comprehensive interrogation of a minimalist synthetic CDR-H3 library and its ability to generate antibodies with therapeutic potential

CM Mahon, MA Lambert, J Glanville, JM Wade, BJ Fennell, MR Krebs, ...
Journal of molecular biology 425 (10), 1712-1730

The restricted DH gene reading frame usage in the expressed human antibody repertoire is selected based upon its amino acid content

J Benichou, J Glanville, ETL Prak, R Azran, TC Kuo, J Pons, C Desmarais, ...
The Journal of Immunology 190 (11), 5567-5577



Dietary gluten triggers concomitant activation of CD4+ and CD8+ $\alpha\beta$ T cells and $\gamma\delta$ T cells in celiac disease

A Han, EW Newell, J Glanville, N Fernandez-Becker, C Khosla, Y Chien, ...
Proceedings of the National Academy of Sciences 110 (32), 13073-13078

The antibody mining toolbox: An open source tool for the rapid analysis of antibody repertoires

S D'Angelo, J Glanville, F Ferrara, L Naranjo, CD Gleasner, X Shen, ...
mAbs 6 (1), 160-172

Multi step selection in Ig H chains is initially focused on CDR3 and then on other CDR regions

G Liberman, J Benichou, L Tsaban, J Glanville, Y Louzoun
Frontiers in immunology 4

Estimate of Within Population Incremental Selection Through Branch Imbalance

The antibody mining toolbox

S D'angelo, J Glanville, F Ferrara, L Naranjo, CD Gleasner, X Shen, ...

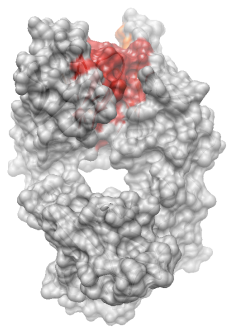
Staphylococcus aureus specific antibodies and uses thereof

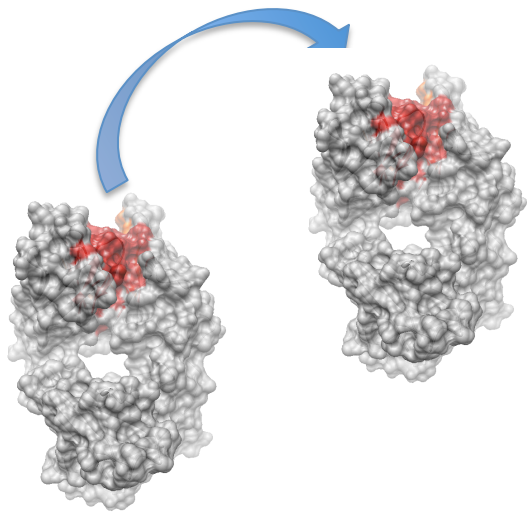
DL Foletti, JFC Riggers, JEG Glanville, LMB Shaughnessy, DL Shelton, ...
US Patent App. 13/719,214

ANTI-NOTCH-1 ANTIBODIES

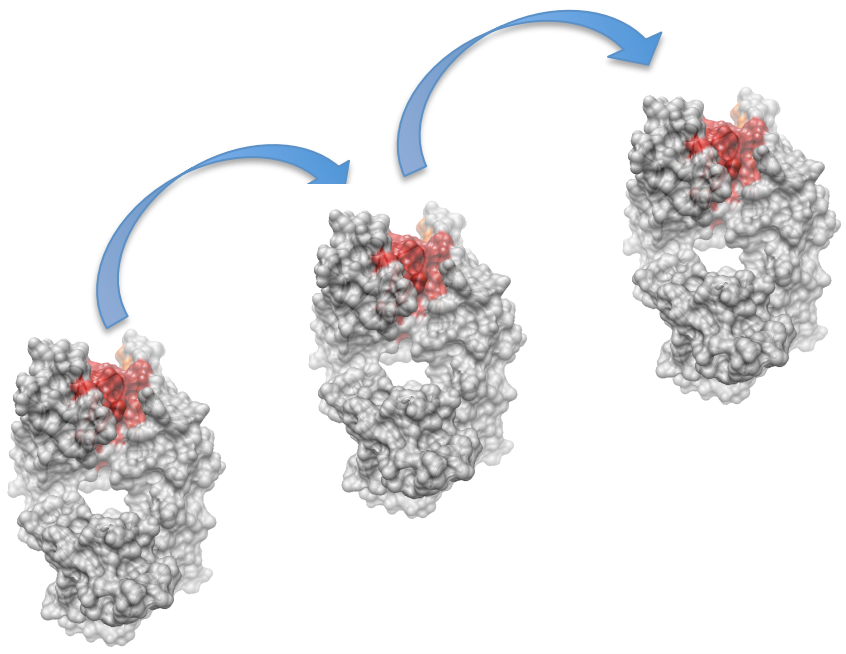
A RAJPAL, DM STONE, JEG GLANVILLE, W ZHAI
WO Patent 2,012,080,891

Adaptive repertoires



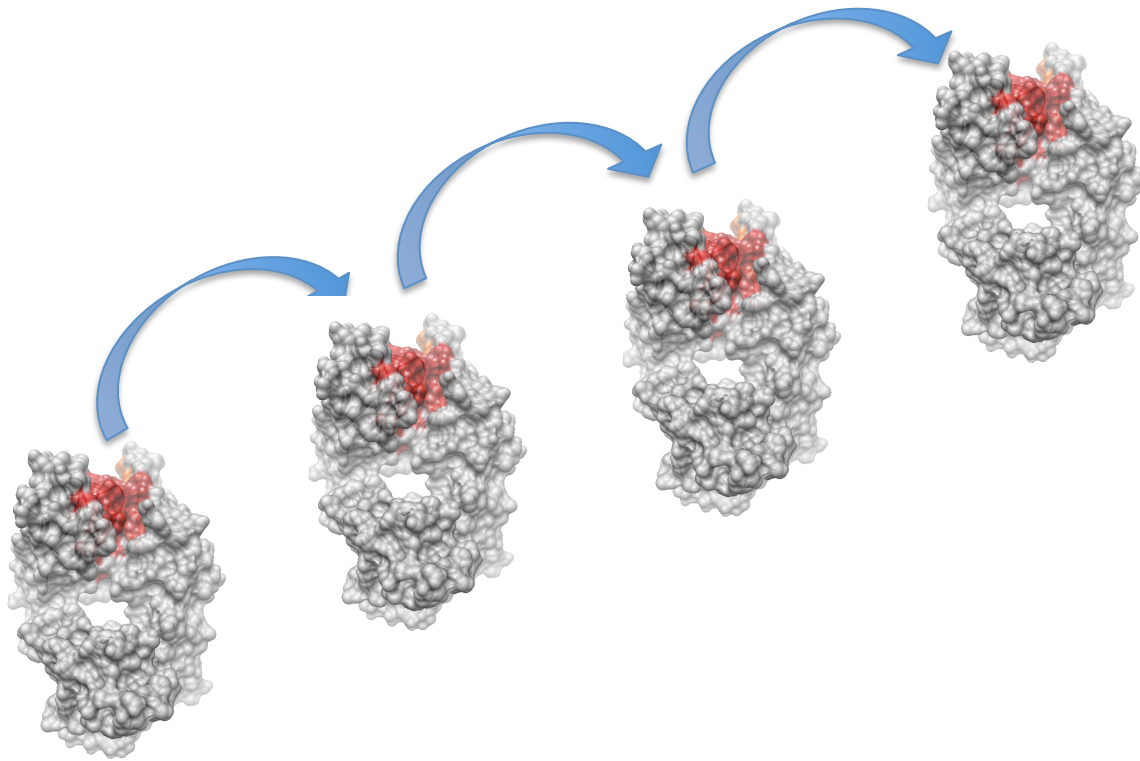


Affinity maturation



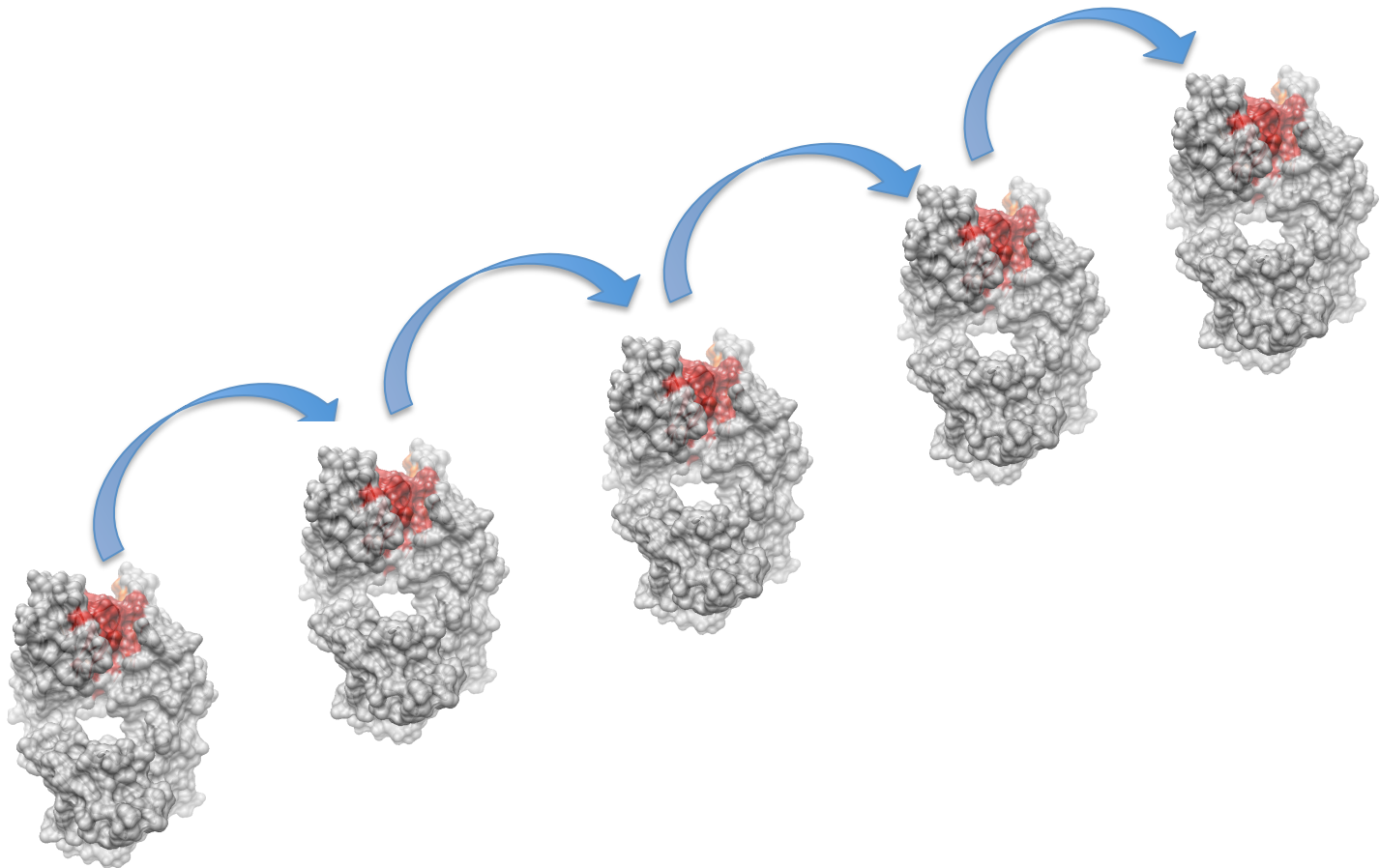
Affinity maturation

Biochemical liability elimination



Affinity maturation

Biochemical liability elimination

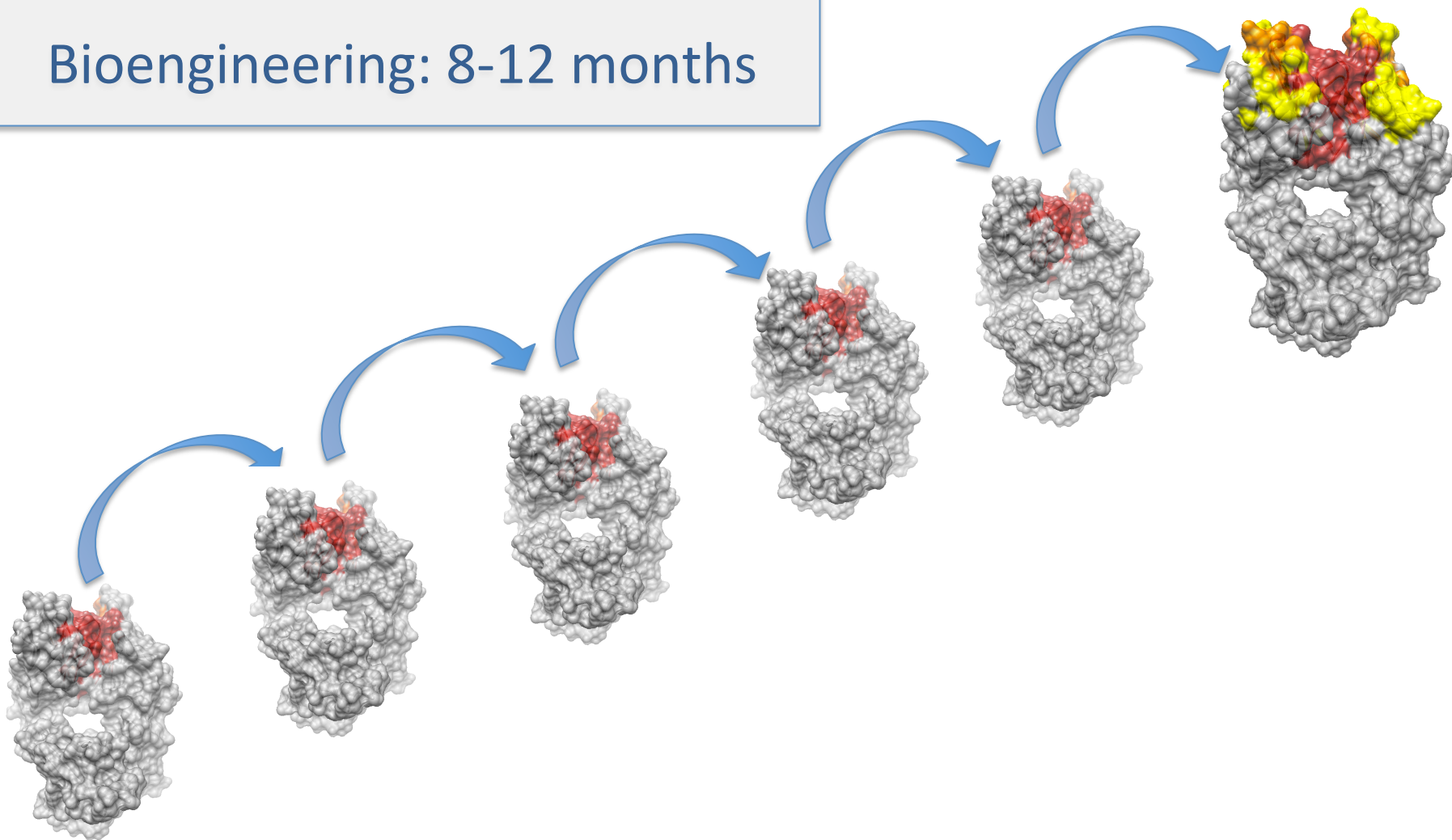


Affinity maturation

Biochemical liability elimination

Off-target binding removal

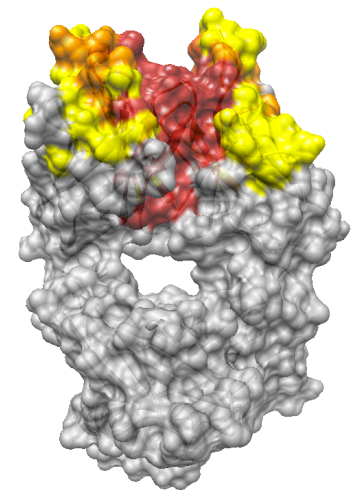
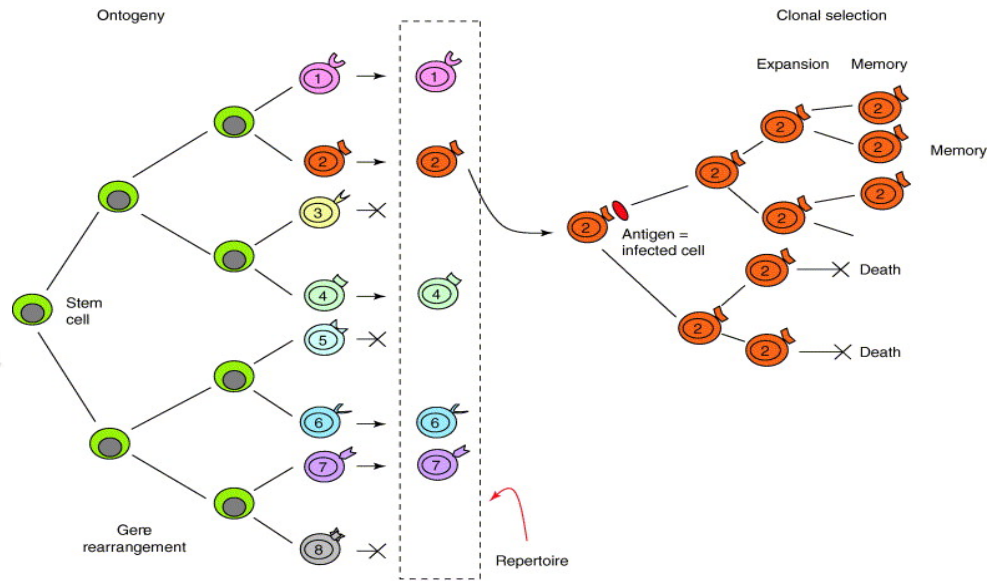
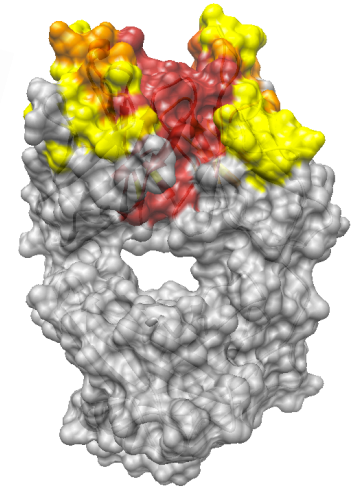
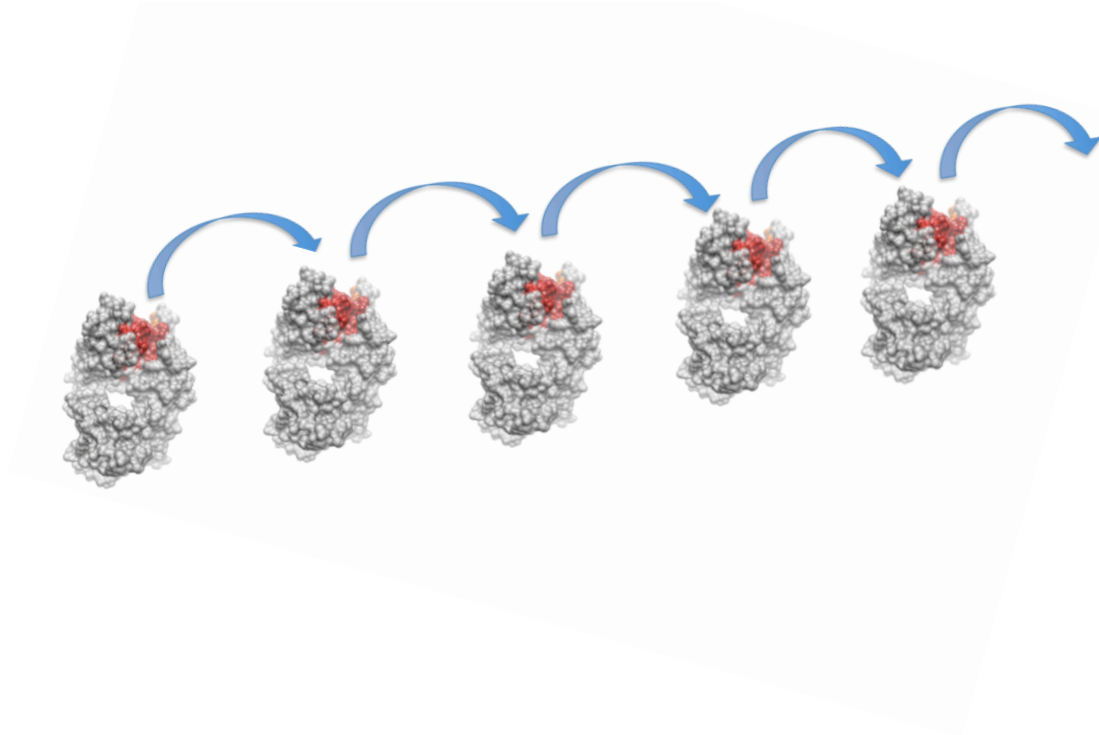
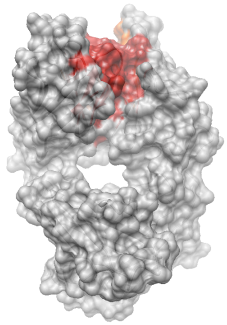
Bioengineering: 8-12 months

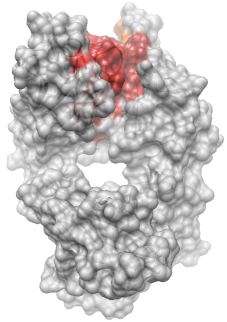


Affinity maturation

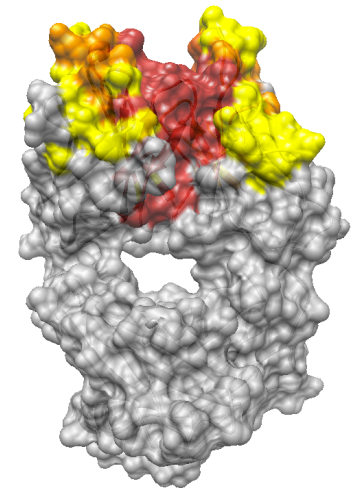
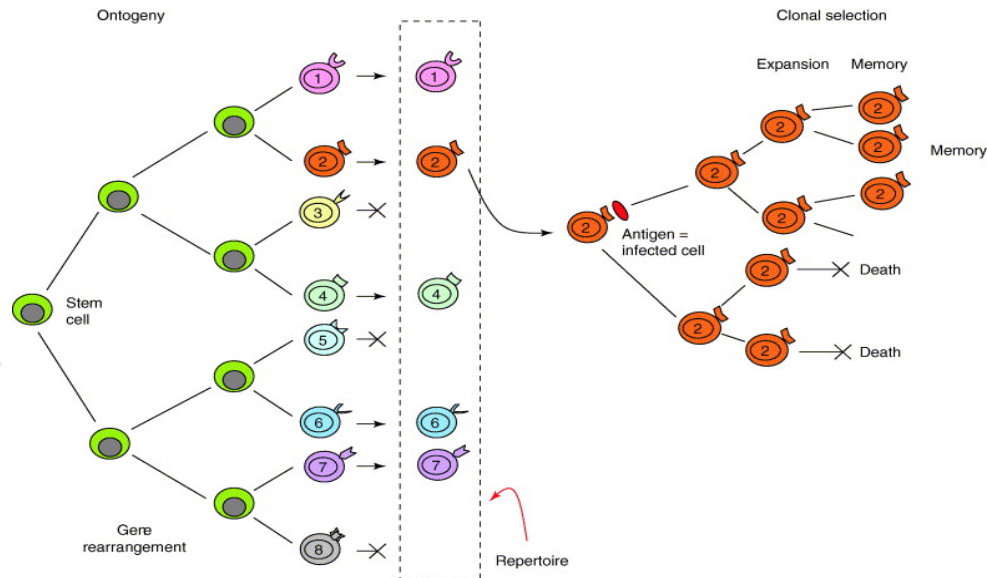
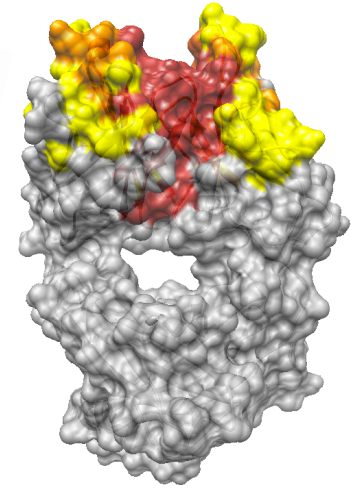
Biochemical liability elimination

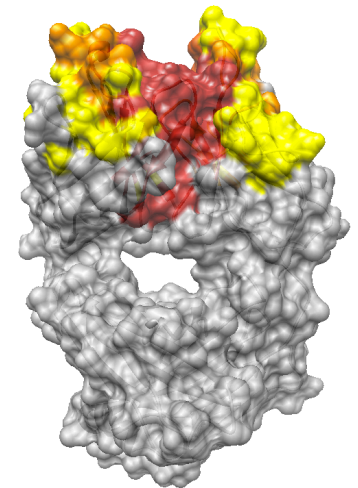
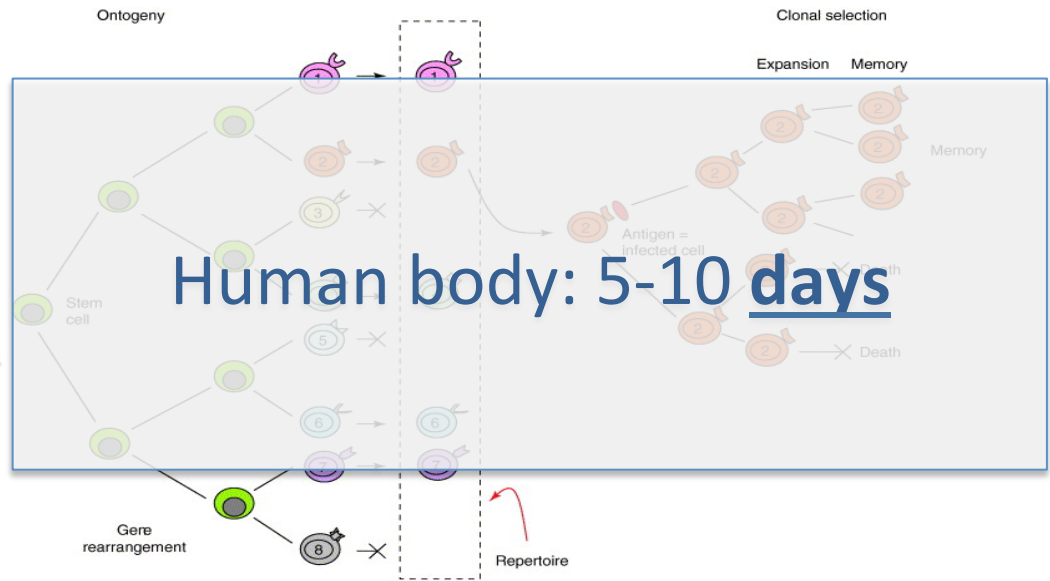
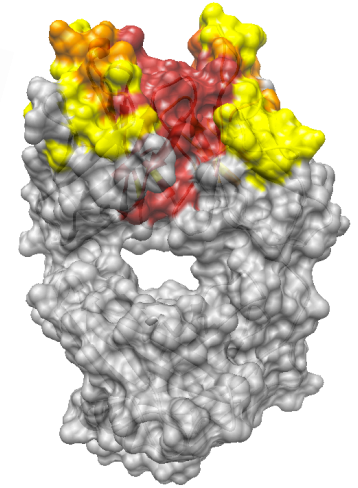
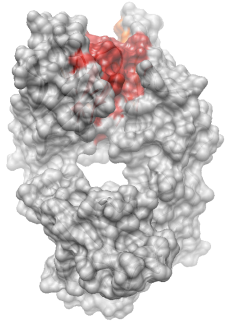
Off-target binding removal

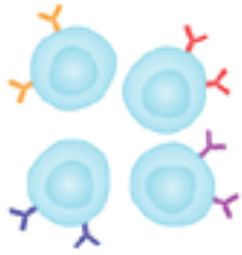




Bioengineering: 8-12 months





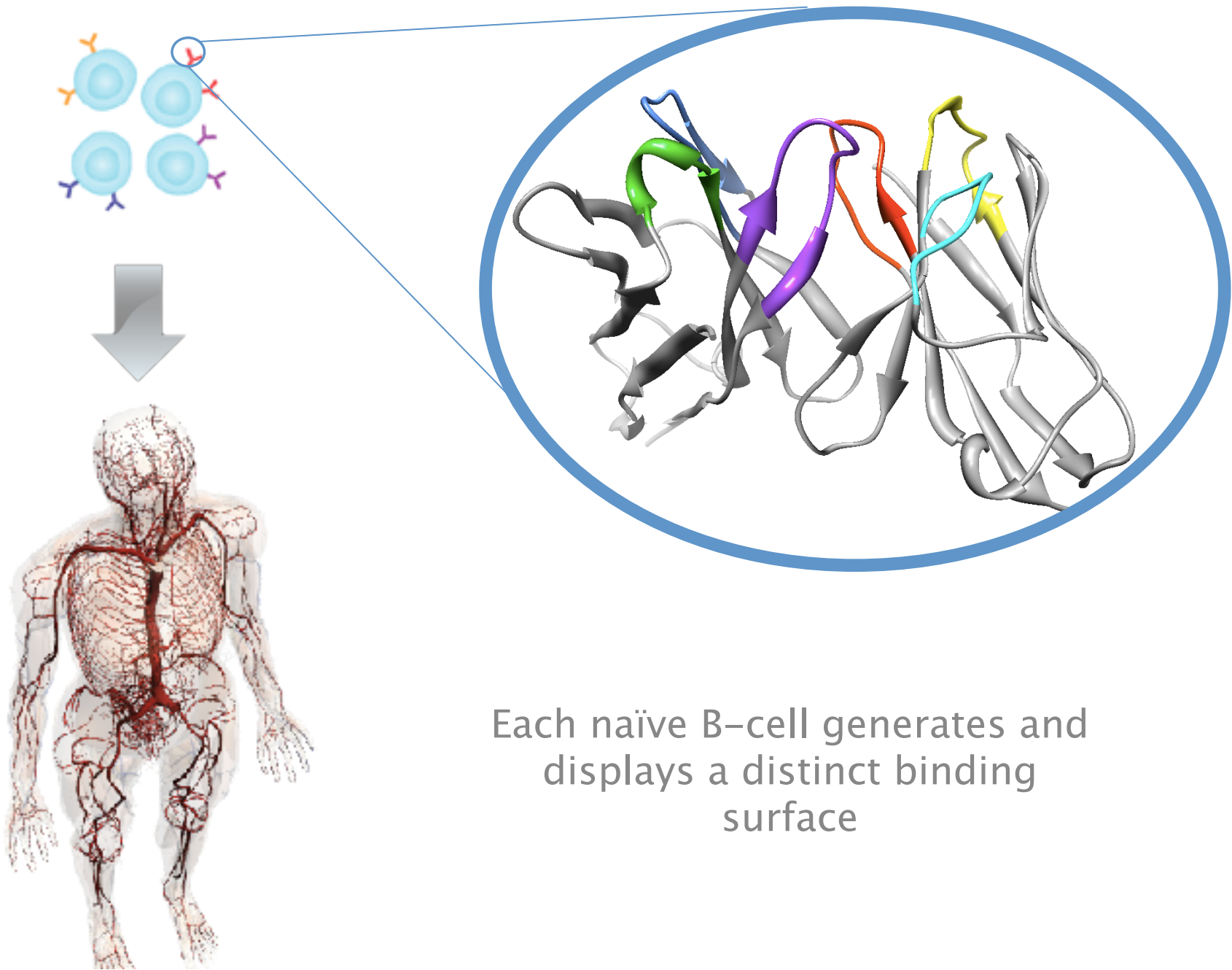


Antibody repertoires

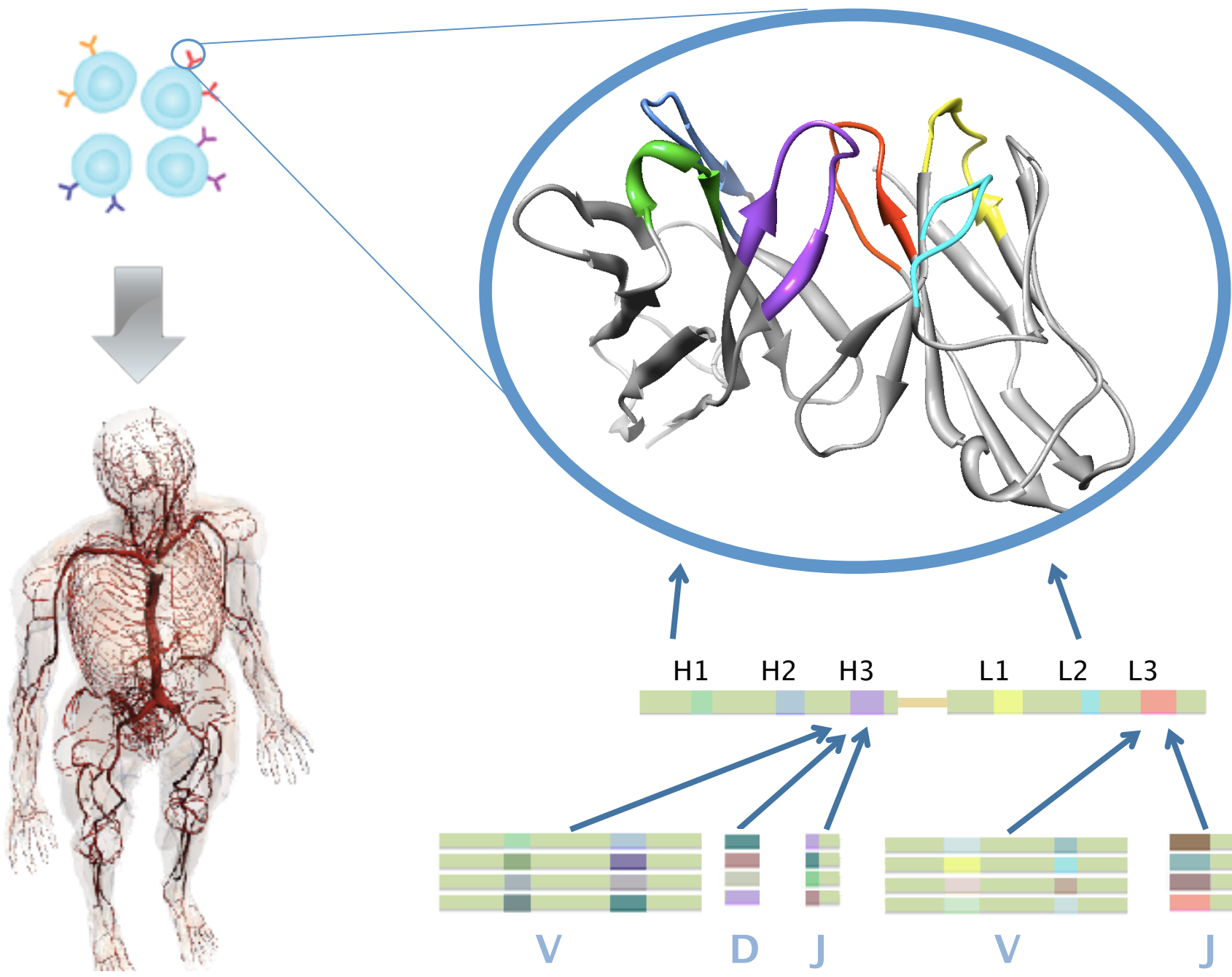


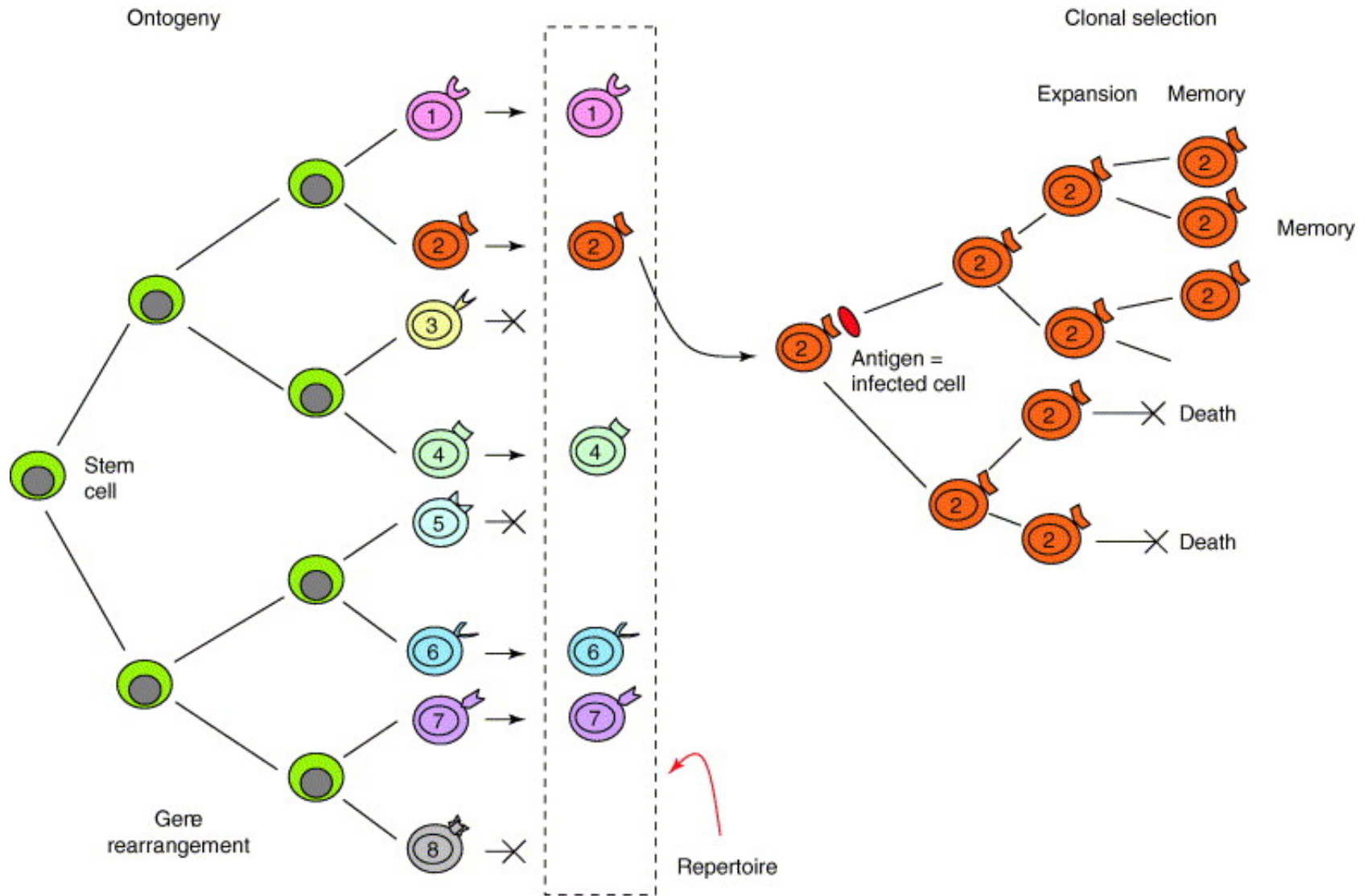
A population of distinct antibodies with a variety of binding specificities

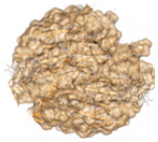




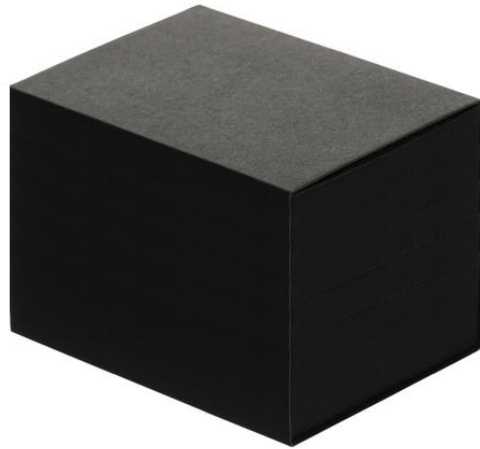
Each naïve B-cell generates and displays a distinct binding surface



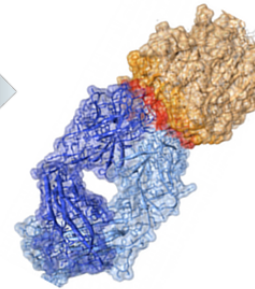




Antigen



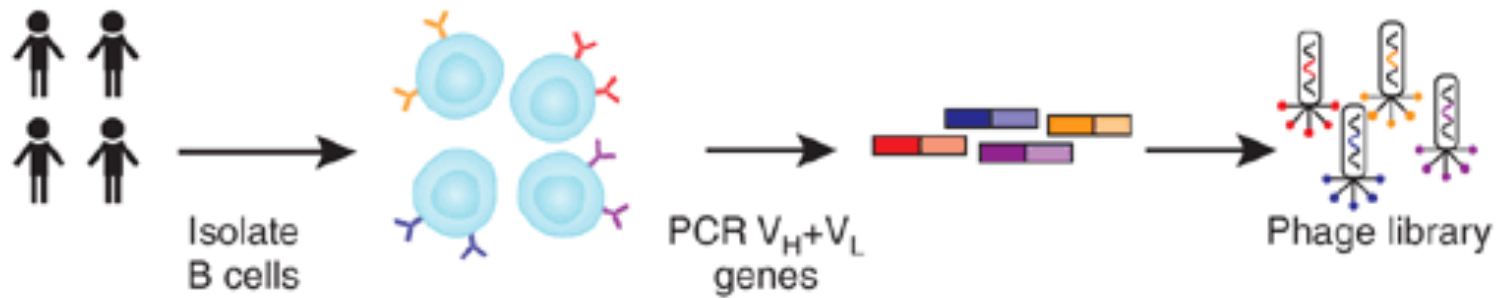
Antibody Repertoire



Anti-antigen Antibody

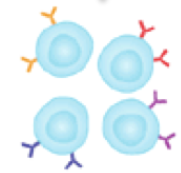
654 human donors

35 billion transformants

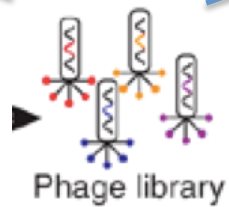
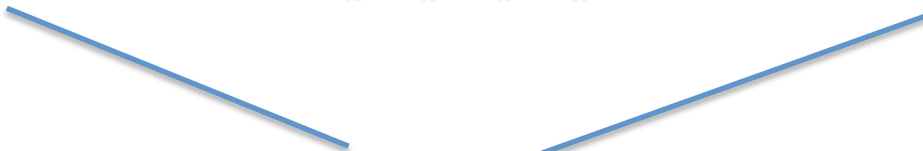




Phage library

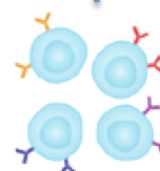


A spleen



Phage library

10-200 hits
+/- 100nM

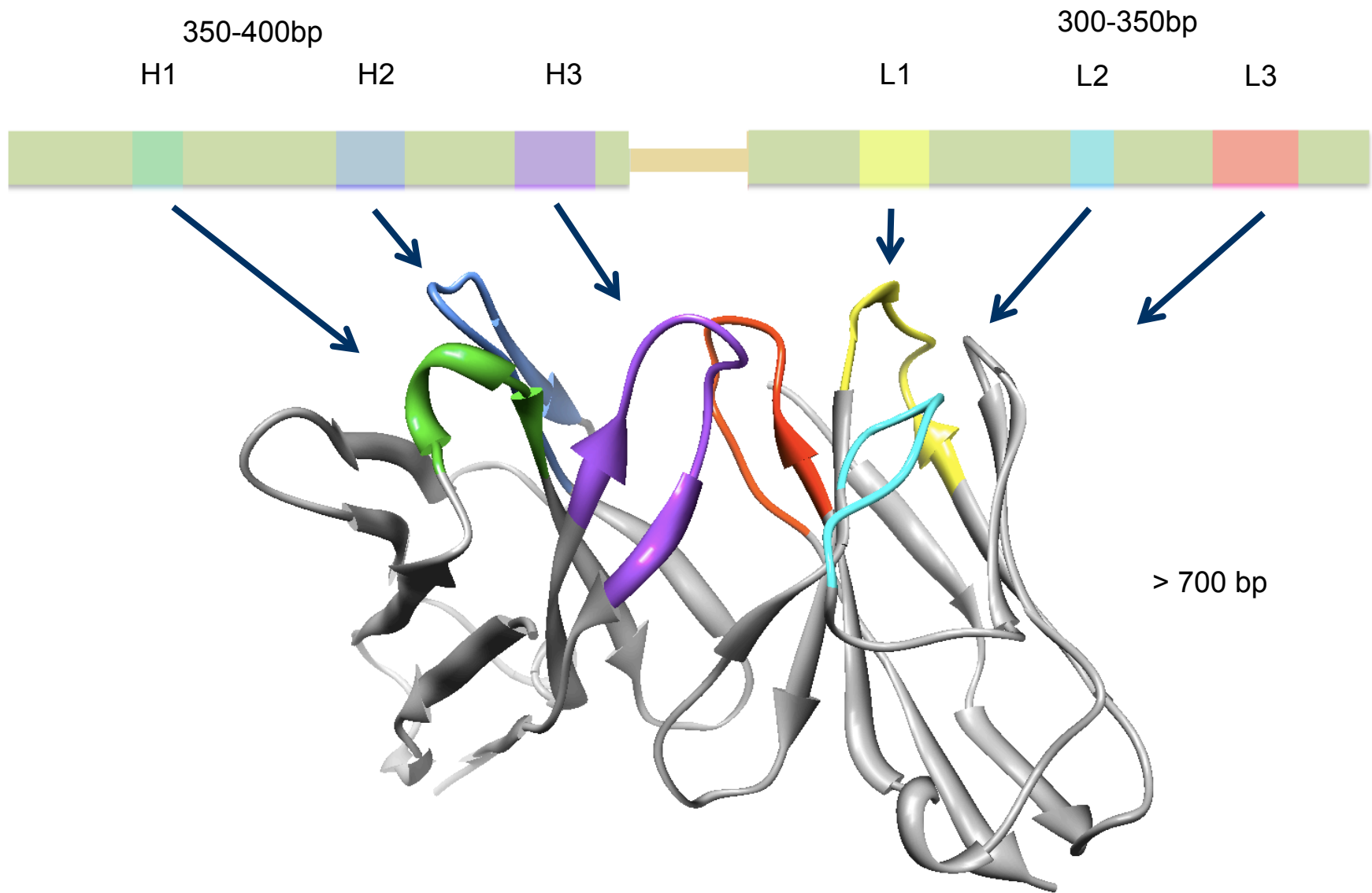


A spleen

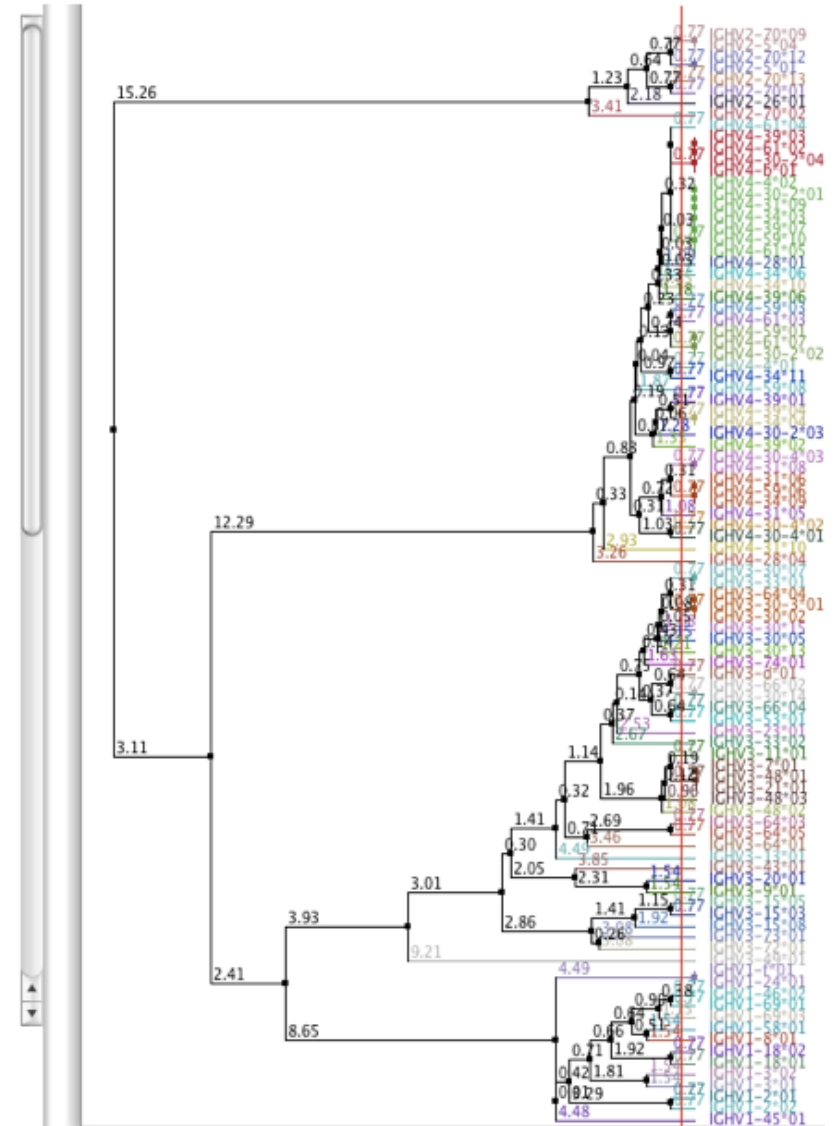
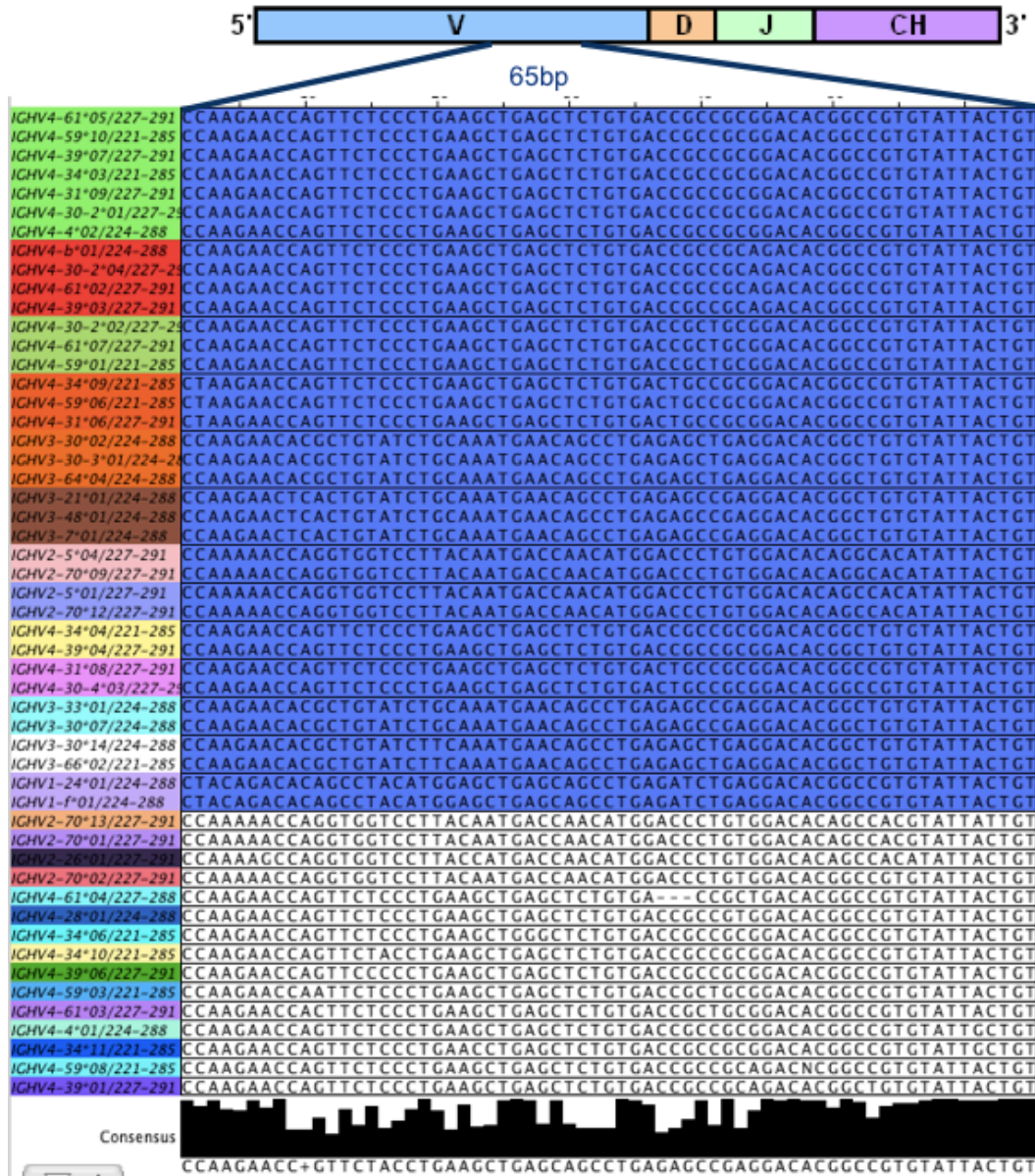
100-300 hits
< 1nM

Sequencing Repertoires

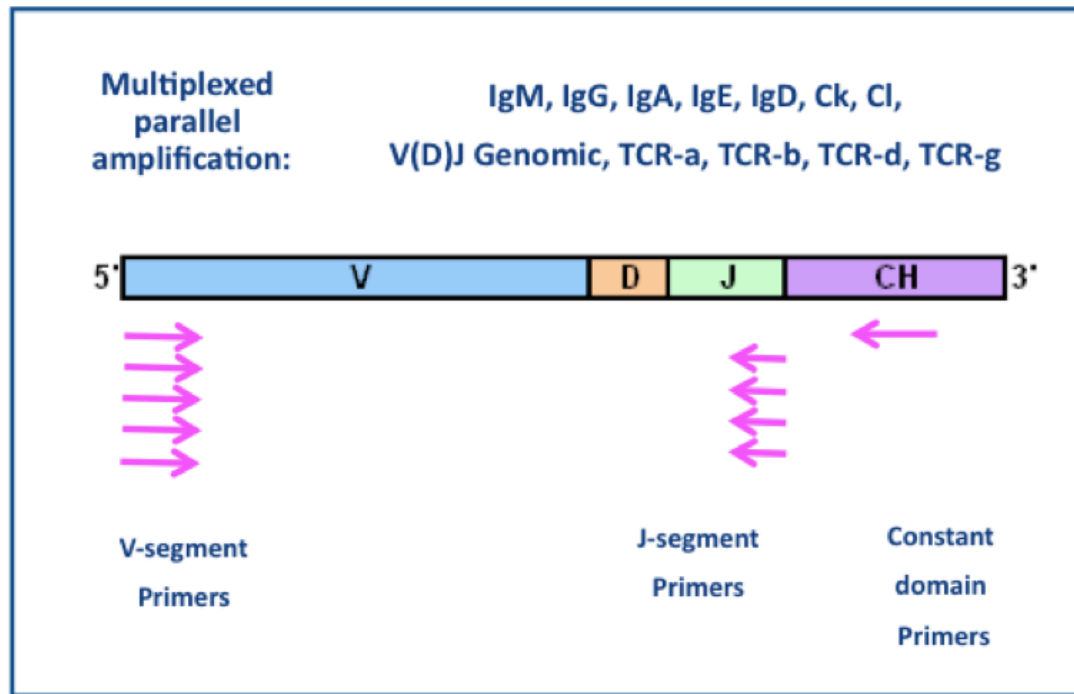
Amplifying repertoires from a synthetic library



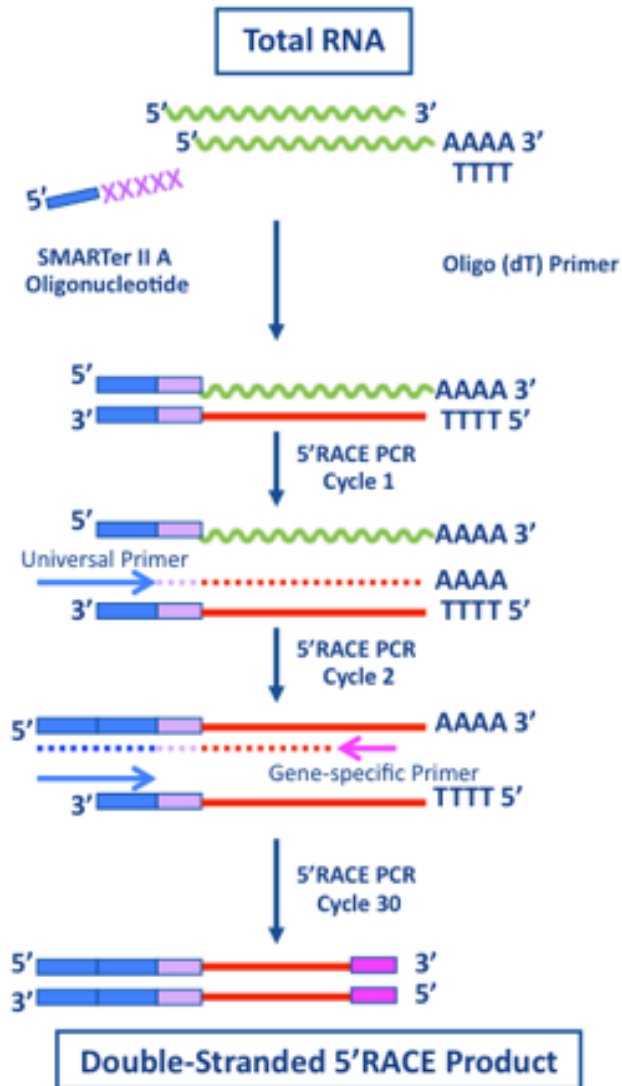
Require long reads without assembly



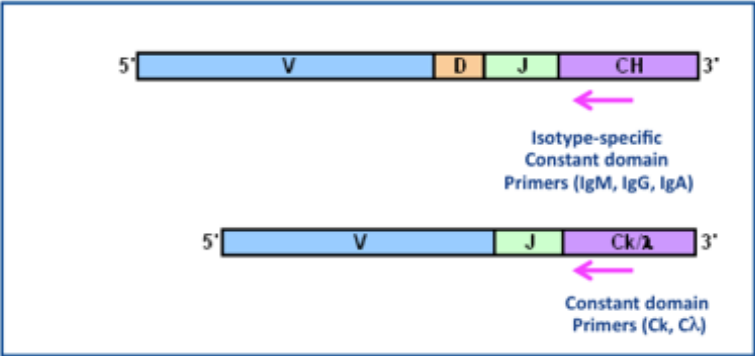
Multiplex primer design & primer bias



Mitigating primer bias with 5'RACE



- Singleplex 5'RACE PCR reaction setup for each constant domain primer
- 5'RACE products gel-purified and pooled for high-throughput sequencing
- Eliminates primer performance bias



Glanville et al, PNAS 2011



VDJFasta new
Brought to you by: jacobglanville

Summary Files Reviews Support Wiki Hosted Apps • Discussion

★ 5.0 Stars (10)
↓ 4 Downloads (This Week)
📅 Last Update: 2013-04-15

Download VDJFasta-1.1.2.tar.gz

Tweet 0 +1 0 /rme 2

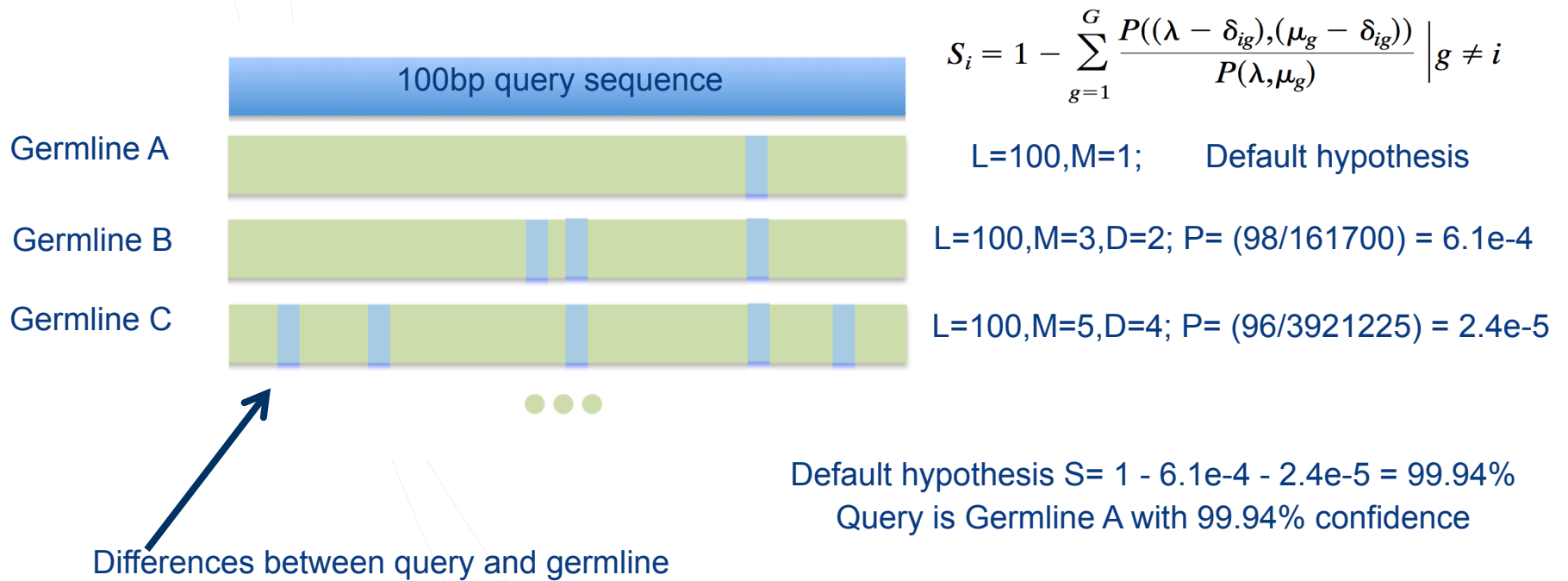
Description

Bioinformatics Perl extension for the analysis of antibody variable domain repertoire mammalian repertoire sequences obtained either by Sanger or 454 sequencing. Me Glanville, Zhai, Berka et al, PNAS 2009.

VDJFasta

Avoiding germline mis-classification

Ask “what are the odds that mutations in very specific positions would cause me to erroneously classify this sequence?”



Germline classification accuracy

Table S4. V(D)J classification accuracy benchmark: V segments

| Mutations | V-seg (%) | V-seg errors (%) |
|-----------|-----------|------------------|
| 1 | 100 | 0.00 |
| 5 | 100 | 0.00 |
| 10 | 100 | 0.00 |
| 15 | 100 | 0.00 |
| 20 | 100 | 0.00 |
| 25 | 99.9 | 0.01 |
| 30 | 99.8 | 0.02 |

V segments are classified by VDJFasta probabilistic classifier (1). This table illustrates the percentage of accuracy in v-segment classification (column 2) and error (column 3) with variable number of mutations in the v-gene (column 1). Classification quality was determined by simulating V(D)J rearrangements, simulating substitutions in the IMGT germ-line reference sequences, and then attempting classification. Classification was accurate if the correct germ line was identified. Classification was in error when the wrong germ line was classified. Classification was "ambiguous" when the correct germ line could not be determined with confidence. Erroneous classification remained below 0.3% for all simulations, indicating that sequences could be underclassified, but were rarely being incorrectly classified. Simulations were performed in the VDJFasta tool "fasta-vdj-sim.pl." Parameters for V were as follows: expectation value cutoff 1e-10, 1e-2, 1e-3, 1e-4; ambiguous-hit score (1) 1e-3, 1e-1, 1e-3, 1e-3; minimum segment alignment length (1) 100, 8, 35, 30.

Table S5. V(D)J classification accuracy benchmark: D and J segments

| Mutations | D-seg (%) | J-seg (%) | D-seg errors (%) | J-seg errors (%) |
|-----------|-----------|-----------|------------------|------------------|
| 0 | 96.3 | 100 | 0.21 | 0.00 |
| 1 | 92.9 | 99.9 | 0.13 | 0.00 |
| 2 | 78.1 | 99.7 | 0.34 | 0.00 |
| 3 | 62.9 | 99.6 | 0.24 | 0.00 |
| 4 | 56.2 | 99.0 | 0.17 | 0.02 |
| 5 | 47.8 | 98.5 | 0.23 | 0.09 |
| 6 | 36.5 | 96.9 | 0.27 | 0.22 |

V(D)J segments are classified by VDJFasta probabilistic classifier (1). This table illustrates the percentage of accuracy in D- and J-segment classification (column 2 and 3) and error (column 3 and 4) with variable number of mutations in each gene segment (column 1). D-segment classification additionally requires that both a V_H and J_H segment have been identified, and that the putative D-segment classification is found between them on the sequence, with up to 8-bp overlap tolerated to account for overalignment of segments to the query. Classification quality was determined by simulating V(D)J rearrangements, simulating substitutions in IMGT germ-line reference sequences, and then attempting classification. Classification was accurate if the correct germ line was identified. Classification was in error when the wrong germ line was classified. Classification was "ambiguous" when the correct germ line could not be determined with confidence. Ambiguous classification rates increased rapidly for D-segments as more SHM was introduced, suggesting that D-segment analysis in antigen-experienced compartments would be challenging. Erroneous classification remained below 0.3% for all simulations, indicating that sequences could be underclassified, but were rarely being incorrectly classified. Simulations were performed in the VDJFasta tool "fasta-vdj-sim.pl." Parameters for D and J, CH1 were as follows: expectation value cutoff 1e-10, 1e-2, 1e-3, 1e-4; ambiguous-hit score (1) 1e-3, 1e-1, 1e-3, 1e-3; minimum segment alignment length (1) 100, 8, 35, 30.

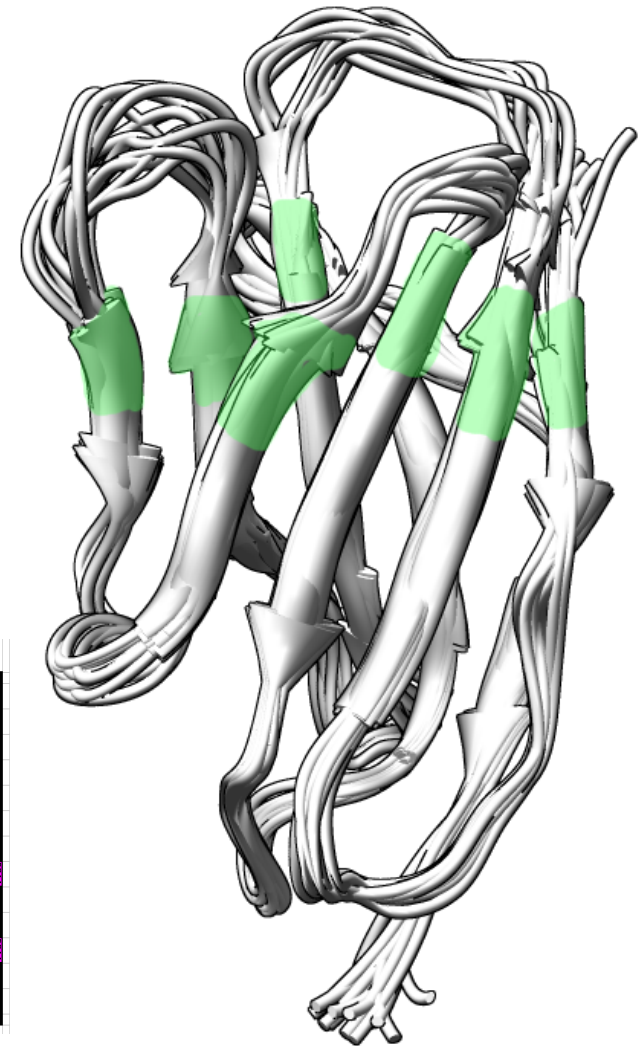
CDR recognition with profile HMMs

HMM CDR recognition was evaluated structurally

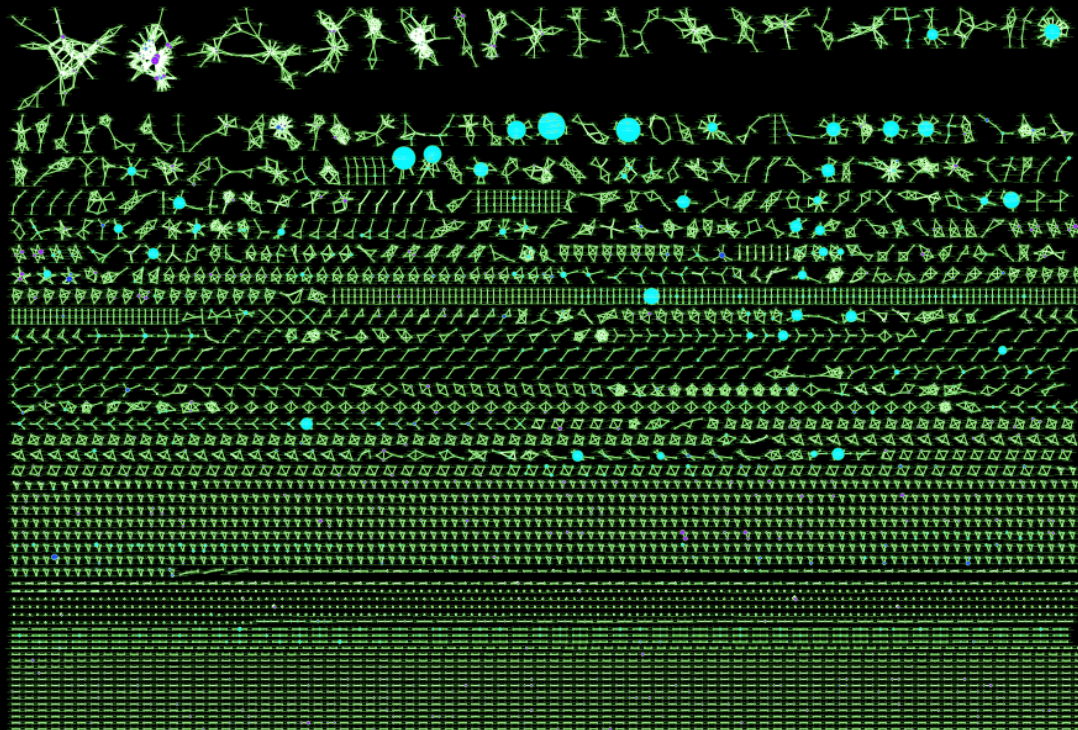
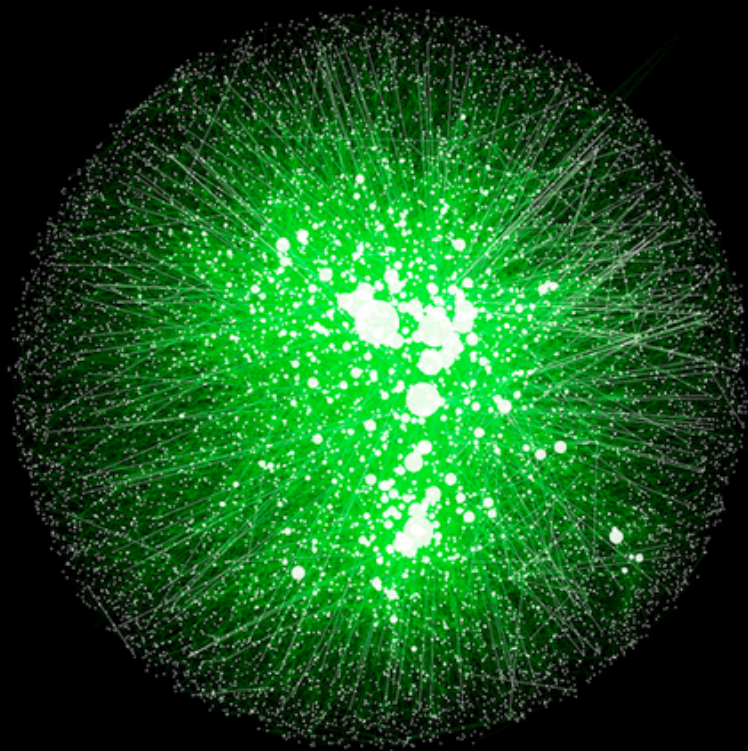
- ▶ 779 non-redundant structures were superposed
- ▶ Sequences of reference structures were extracted
- ▶ Reference structure sequences were aligned to HMM
- ▶ HMM-Predicted boundary positions were compared to structure

HMM CDR recognition was highly accurate

- ▶ 99.74% boundary recognition success
- ▶ Two miscalls were both in C-terminal H3
- ▶ One was a catalytic antibody



| | Framework 1 | CDR 1 | Framework 2 | CDR 2 | Framework 3a | Framework 3b | CDR 3 | Framework 4 |
|-------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| AHo | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N |
| Kabat VA | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N |
| Kabat Vc | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N |
| Chotia VL (preB9) | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N |
| Chotia VL 89-97 | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N |
| Kabat VH | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N |
| Chotia VH | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N |
| Kabat TCR Va | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N |
| Kabat TCR Vb | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N |
| Kabat TCR Vy | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N |
| Kabat TCR Vb | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N |
| IMGT | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N |
| AHo | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N | D-N-N-N-P-P-P-P-S-S | F-N-N-N-N-N-N-N-N-N |

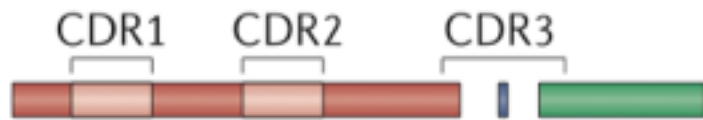


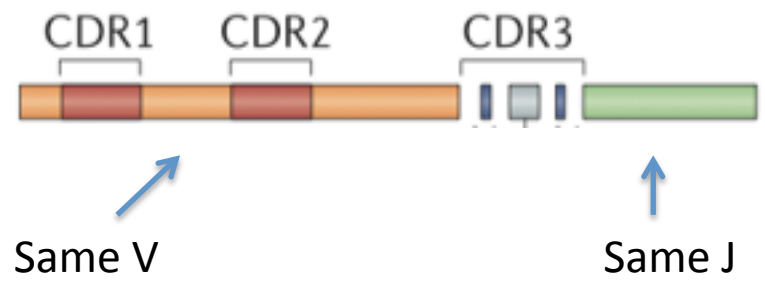
Repertoire

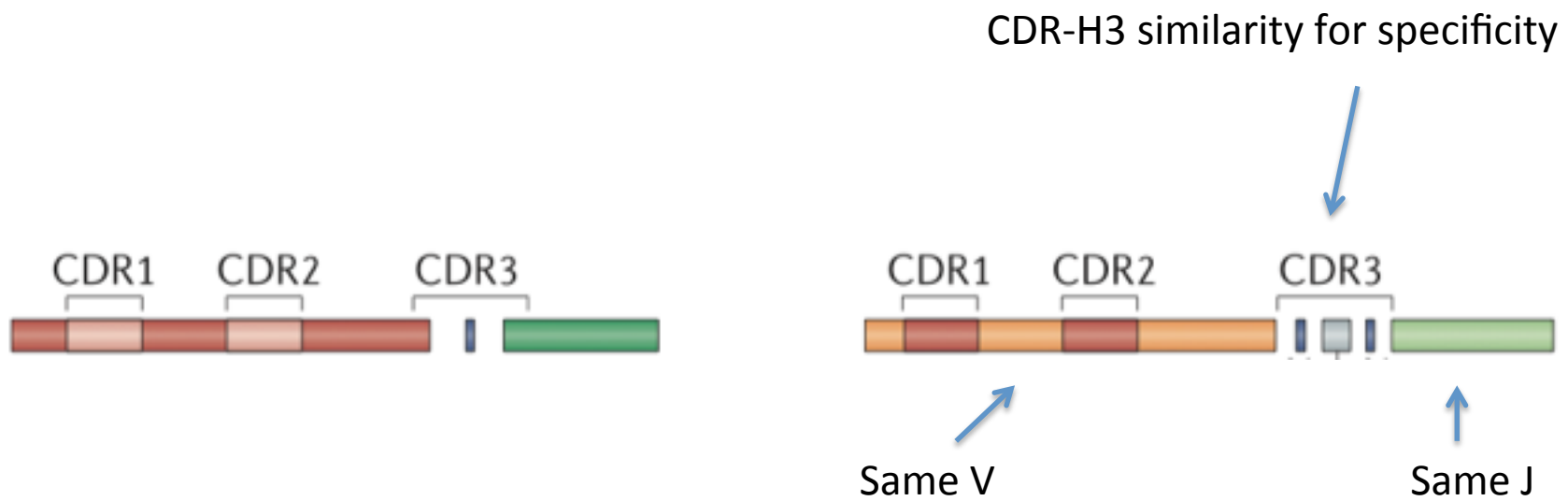
-->

Clonal lineages

Can we do something smarter than alignment distance?



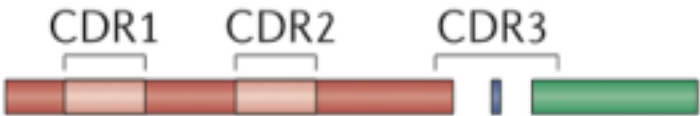




Xu, John L., and Mark M. Davis. "Diversity in the CDR3 Region of VH Is Sufficient for Most Antibody Specificities." *Immunity* 13.1 (2000): 37-45.

SHM for directionality

CDR-H3 similarity for specificity



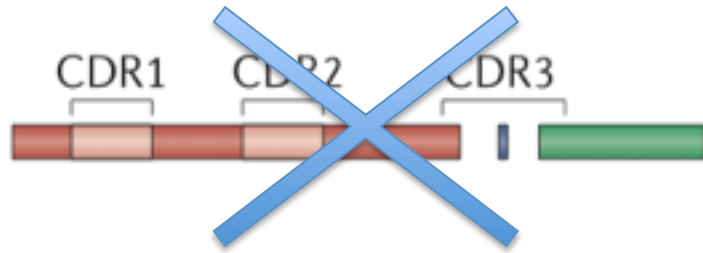
Same V

Same J

Spurn the light chain

SHM for directionality

CDR-H3 similarity for specificity



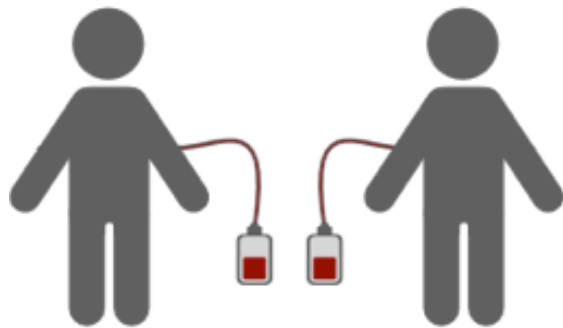
Same V

Same J

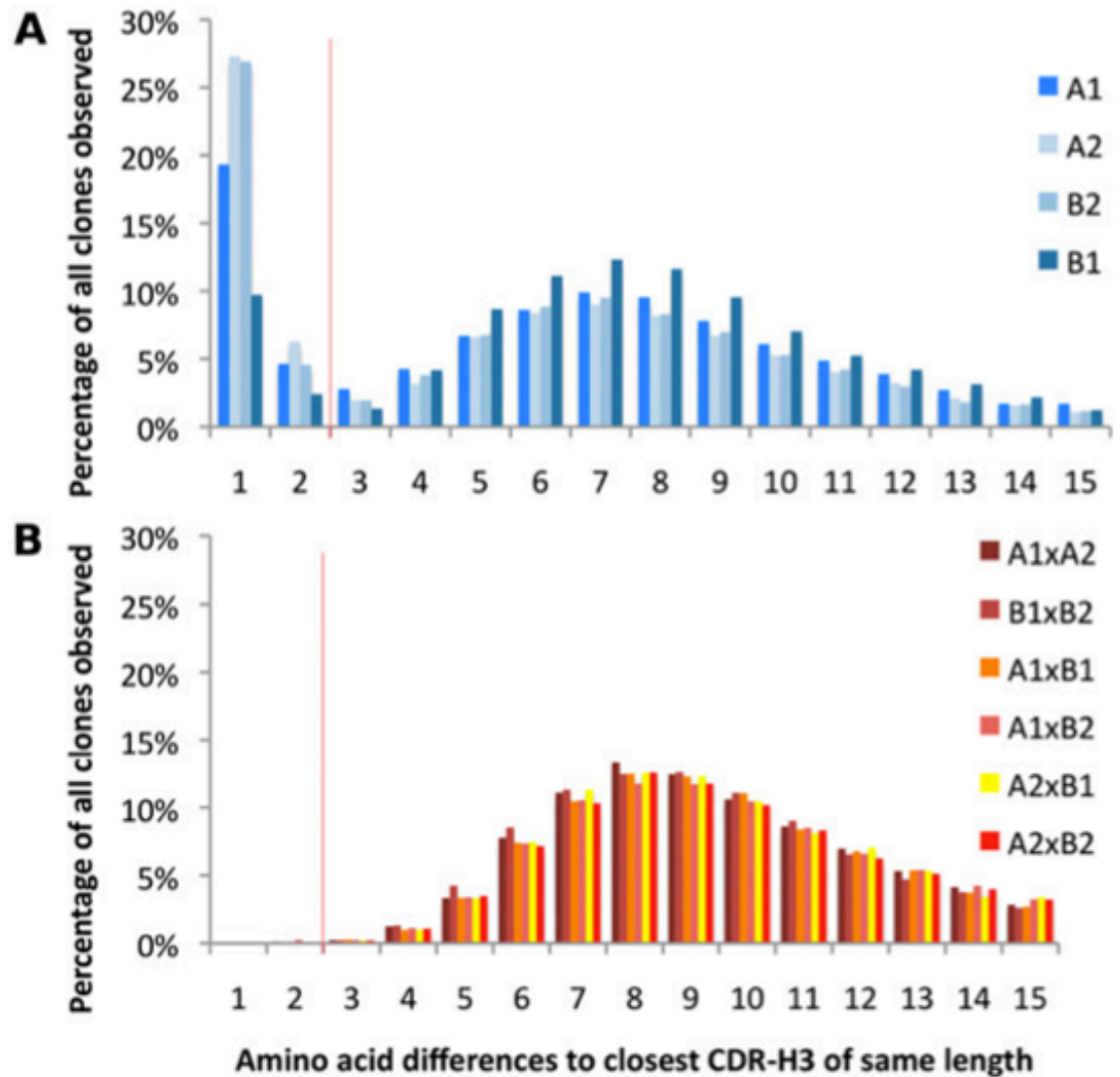
Empirical clone distance cutoff determination



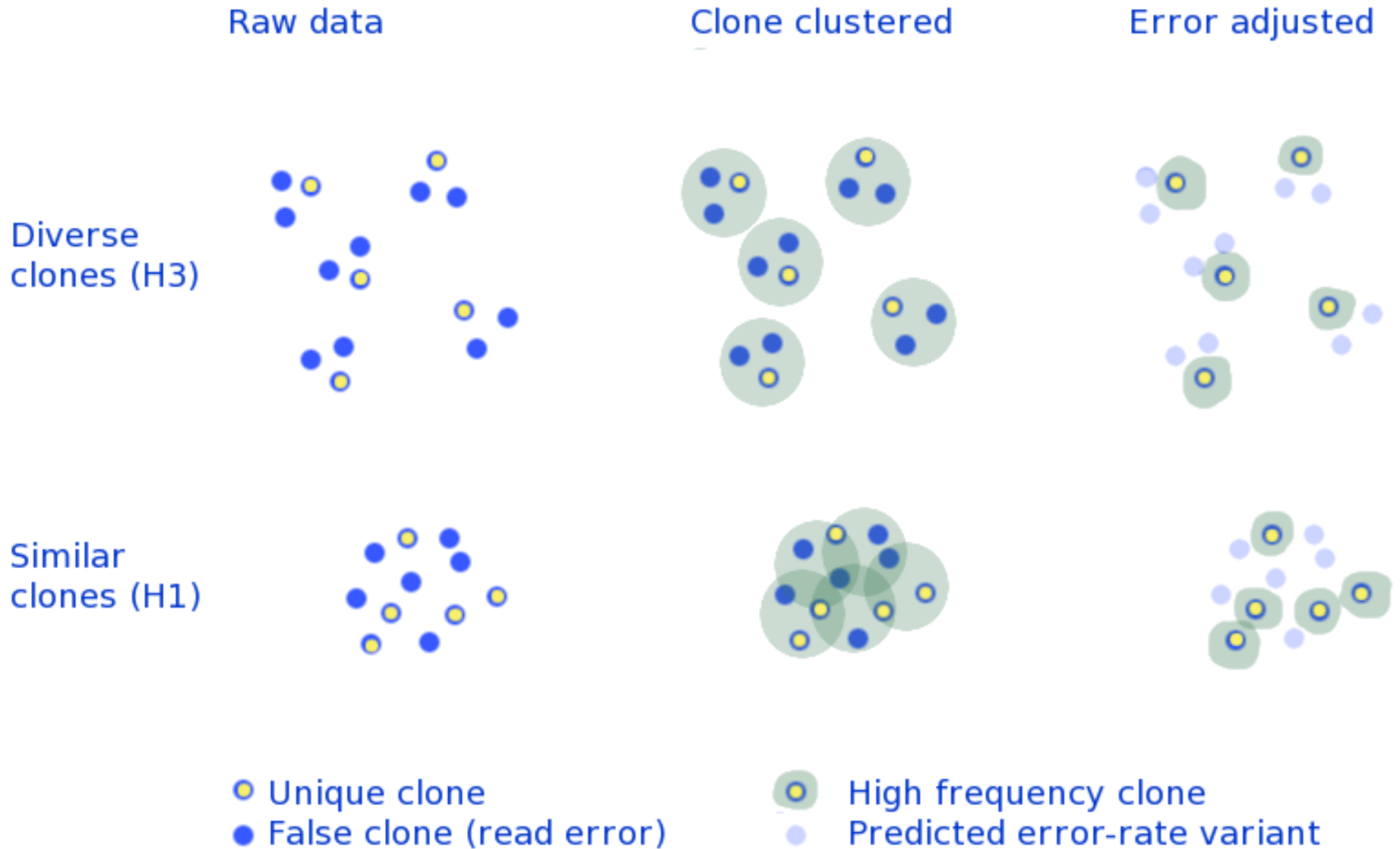
WITHIN
INDIVIDUAL

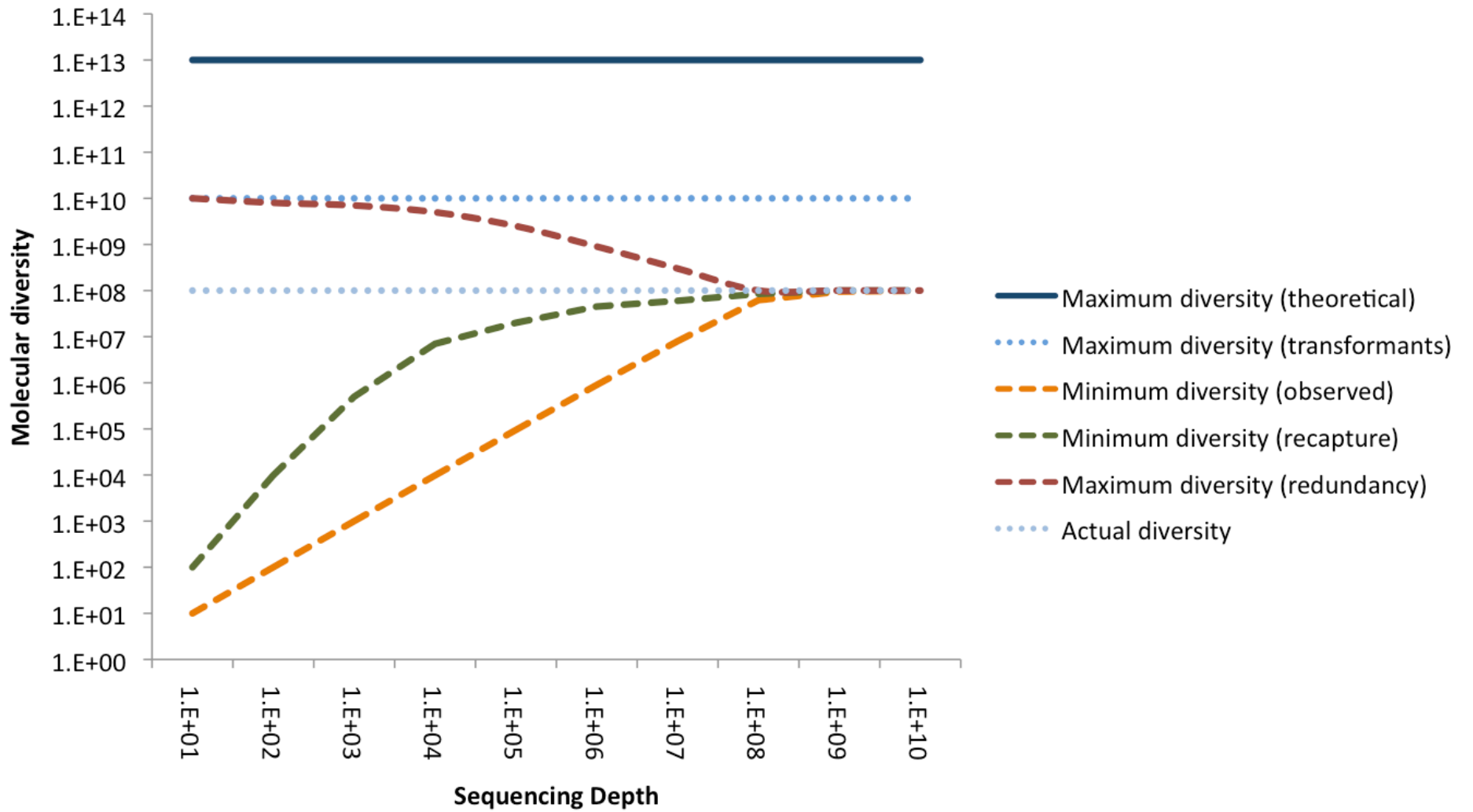


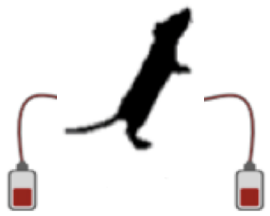
BETWEEN
INDIVIDUALS

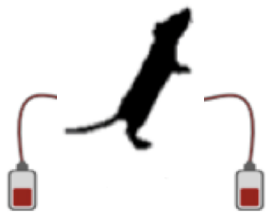


Clone clustering and error clustering









10^{14}



Theoretical IgH naïve V(D)J repertoire

10^{13}

10^{12}

10^{11}

10^{10}

10^9

10^8

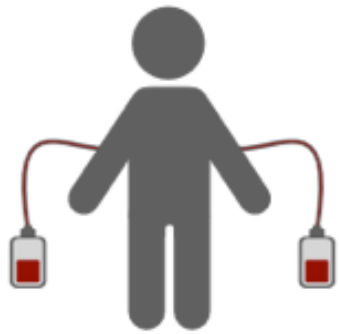
Total B-cells in a mouse

10^7

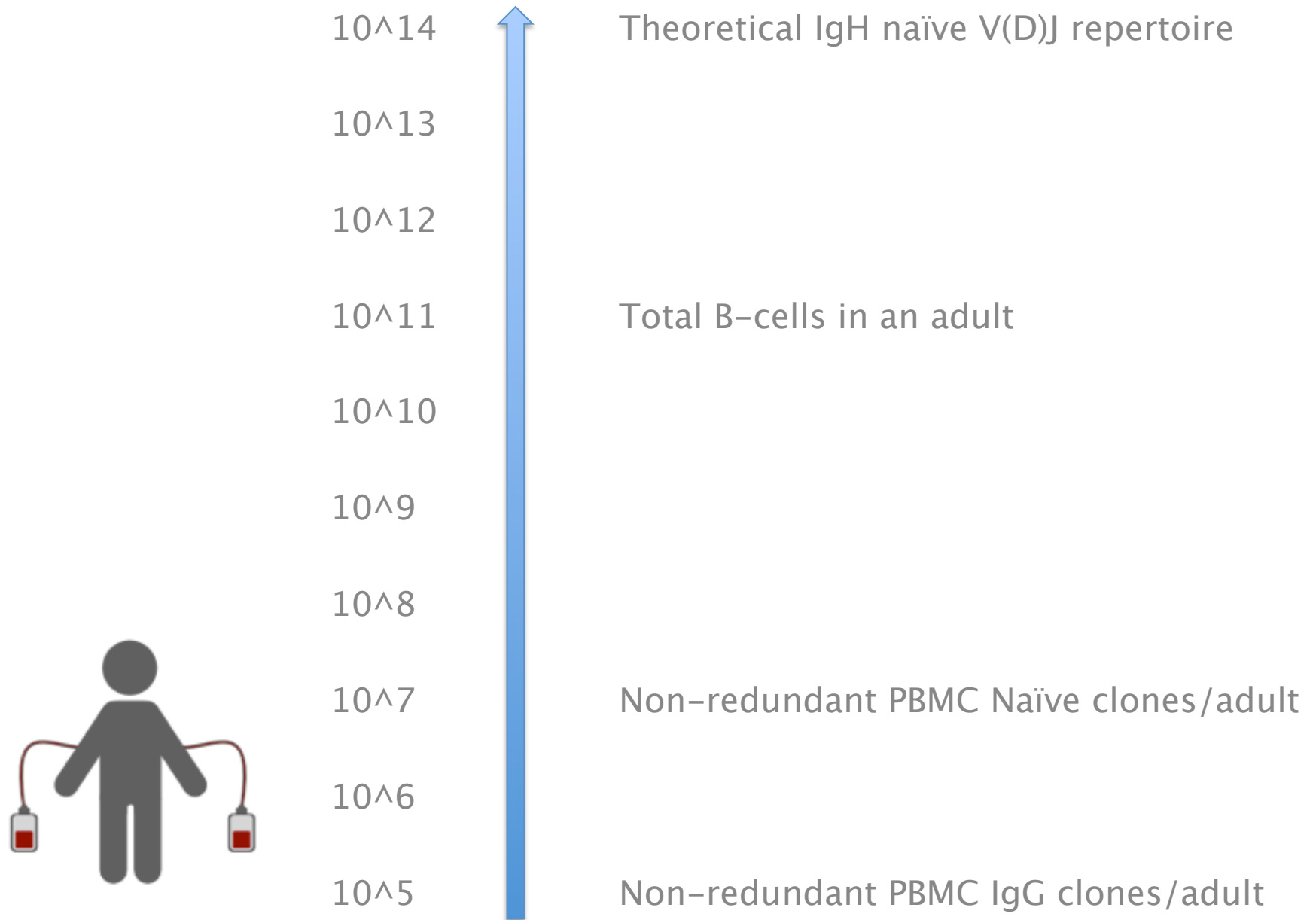
Unique clones in a mouse

10^6

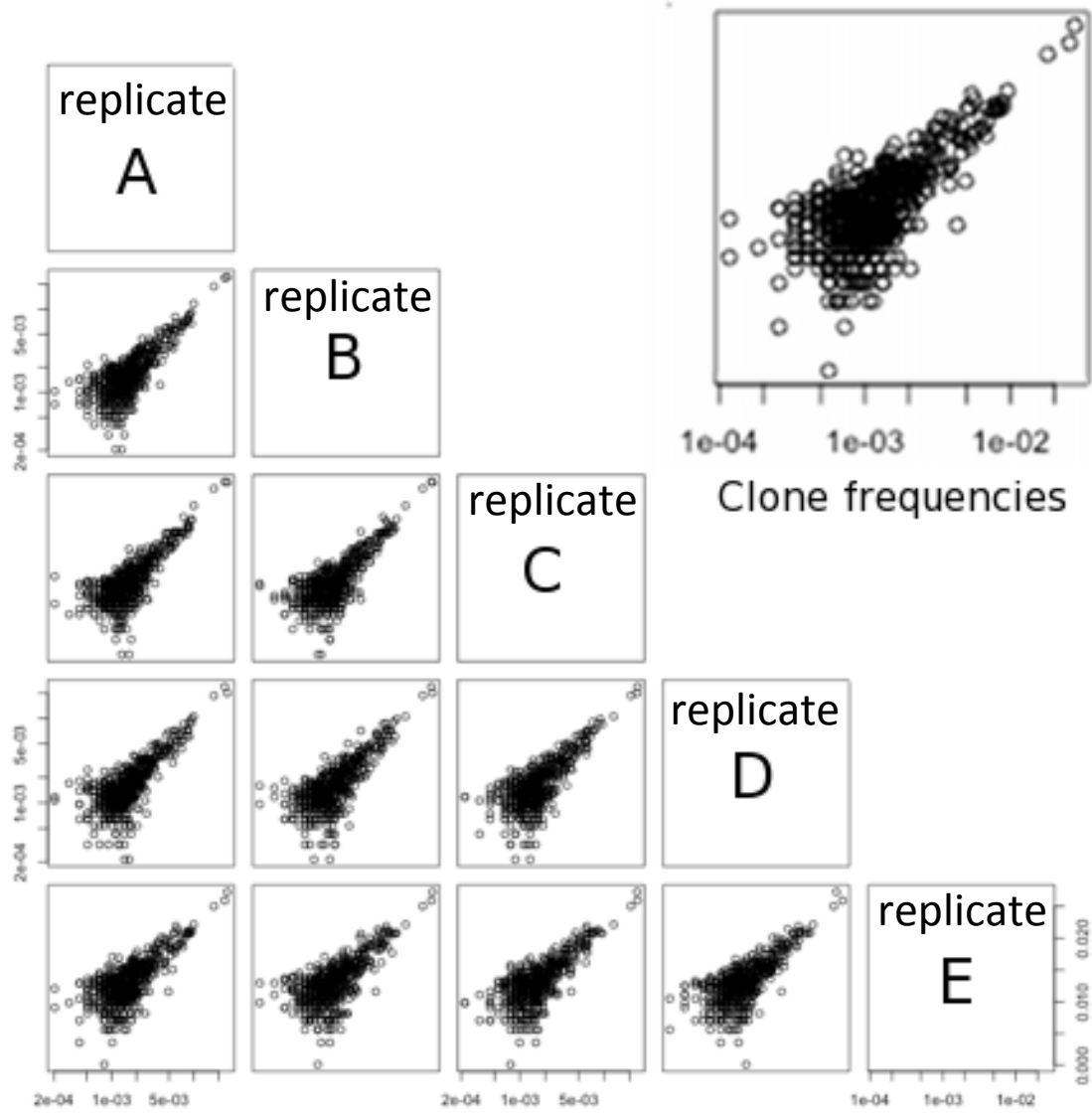
10^5



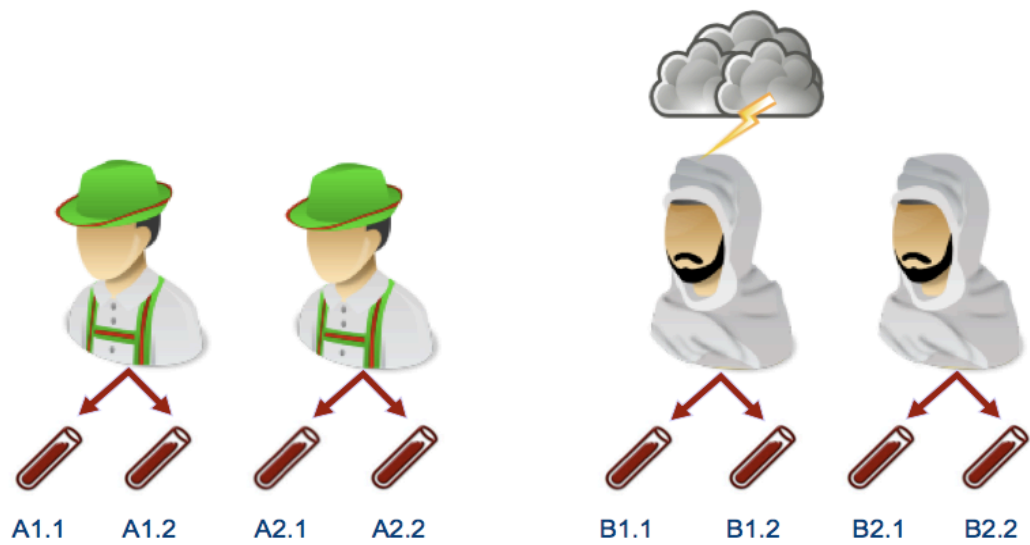
Individual adults only sample a small part of the antibody repertoire



Redundant clones constrain library diversity

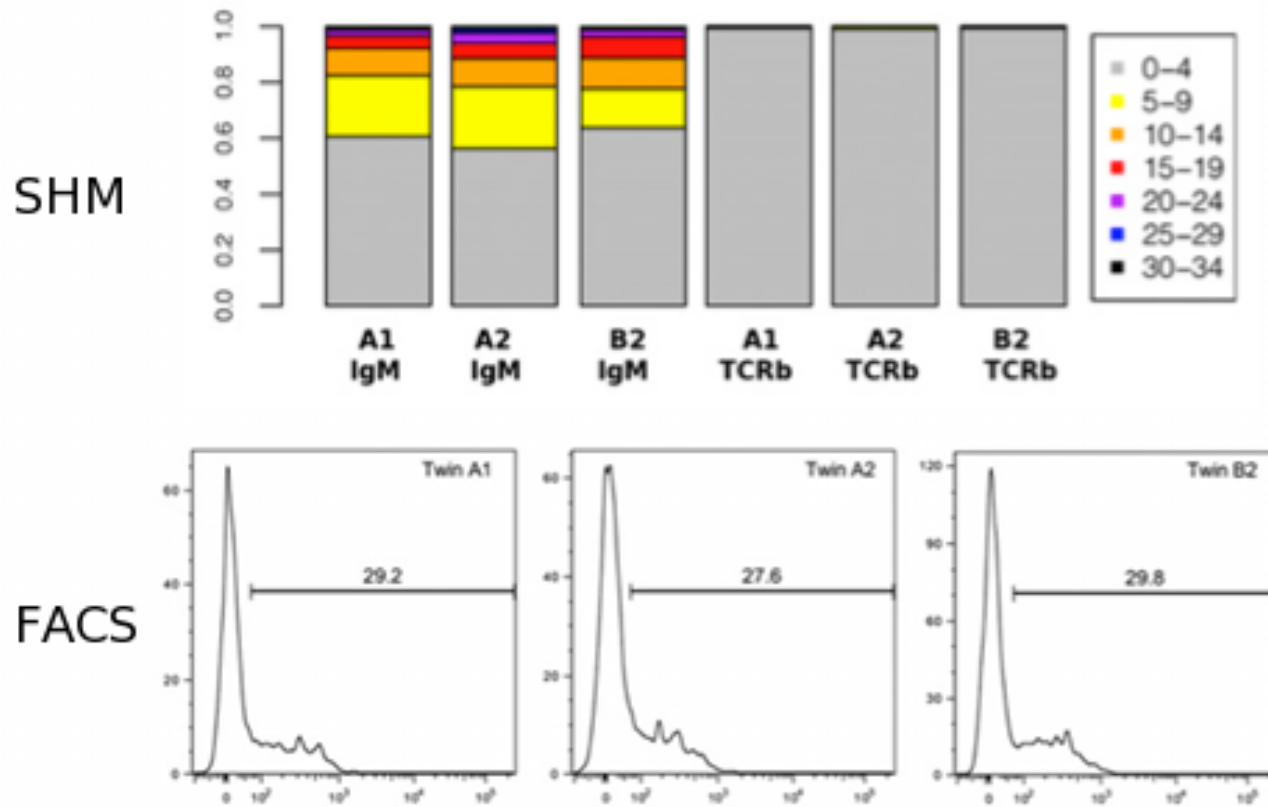


Twins

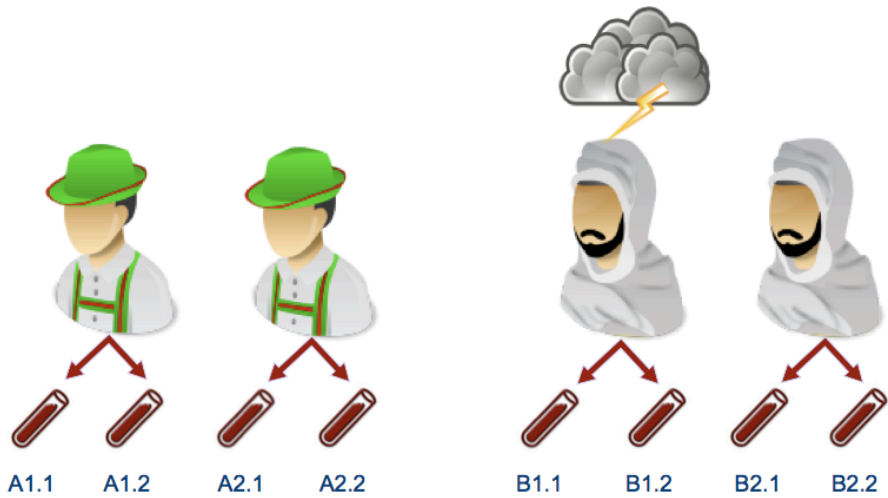
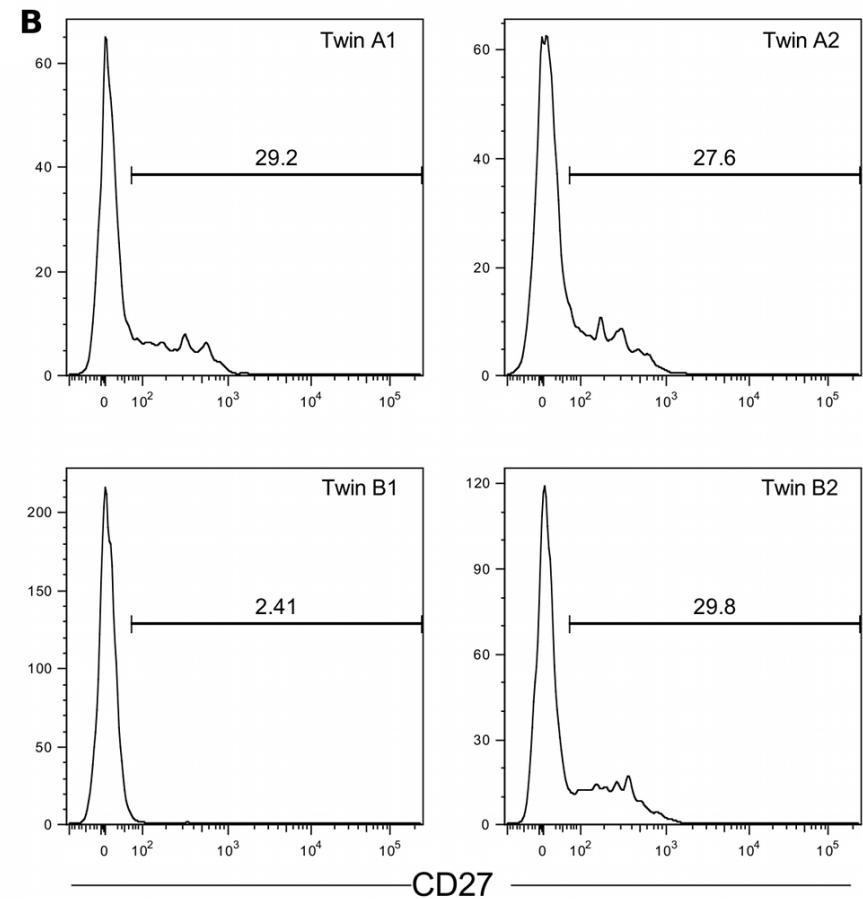
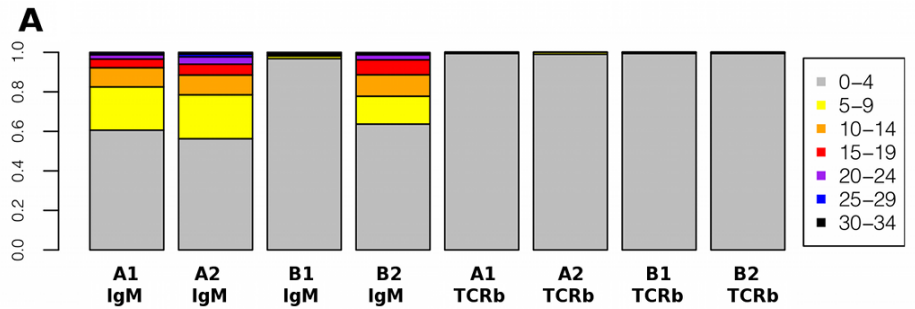


<https://www.youtube.com/watch?v=wNW0vZiaMuA>

Detecting somatic hypermutation



Lymphocyte depletion in affected twin



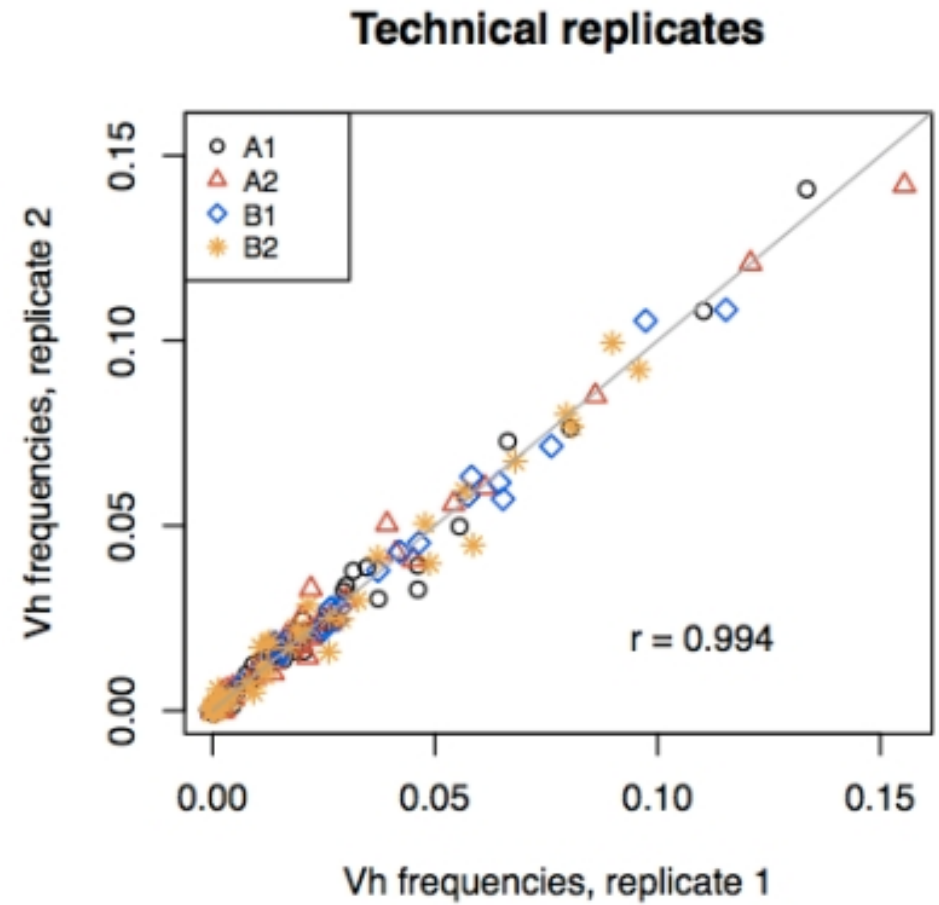
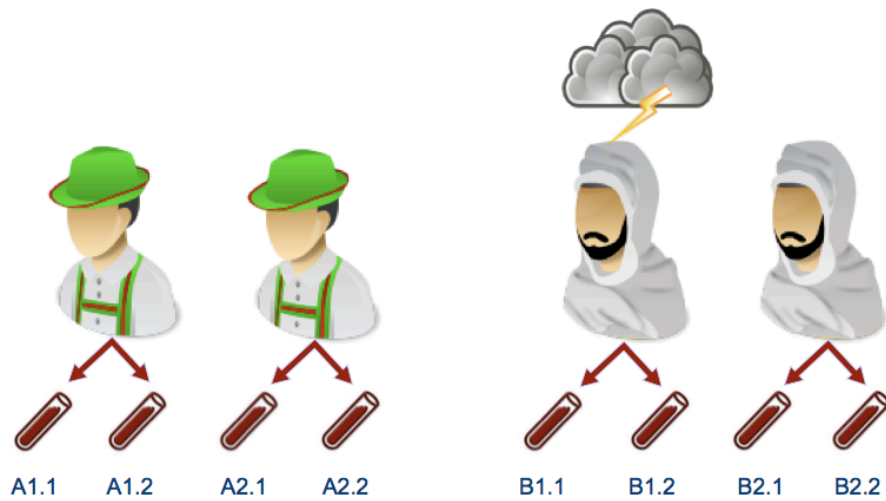
CD27

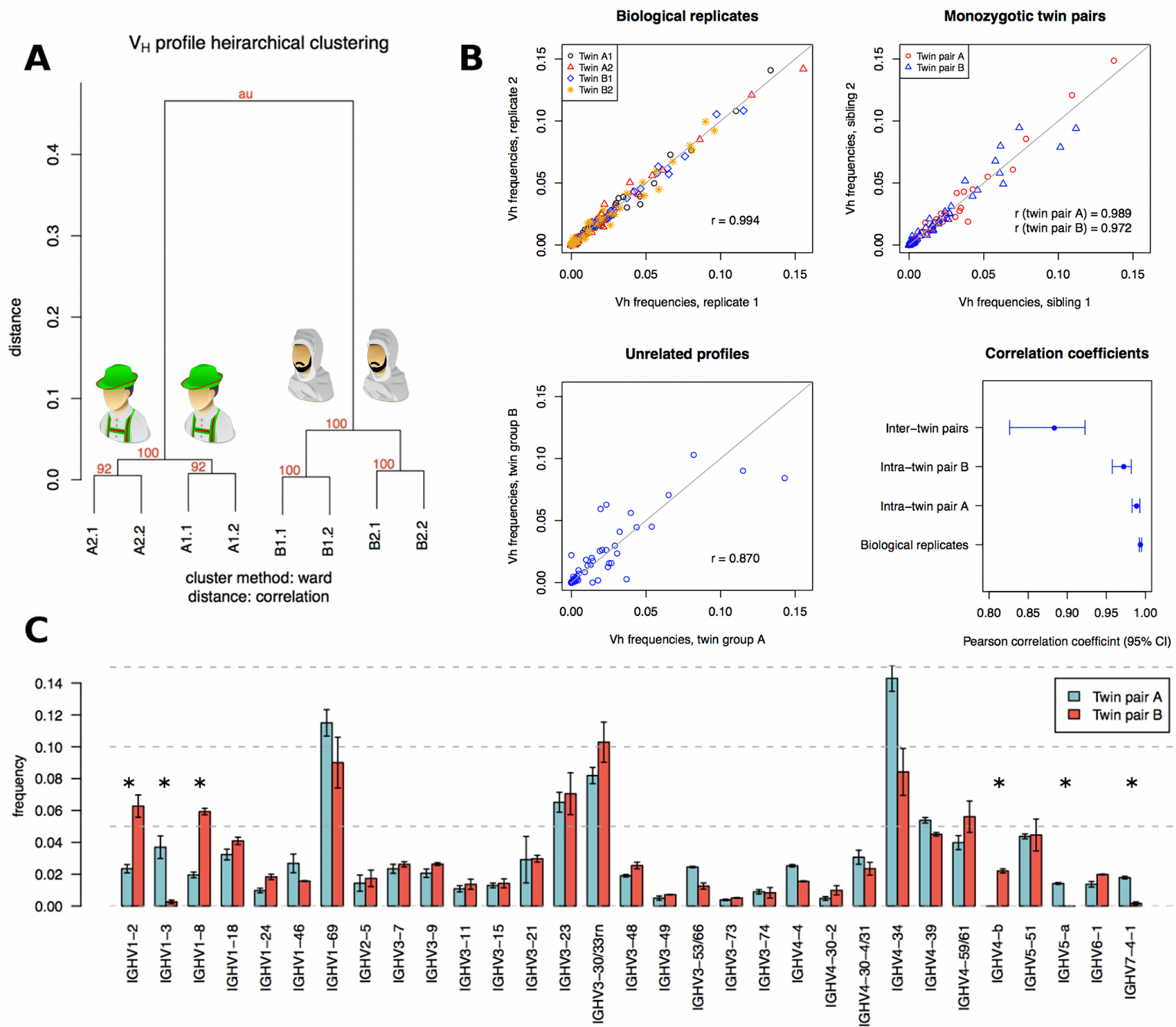
MINIMAL CLONAL OVERLAP

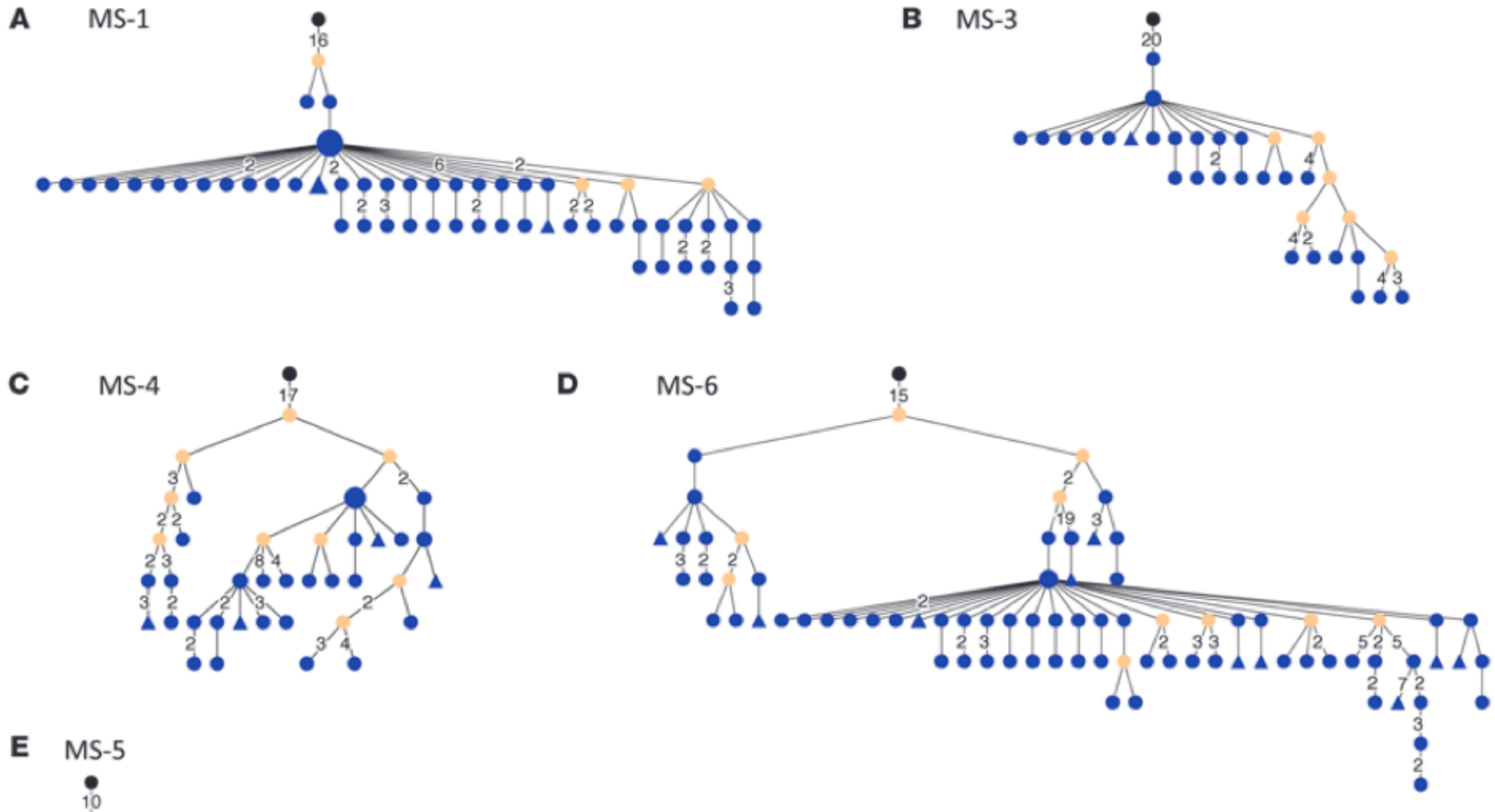
| | | A1 | | | A2 | | | B1 | | | B2 | | |
|----|-----|------|------|------|-------|------|------|-------|------|------|------|------|------|
| | | IgM | IgG | IgA | IgM | IgG | IgA | IgM | IgG | IgA | IgM | IgG | IgA |
| A1 | IgM | 9131 | 308 | 28 | | | | | | 2 | 2 | 1 | |
| | IgG | 8% | 3761 | 143 | | | | | | | 1 | 1 | |
| | IgA | 0.5% | 4% | 5604 | | | | 7 | | | 1 | 2 | |
| A2 | IgM | 0.0% | 0.0% | 0.0% | 10096 | 53 | 41 | | | 2 | 1 | | |
| | IgG | 0.0% | 0.0% | 0.0% | 4% | 1438 | 87 | | 1 | | | | |
| | IgA | 0.0% | 0.0% | 0.0% | 1% | 6% | 3163 | | | | 1 | | |
| B1 | IgM | 0.0% | 0.0% | 0.1% | 0.0% | 0.0% | 0.0% | 17946 | 39 | 23 | 3 | | |
| | IgG | 0.0% | 0.0% | 0.0% | 0.0% | 0.2% | 0.0% | 6% | 652 | 20 | | | |
| | IgA | 0.1% | 0.0% | 0.0% | 0.1% | 0.0% | 0.0% | 1% | 3% | 1565 | | | |
| B2 | IgM | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 8967 | 32 | 38 |
| | IgG | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 2% | 1414 | 61 |
| | IgA | 0.0% | 0.0% | 0.1% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 2% | 4% | 2089 |

Fig. 4. Unique V(D)J clone overlap between IgM and class-switched IgG and IgA. Unique counts shown in upper diagonal, percent overlap $[(A \cap B)/\min(A,B)]$ in the lower diagonal, and unique clones in sample along diagonal. In the upper diagonal, counts for twins A are in green, for twins B are in orange. \log_{10} heatmap colors sequence counts. Percentage of overlap in the bottom diagonal is dark gray if over 1%, light gray if 0.1–1%, and white if less than 0.1%.

CLONAL NON-REDUNDANCY REMOVES NOISE







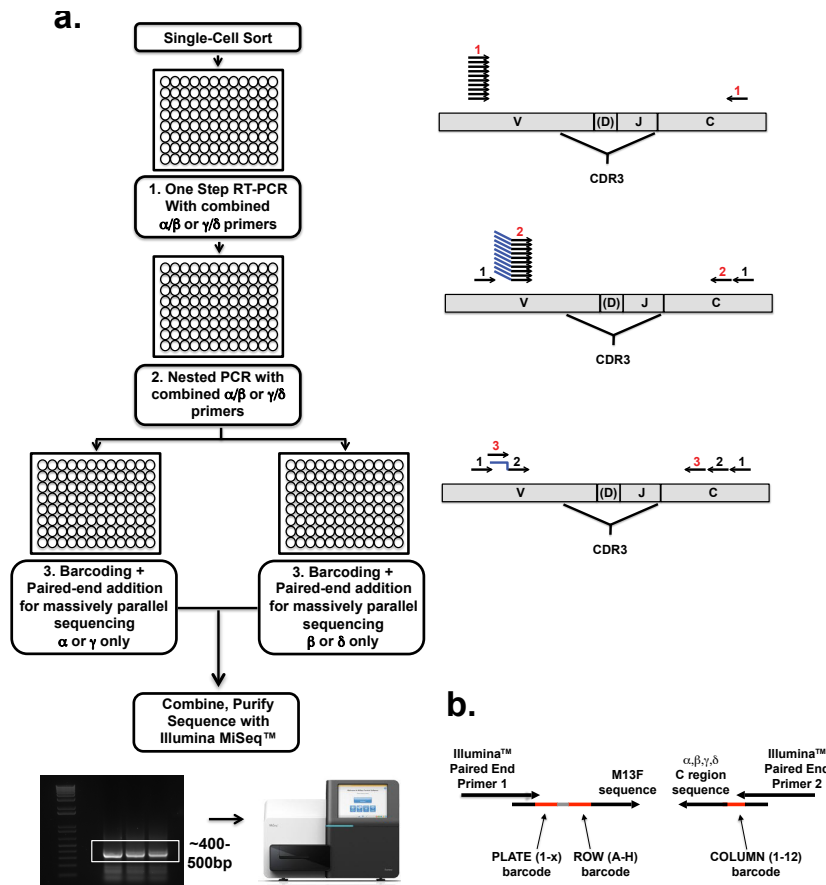
B cell exchange across the blood-brain barrier in multiple sclerosis

H.-Christian von Büdingen,¹ Tracy C. Kuo,² Marina Sirota,² Christopher J. van Belle,¹ Leonard Apeltsin,¹ Jacob Glanville,³ Bruce A. Cree,¹ Pierre-Antoine Gourraud,¹ Amy Schwartzburg,¹ Gabriella Huerta,² Dilduz Telman,² Purnima D. Sundar,² Tyler Casey,¹ David R. Cox,² and Stephen L. Hauser¹

Celiac

Single Cell TCR sequencing

Up to 50 96-well plates in one run



| | |
|----------|------------|
| 10/01/12 | 7 million |
| 12/19/12 | 7 million |
| 01/10/13 | 4 million |
| 03/13/13 | 4 million |
| 04/16/13 | 10 million |
| 06/04/13 | 7 million |

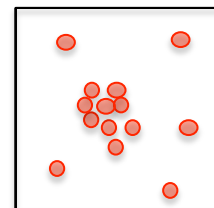
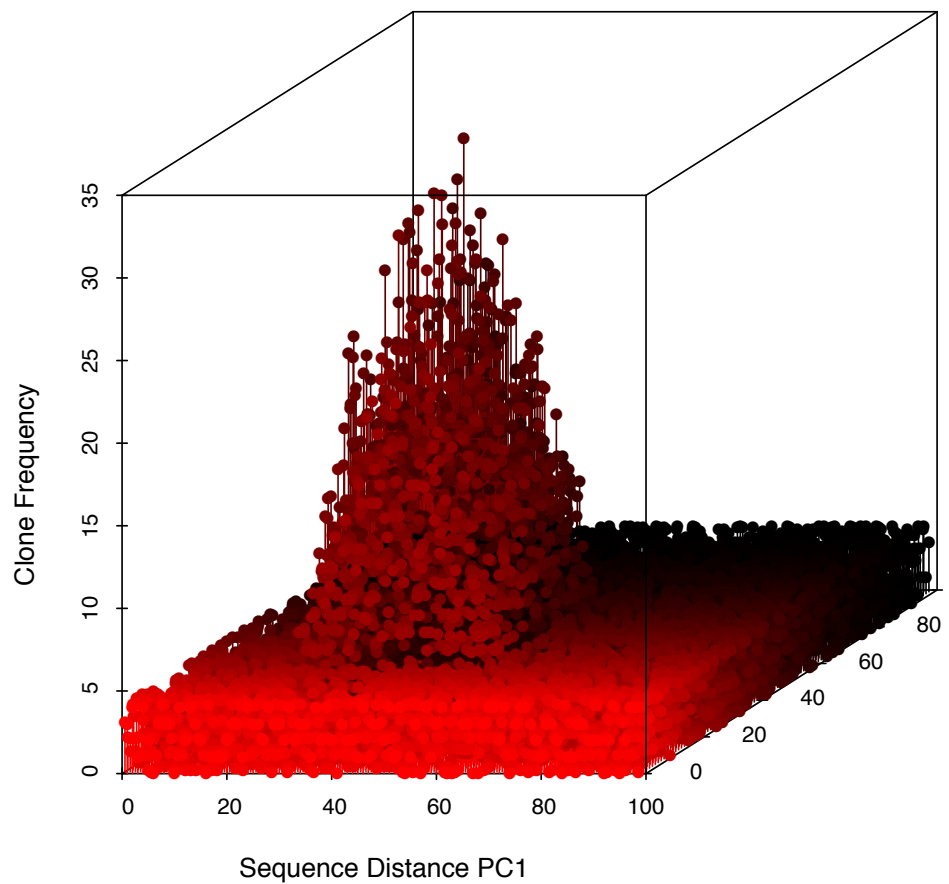


- ① Custom demultiplexing solution
- ② VDJFasta algorithm adaptation
- ③ Amazon cloud distribution
- ④ Single cell analysis

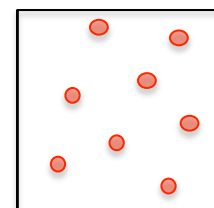
Dietary gluten triggers concomitant activation of CD4⁺ and CD8⁺ $\alpha\beta$ T cells and $\gamma\delta$ T cells in celiac disease

Arnold Han^{a,b}, Evan W. Newell^{b,c}, Jacob Glanville^d, Nielsen Fernandez-Becker^a, Chaitan Khosla^{e,f}, Yueh-hsiu Chien^{b,d}, and Mark M. Davis^{b,d,g,1}

TCR similarity landscape

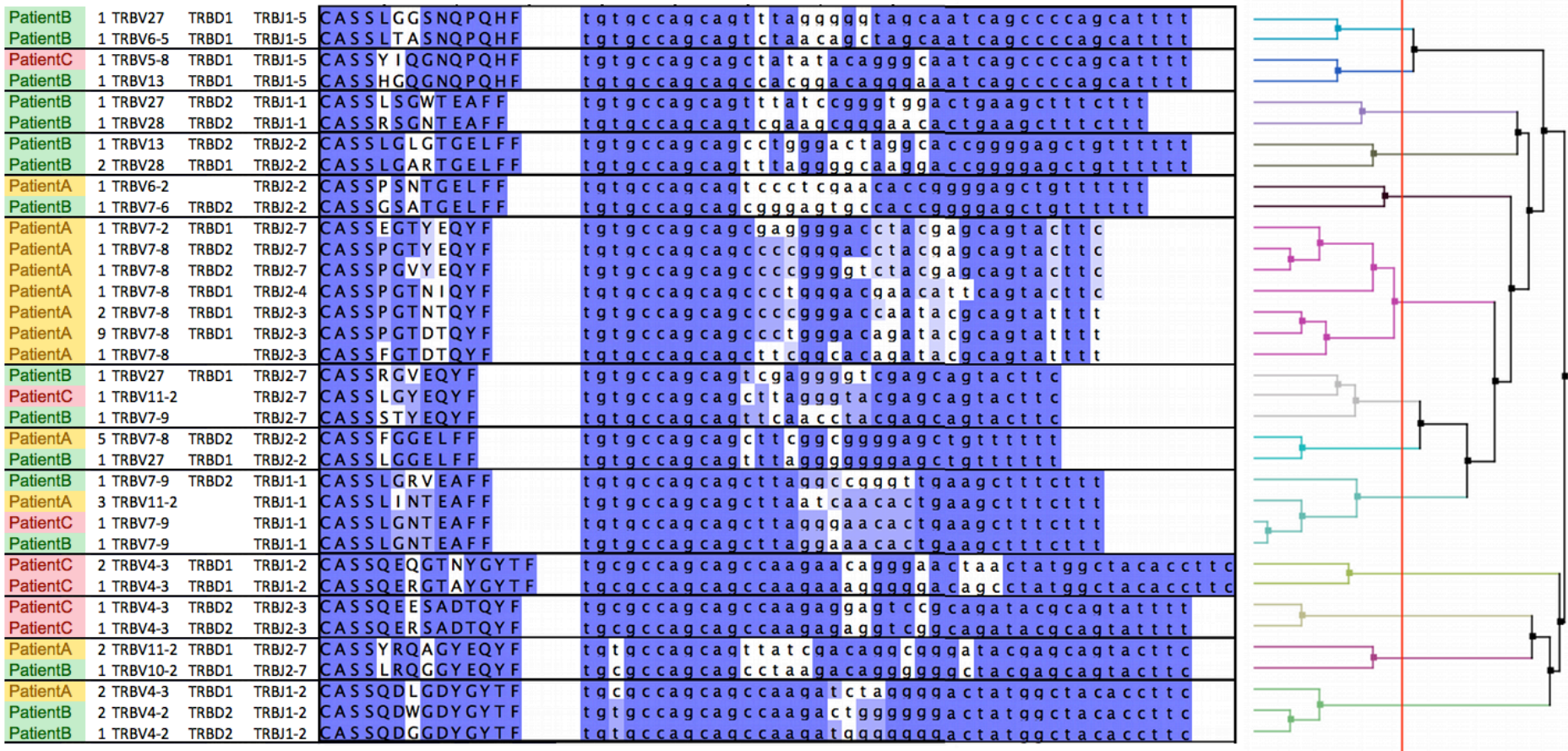


selected



unselected

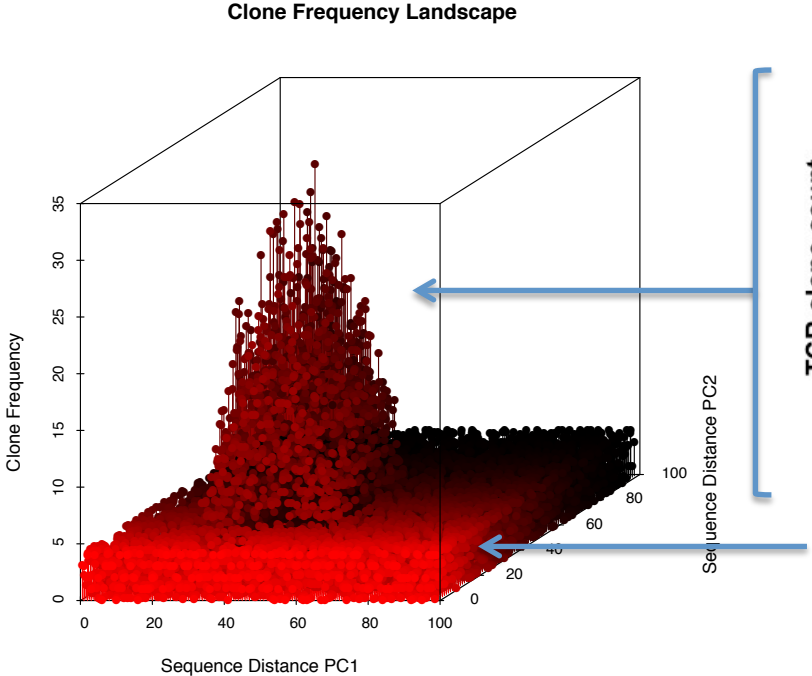
Are these clones more similar to each other than a random handful of T-cells?



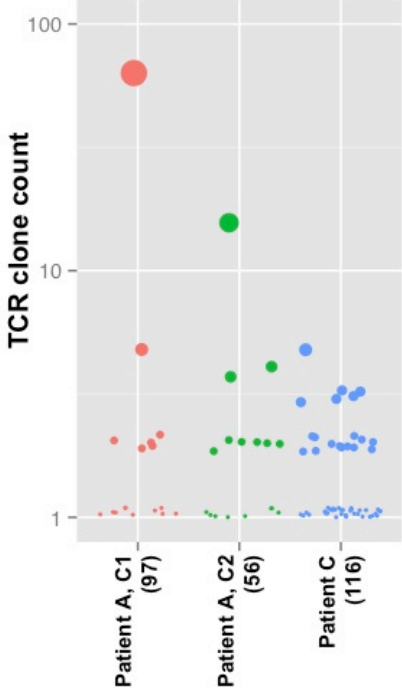
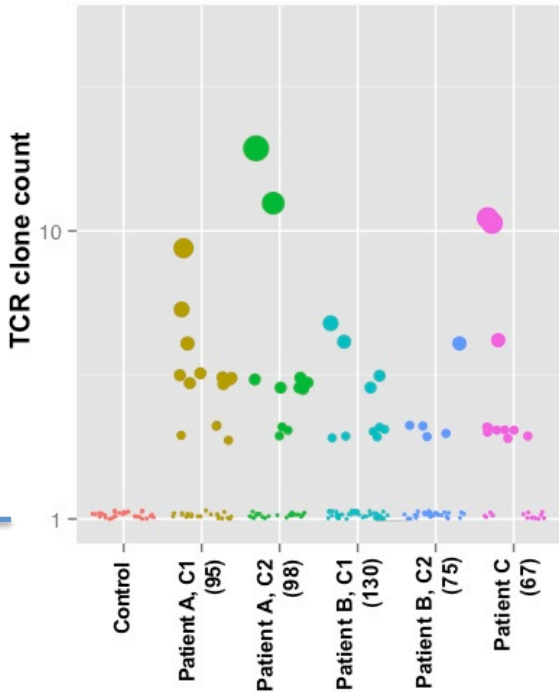
Dietary gluten triggers concomitant activation of CD4⁺ and CD8⁺ αβ T cells and γδ T cells in celiac disease

Arnold Han^{a,b}, Evan W. Newell^{b,c}, Jacob Glanville^d, Nielsen Fernandez-Becker^a, Chaitan Khosla^{e,f}, Yueh-hsiu Chien^{b,d}, and Mark M. Davis^{b,d,g,1}

There is TCR-driven clonal enrichment in the celiac CD8 samples



Simulated example landscape



Dispersion and convergence

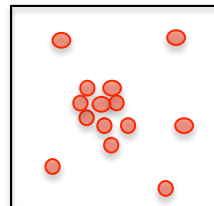
For selected population

For each paratope in set find distance to next closest sequence

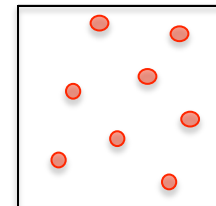
For unselected population

For each paratope in set find distance to next closest sequence

Repeat 10,000 times

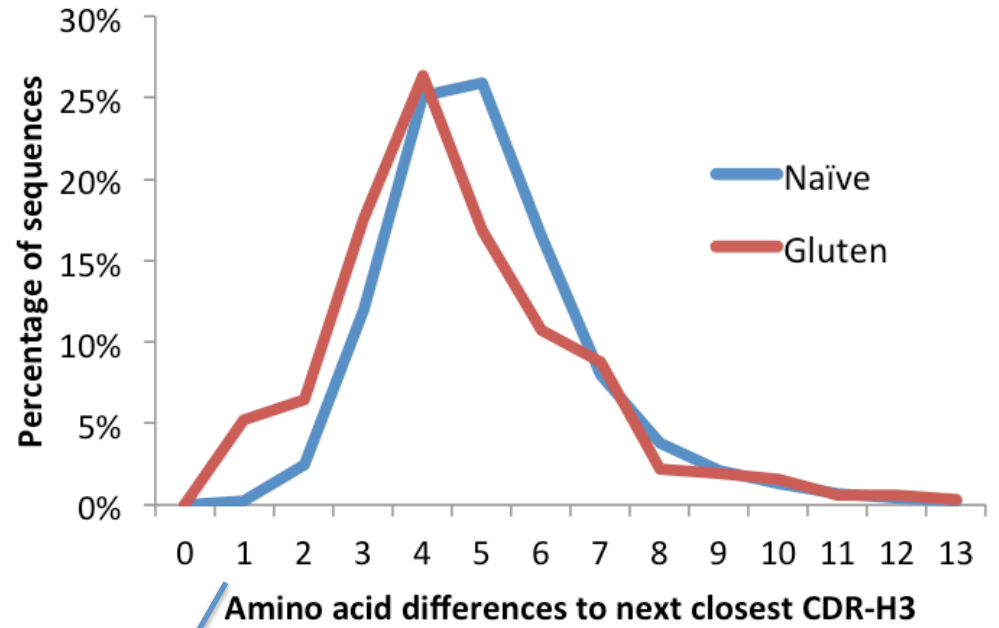


conversion

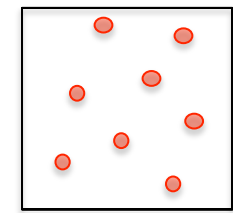
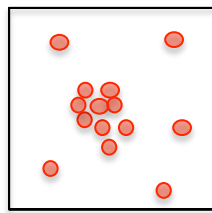


dispersion

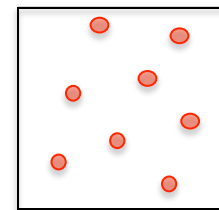
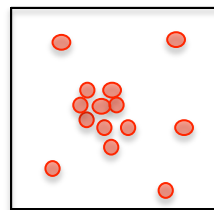
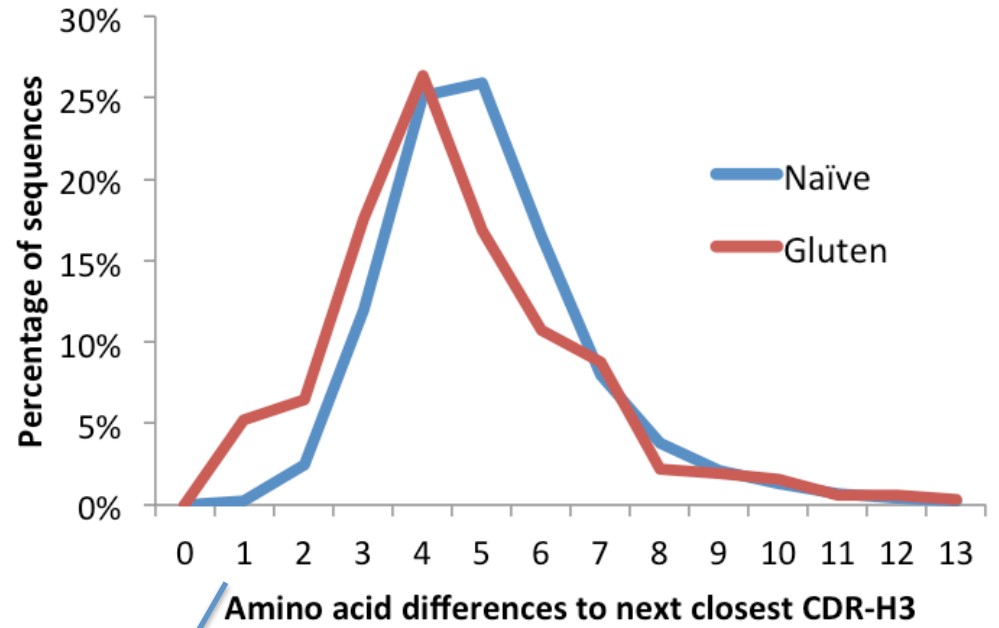
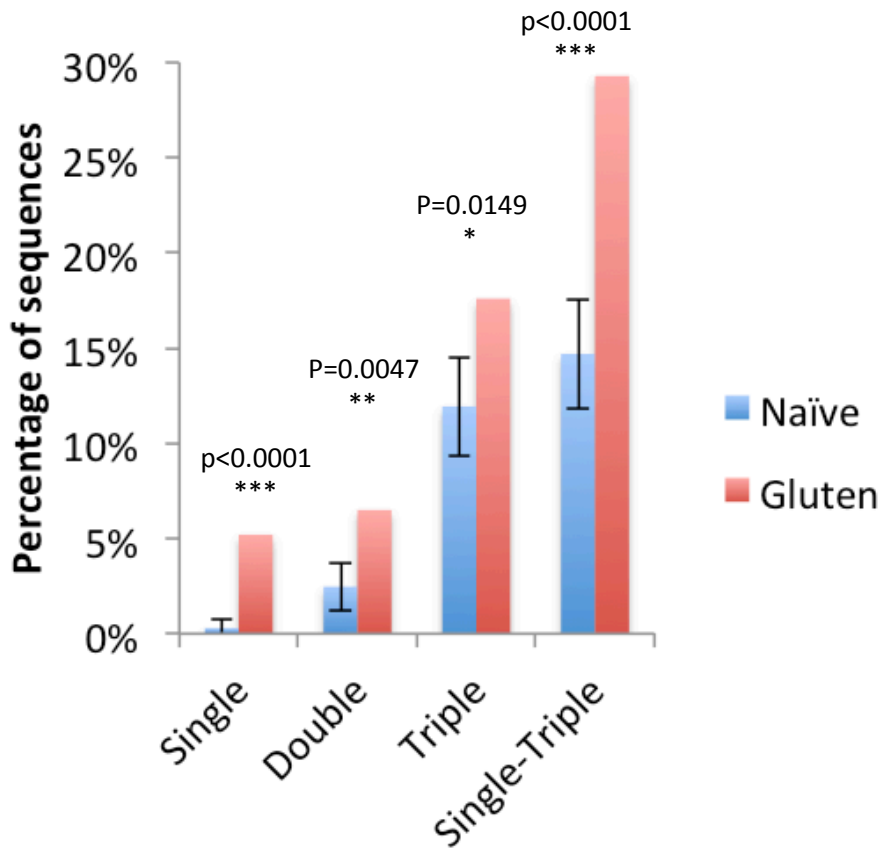
Dispersion and convergence



Amino acid differences to next closest CDR-H3



Dispersion and convergence



1. Can we detect convergent specificities before antigen is known?

Syndrome X

Healthy controls



Vaccines



unprotected



unprotected



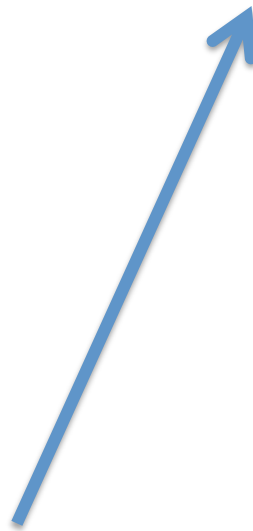
vaccinated



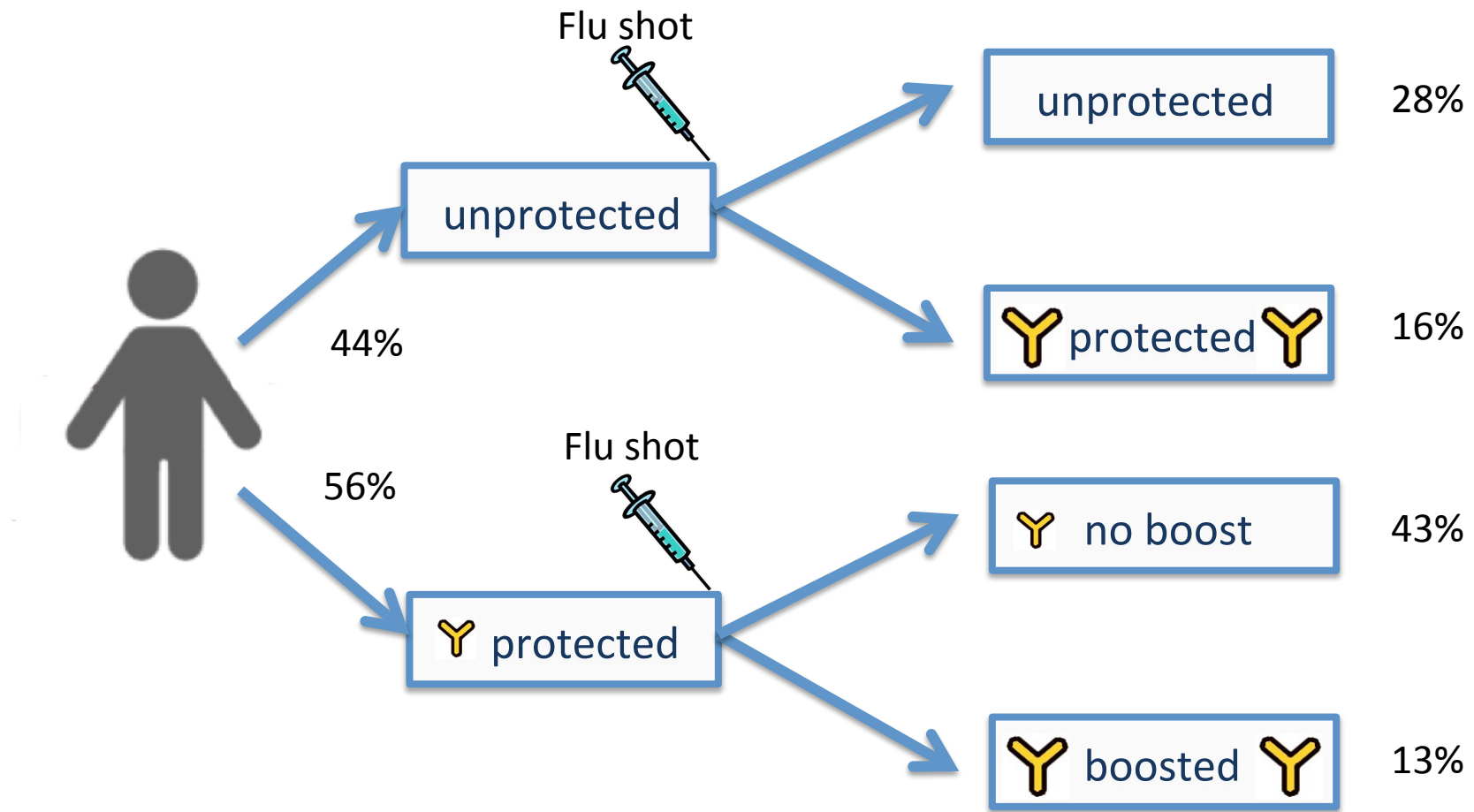
unprotected



vaccinated

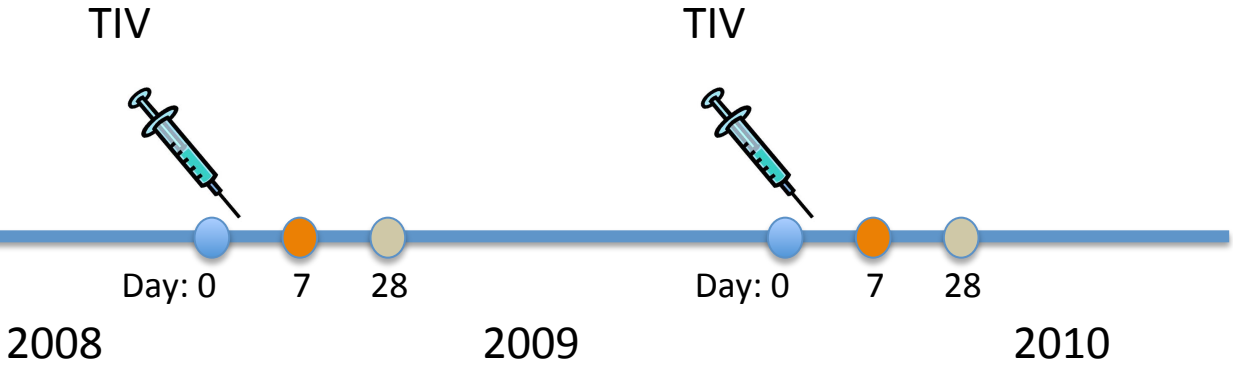
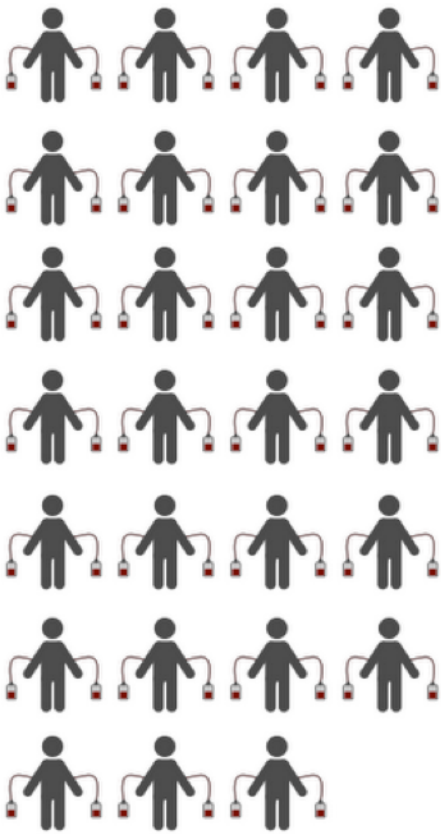


Y protected Y



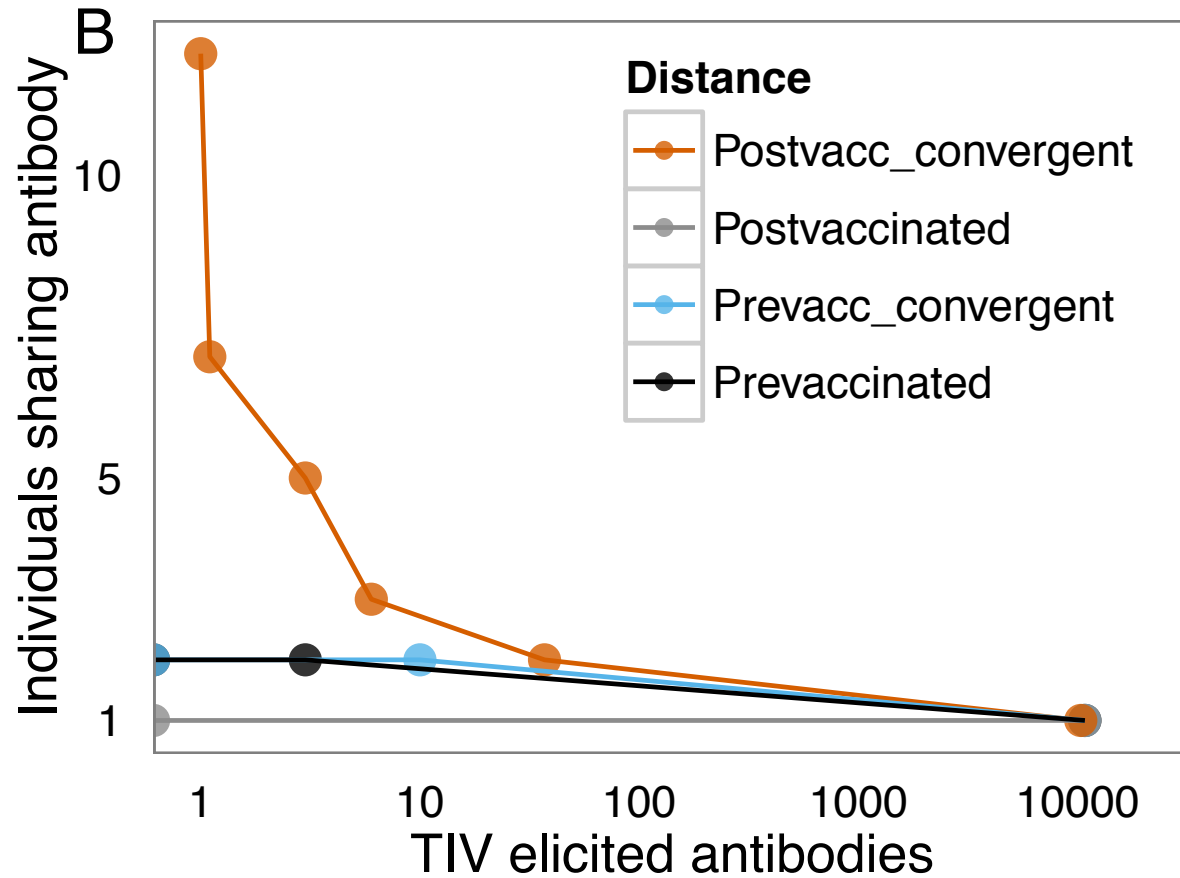
Percentages obtained from 162 assays, 2008-2009 and 2009-2010 flu season

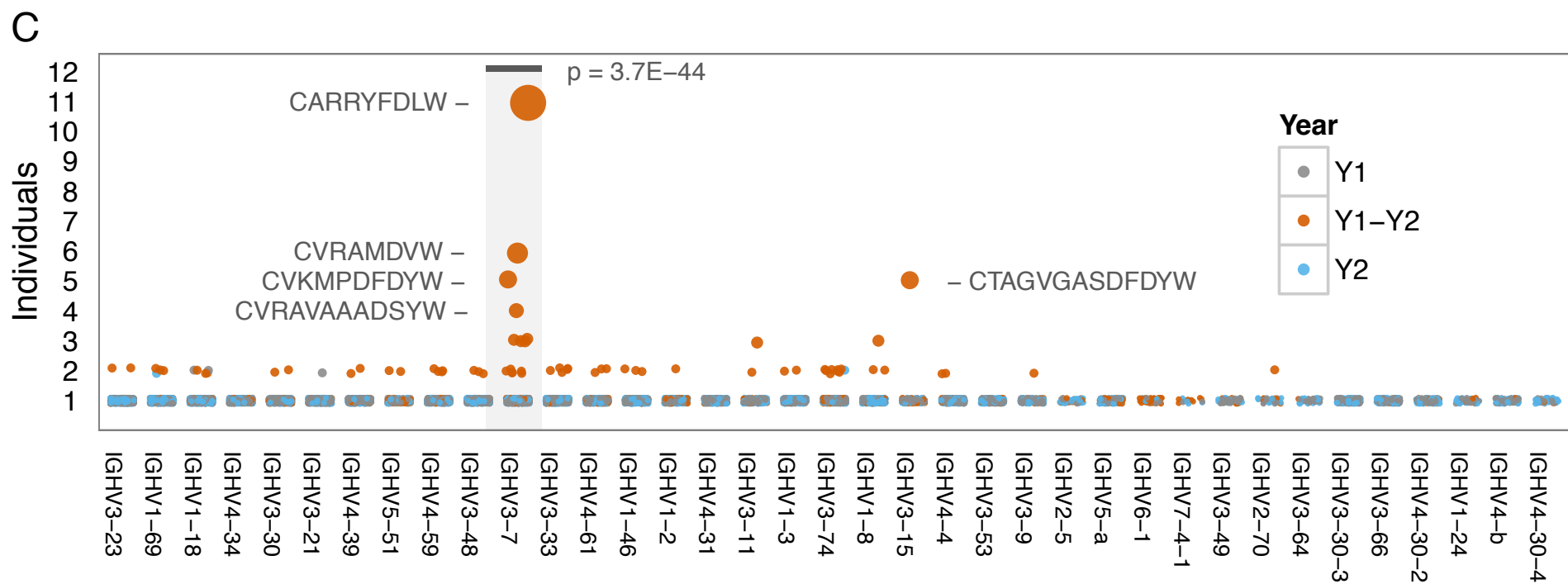
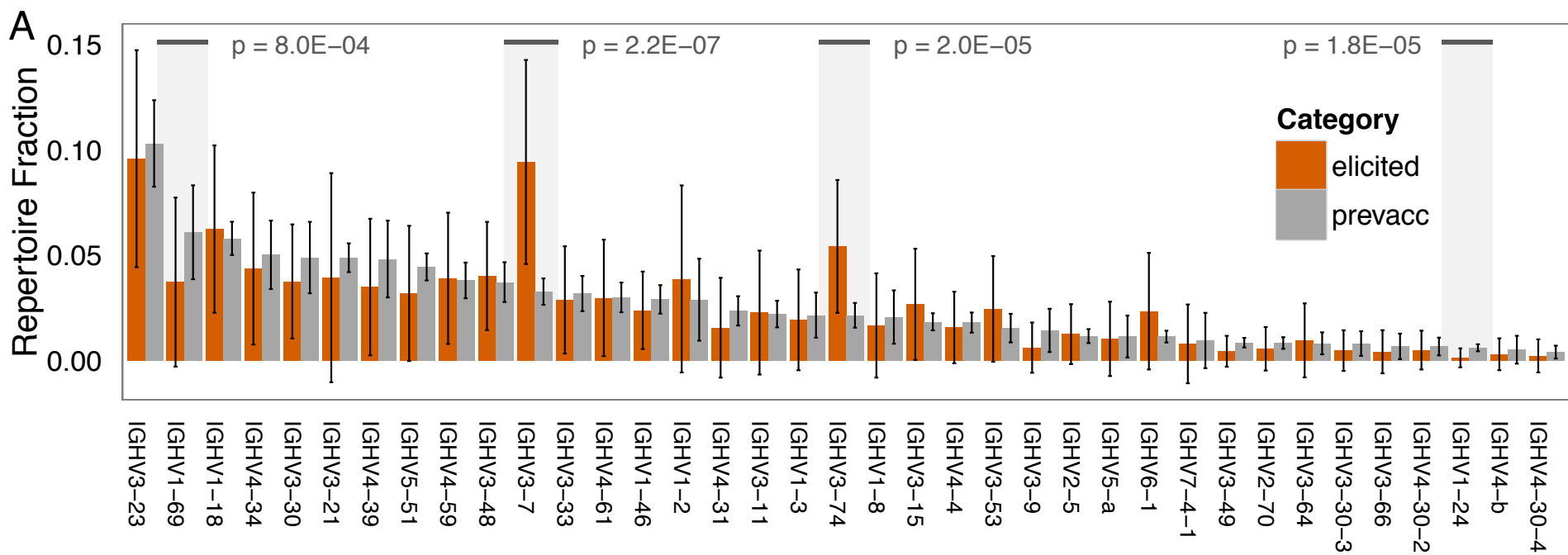
27 subjects



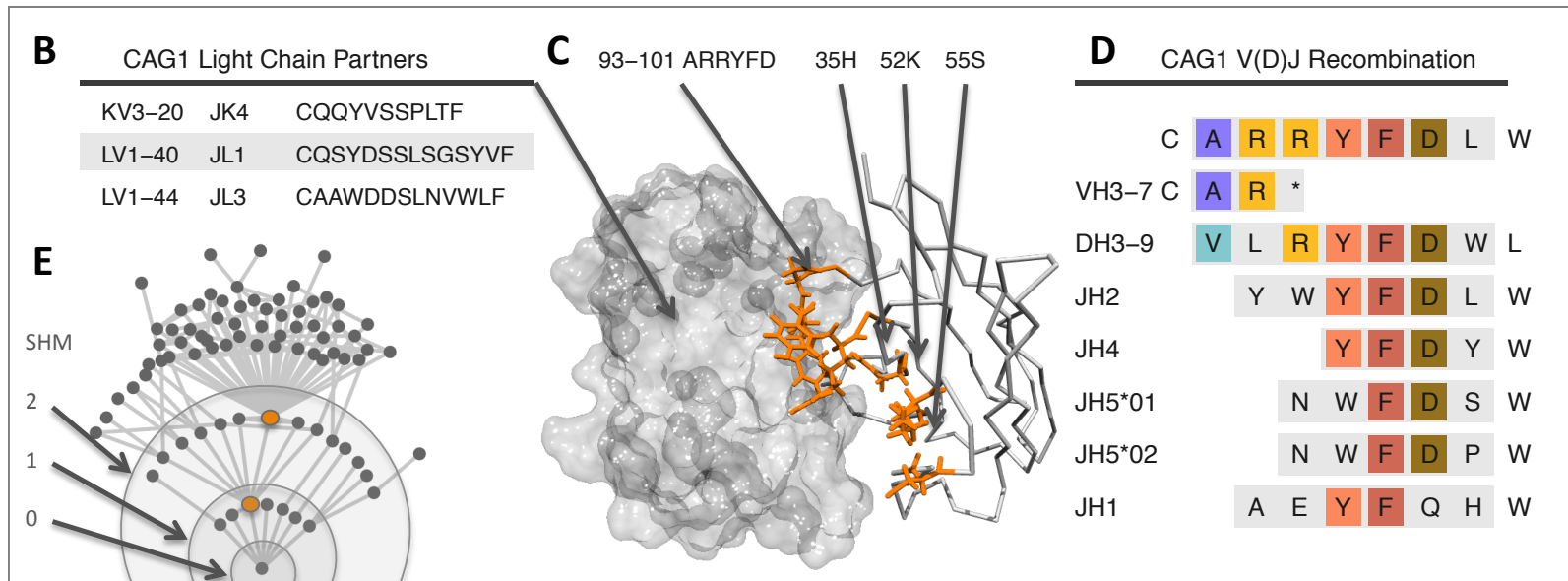
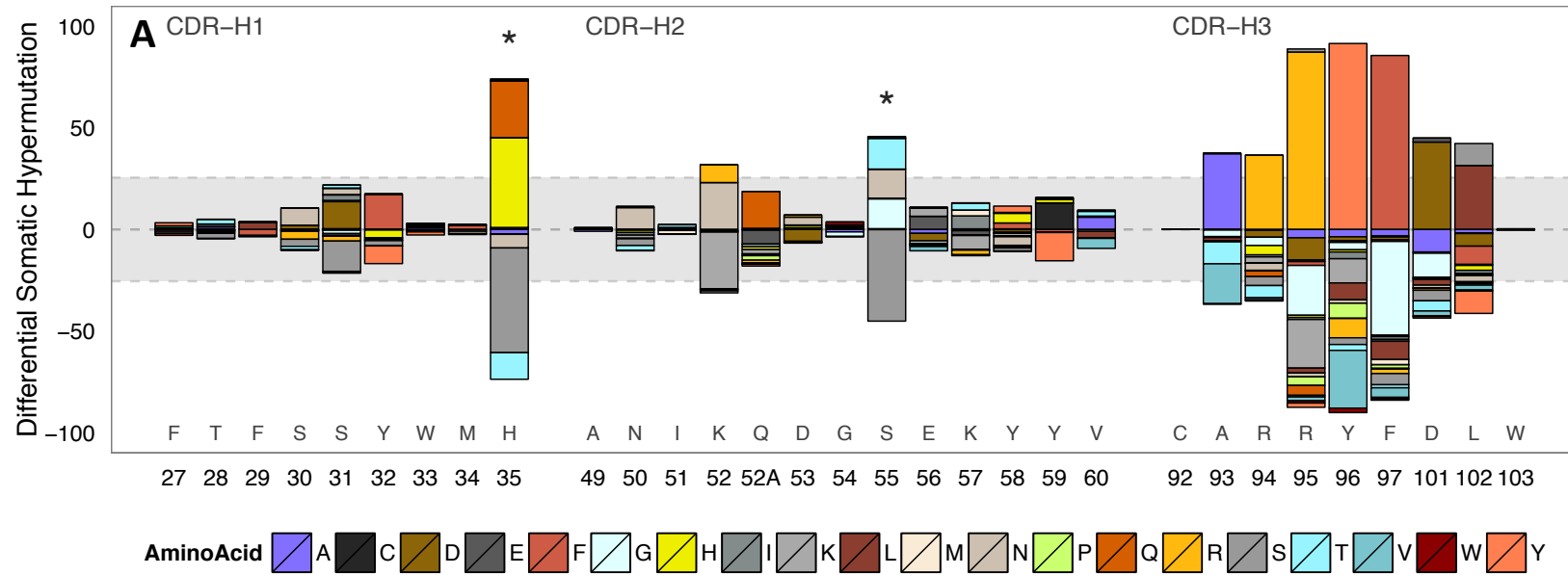
- 27 people vaccinated on 2 consecutive years
- Fv cDNA and gDNA sequenced at day 0, 7, 28
- 50k reads/subject

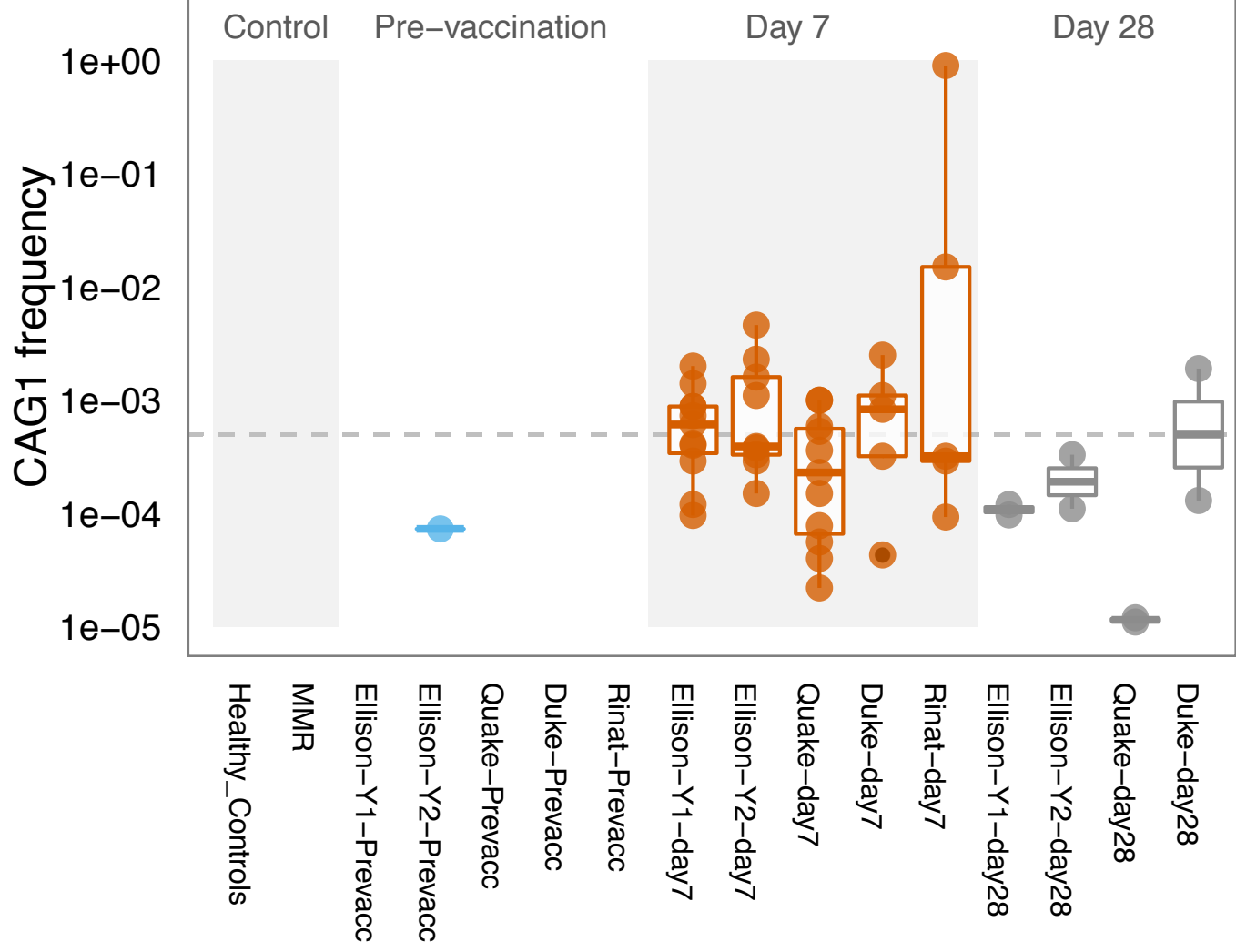
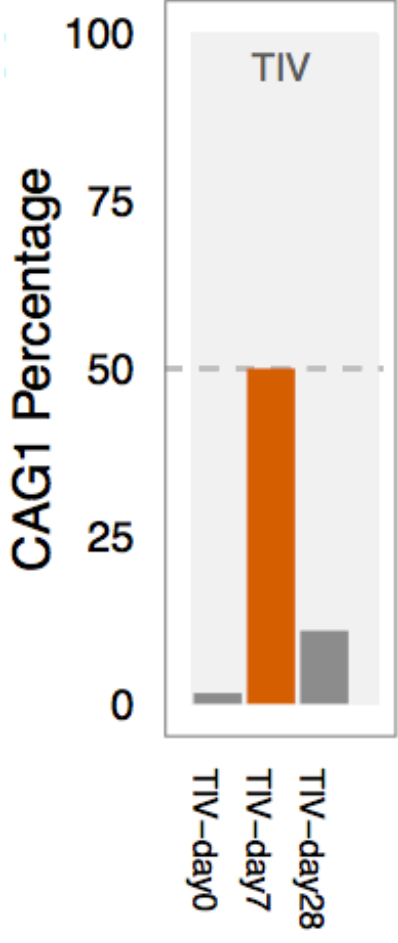
Convergent antibody groups are elicited after TIV vaccination

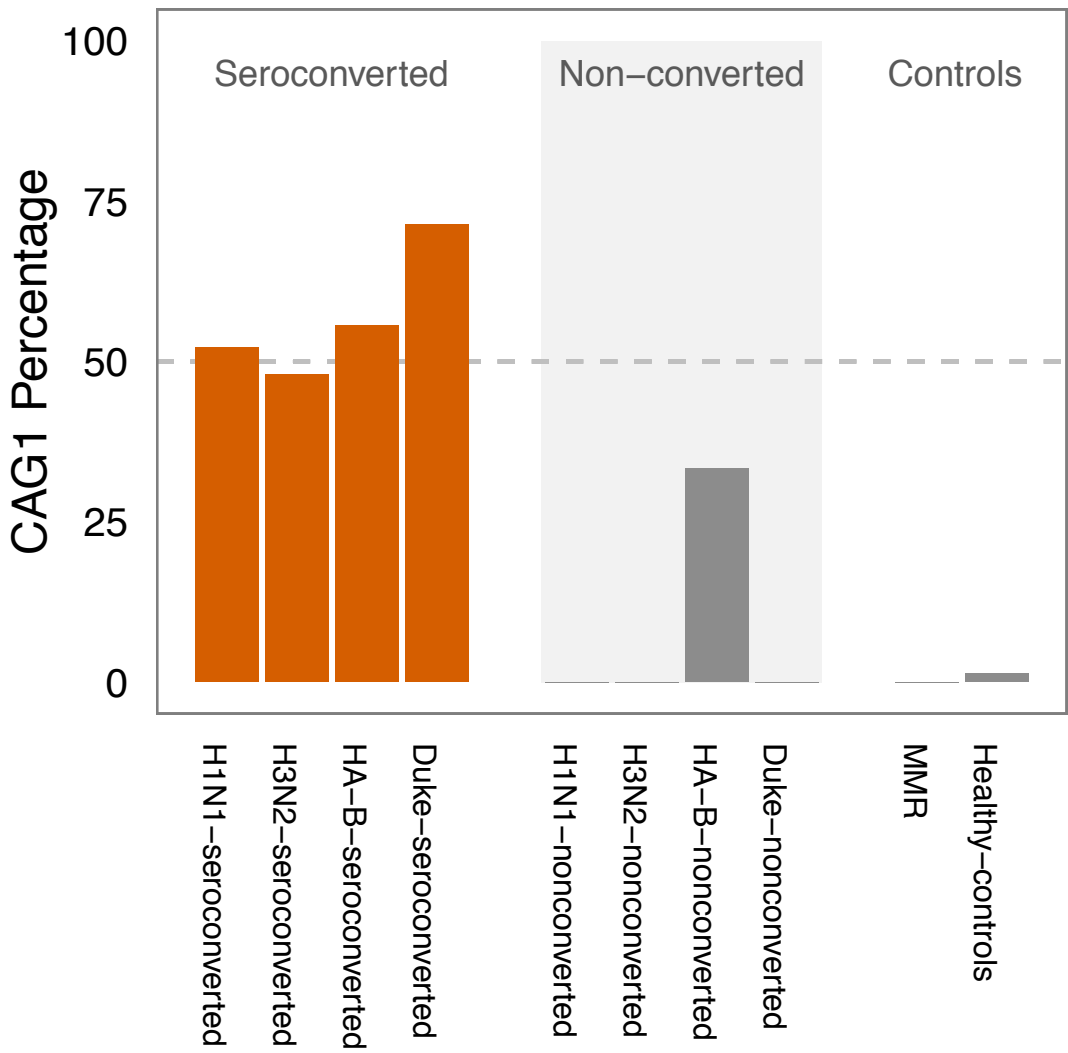




CAG1 – convergent somatic hypermutation and V(D)J rearrangements

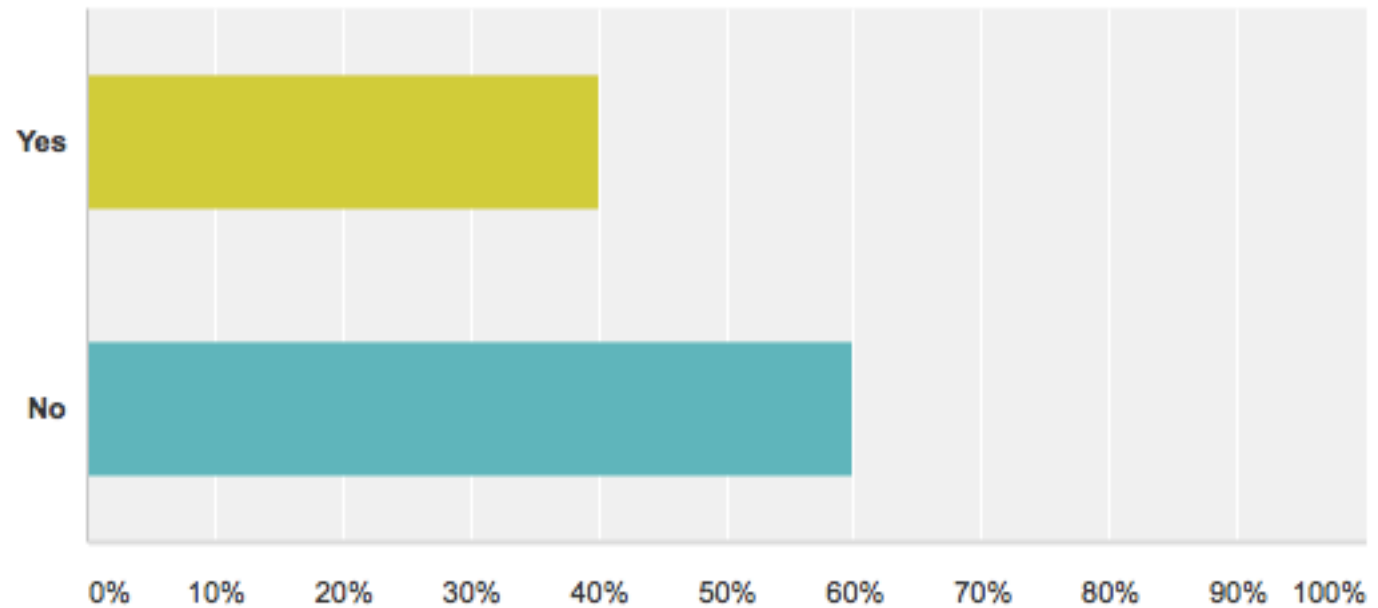






Do you already have a dataset that you would like to work on for this course?

Answered: 10 Skipped: 0



| Answer Choices | Responses |
|----------------|-----------|
| Yes | 40.00% 4 |
| No | 60.00% 6 |
| Total | 10 |

Repertoire data stores

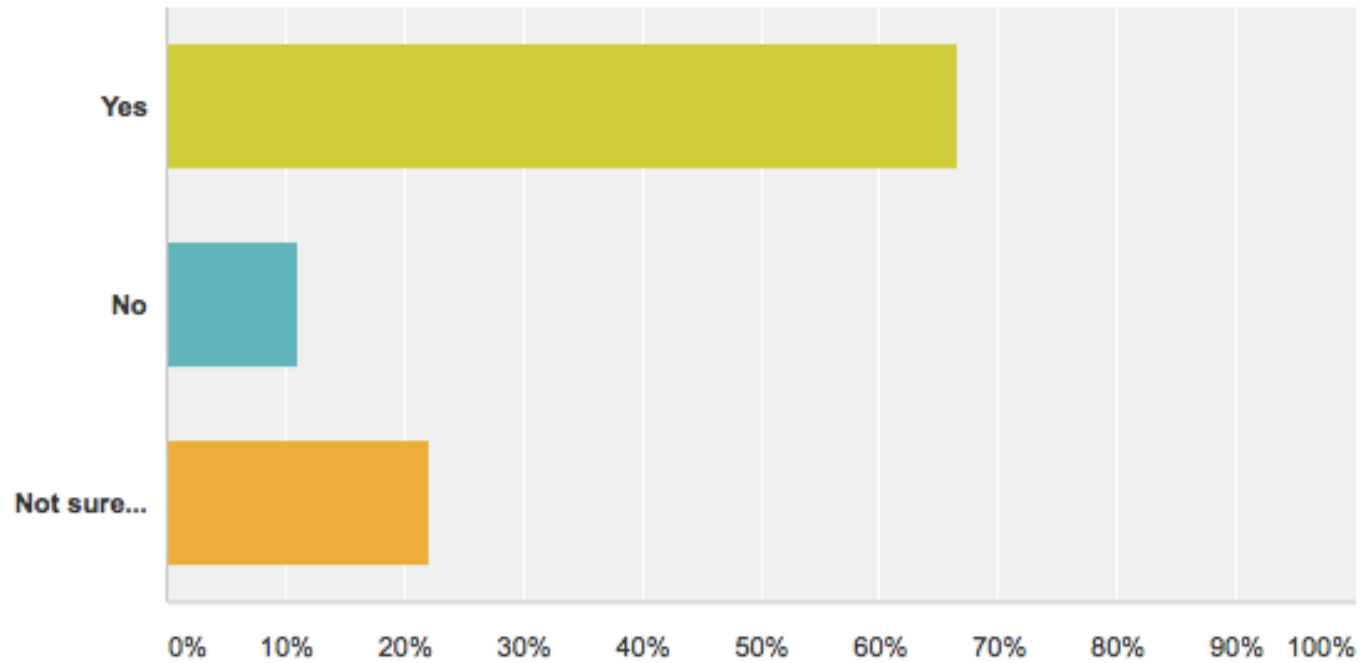
NCBI Trace Archives

Many labs on campus

Simulated data

Would you be interested in being partnered with a mentor?

Answered: 9 Skipped: 1



| Answer Choices | Responses |
|----------------|-----------|
| Yes | 66.67% 6 |
| No | 11.11% 1 |
| Not sure... | 22.22% 2 |
| Total | 9 |

Repertoire labs on campus

Boyd

Davis

Quake

(many others)

Coding

Poll results

Crowdsourcing assignments

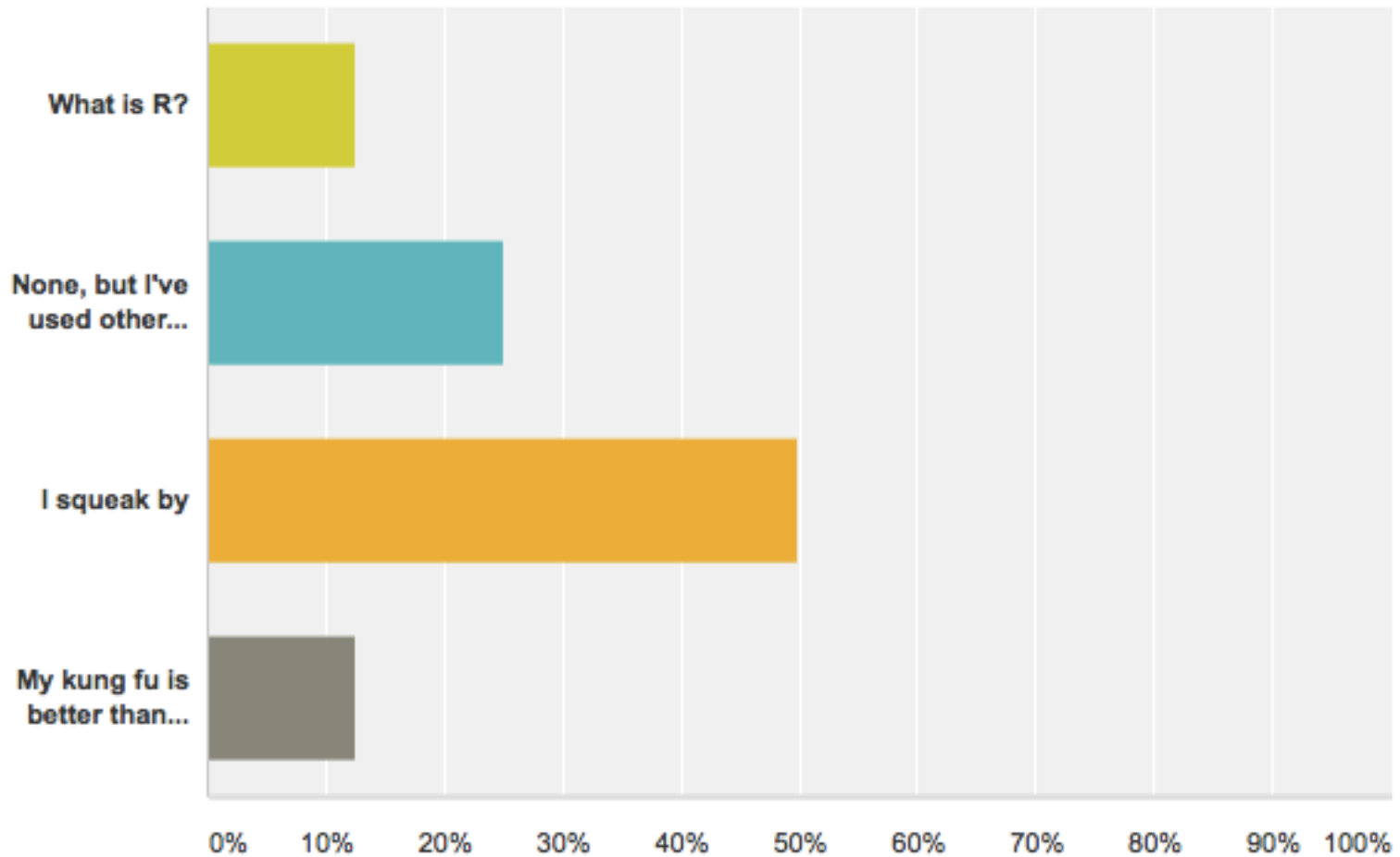
Crowdsourcing assignment #1 – figures

Example assignment #1 template & code walk-through

Immunocode

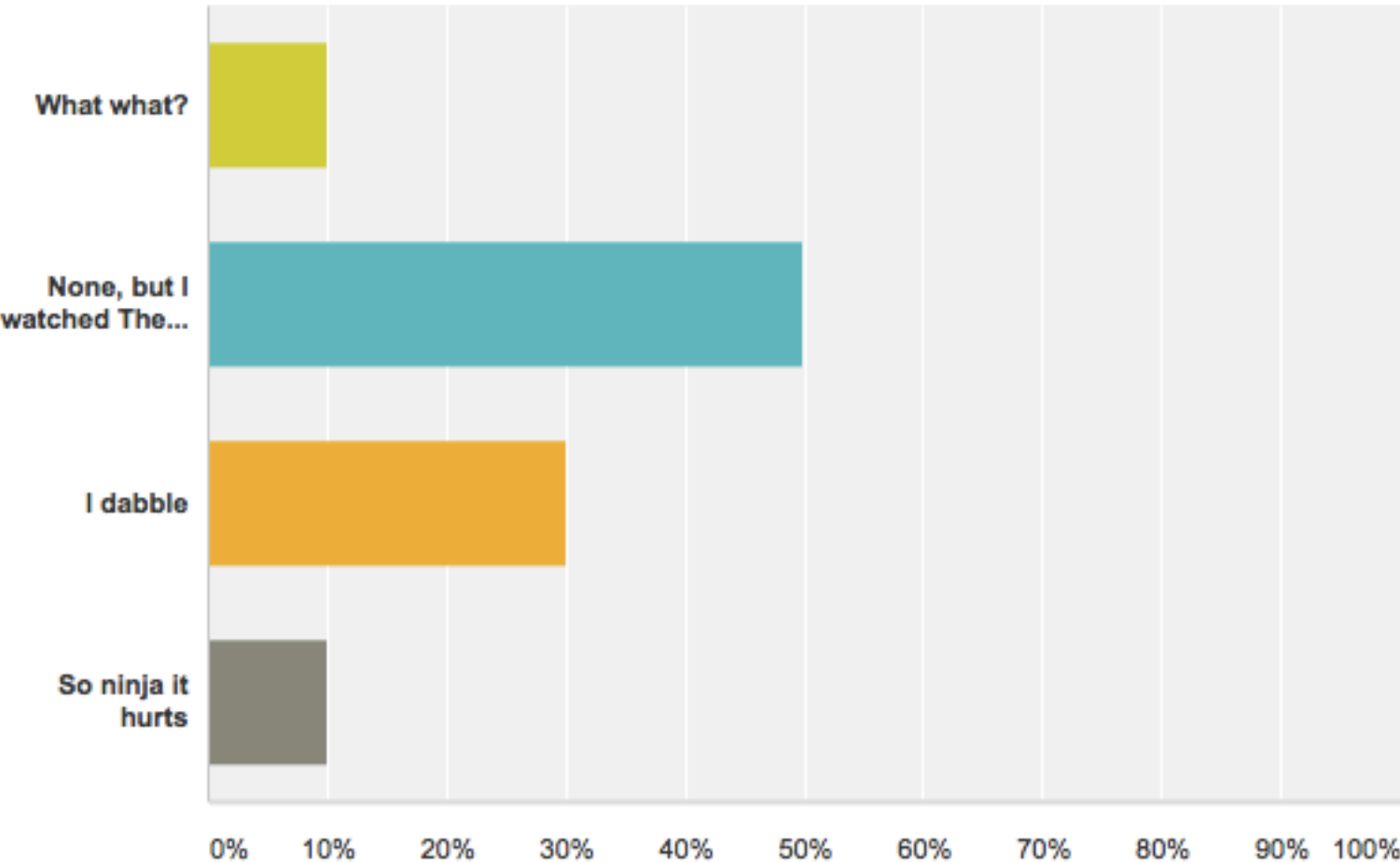
How much experience have you had with R?

Answered: 8 Skipped: 2



How much experience do you have at the command line?

Answered: 10 Skipped: 0





SWIRL: Statistics With InteRactive Learning

by JACOB GLANVILLE on OCTOBER 31, 2013 · [LEAVE A COMMENT](#)

“The swirl R package is designed to simultaneously teach users statistics and the R programming language. If you are new to the R programming language, have no fear. We will walk you through each of the steps required to begin using the swirl instructional platform today!” – [SWIRL](#)

[Continue Reading →](#)

Immunol 310: Mass Cytometry

by JACOB GLANVILLE on OCTOBER 30, 2013 · [LEAVE A COMMENT](#)

“Flow cytometry bioinformatics is the application of bioinformatics to flow cytometry data, which involves storing, retrieving, organizing and analyzing flow cytometry data using extensive computational resources and tools. Flow cytometry bioinformatics requires extensive use of and contributes to the development of techniques from computational statistics and machine learning. Flow cytometry and related methods allow the [...]

[Continue Reading →](#)

Immunol 310: expression analysis

[Upload New Code](#)

CATEGORIES

- [Immuno 310](#)
- [Perl](#)
- [R](#)
- [Unix](#)
- [web-based](#)

POSTS

| APRIL 2014 | | | | | | |
|-----------------------|----|----|----|----|----|----|
| M | T | W | T | F | S | S |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 21 | 22 | 23 | 24 | 25 | 26 | 27 |
| 28 | 29 | 30 | | | | |
| « Oct | | | | | | |

RECENT POSTS

[SWIRL: Statistics With InteRactive Learning](#)

Crowdsourcing assignments

Wednesday: assignments handed out

All assignments are of a similar theme

Each person gets a slightly different assignment

Collaboration is fine

Following Wednesday: assignments reviewed

Submit working code the night before

Provide code review

Code should work at command-line

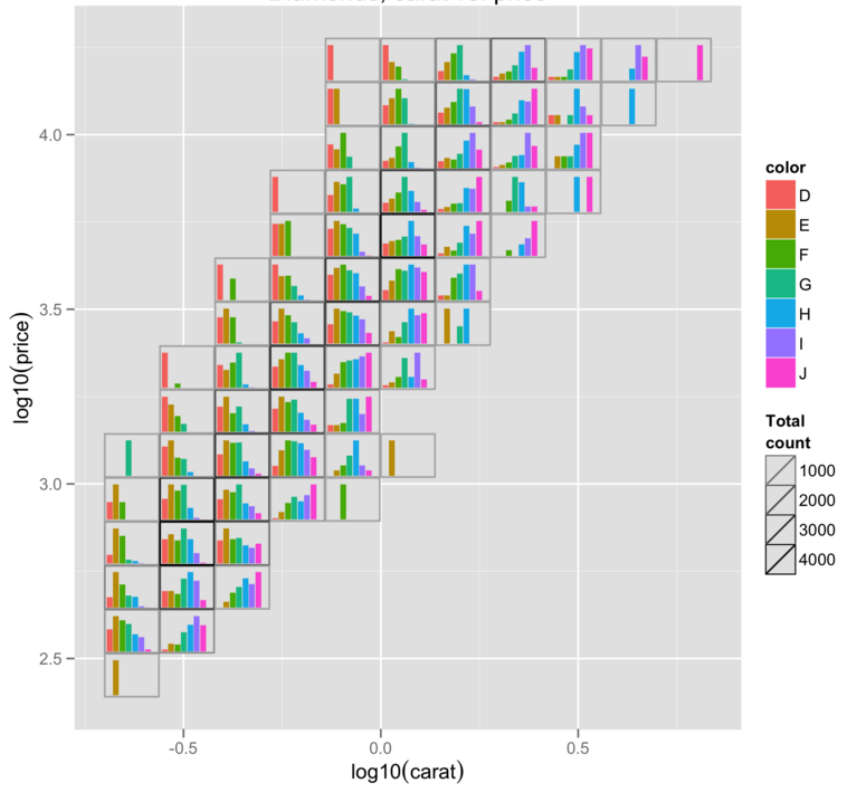
Code should be useful

All code is scored and shared on immunocode

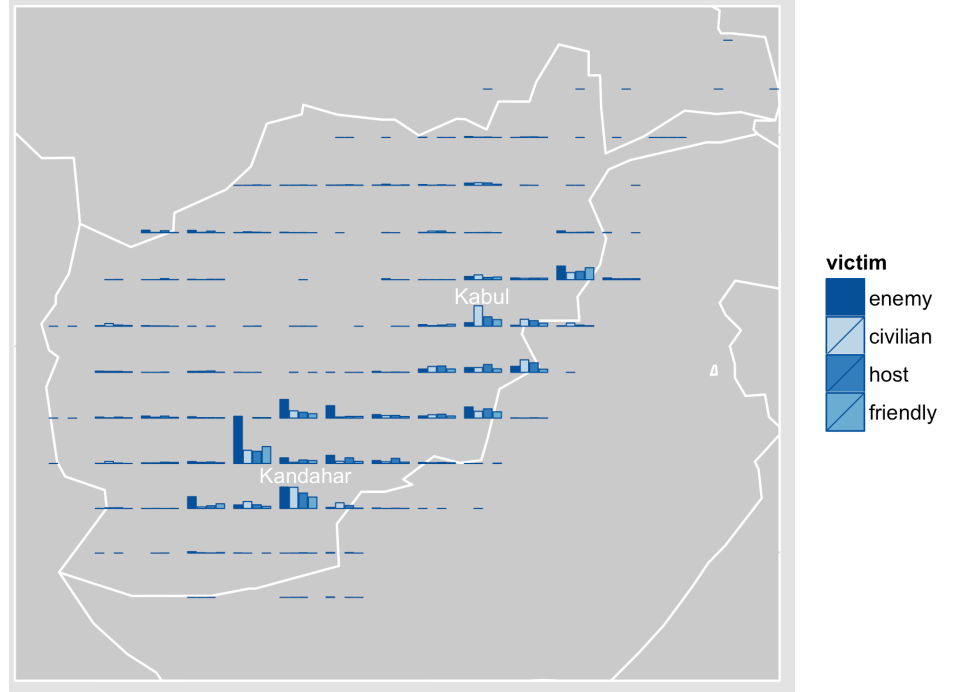
Crowdsourcing assignments

- ① figures with ggplot2
- ② data reformatting with plyr
- ③ statistics
- ④ clustering
- ⑤ pattern recognition algorithms
- ⑥ custom CSI

Diamonds, carat vs. price



Afghan War Diary - Casualty counts by type and location



Geoms

Geoms, short for geometric objects, describe the type of plot you will produce.

- `geom_abline`
Line specified by slope and intercept.
- `geom_area`
Area plot.
- `geom_bar`
Bars, rectangles with bases on x-axis
- `geom_bin2d`
Add heatmap of 2d bin counts.
- `geom_blank`
Blank, draws nothing.
- `geom_boxplot`
Box and whiskers plot.
- `geom_contour`
Display contours of a 3d surface in 2d.
- `geom_crossbar`
Hollow bar with middle indicated by horizontal line.
- `geom_density`
Display a smooth density estimate.
- `geom_pointrange`
An interval represented by a vertical line, with a point in the middle.
- `geom_polygon`
Polygon, a filled path.
- `geom_quantile`
Add quantile lines from a quantile regression.
- `geom_raster`
High-performance rectangular tiling.
- `geom_rect`
2d rectangles.
- `geom_ribbon`
Ribbons, y range with continuous x values.
- `geom_rug`
Marginal rug plots.
- `geom_segment`
Single line segments.



- `geom_dotplot`
Dot plot
- `geom_errorbar`
Error bars.
- `geom_errorbarh`
Horizontal error bars
- `geom_freqpoly`
Frequency polygon.
- `geom_hex`
Hexagon binning.
- `geom_histogram`
Histogram
- `geom_hline`
Horizontal line.
- `geom_jitter`
Points, jittered to reduce overplotting.
- `geom_line`
Connect observations, ordered by x value.
- `geom_linerange`
An interval represented by a vertical line.
- `geom_segment`
Single line segments.
- `geom_smooth`
Add a smoothed conditional mean.
- `geom_step`
Connect observations by stairs.
- `geom_text`
Textual annotations.
- `geom_tile`
Tile plane with rectangles.
- `geom_violin`
Violin plot.
- `geom_vline`
Line, vertical.



ggplot2: assignments

- | | |
|---------------|---------------------|
| 1. Area | 8. Quantile |
| 2. Bar | 9. Dotplot |
| 3. Boxplot | 10. Histogram |
| 4. Contour | 11. Line |
| 5. Density | 12. Smooth |
| 6. jitter | 13. Violin |
| 7. Pointrange | 14+: something evil |

(Handed out by order in sign-in sheet)

<http://docs.ggplot2.org>

```
# Author: Jacob Glanville
# Contact: jakeg@stanford.edu

# The purpose of this code is to generate a scatterplot using ggplot.
# Inputs: a tab-delimited file of X and Y coordinates
# Outputs: a scatterplot

##### Libraries #####
library(ggplot2)
##### Arguments #####
cmd_args = commandArgs();
myfile = cmd_args[6]
##### Inputs #####
mytable=read.table(myfile, header=TRUE)
##### Outputs #####
output_pdf=paste("ggplot-scatter-",myfile, ".pdf", sep="")

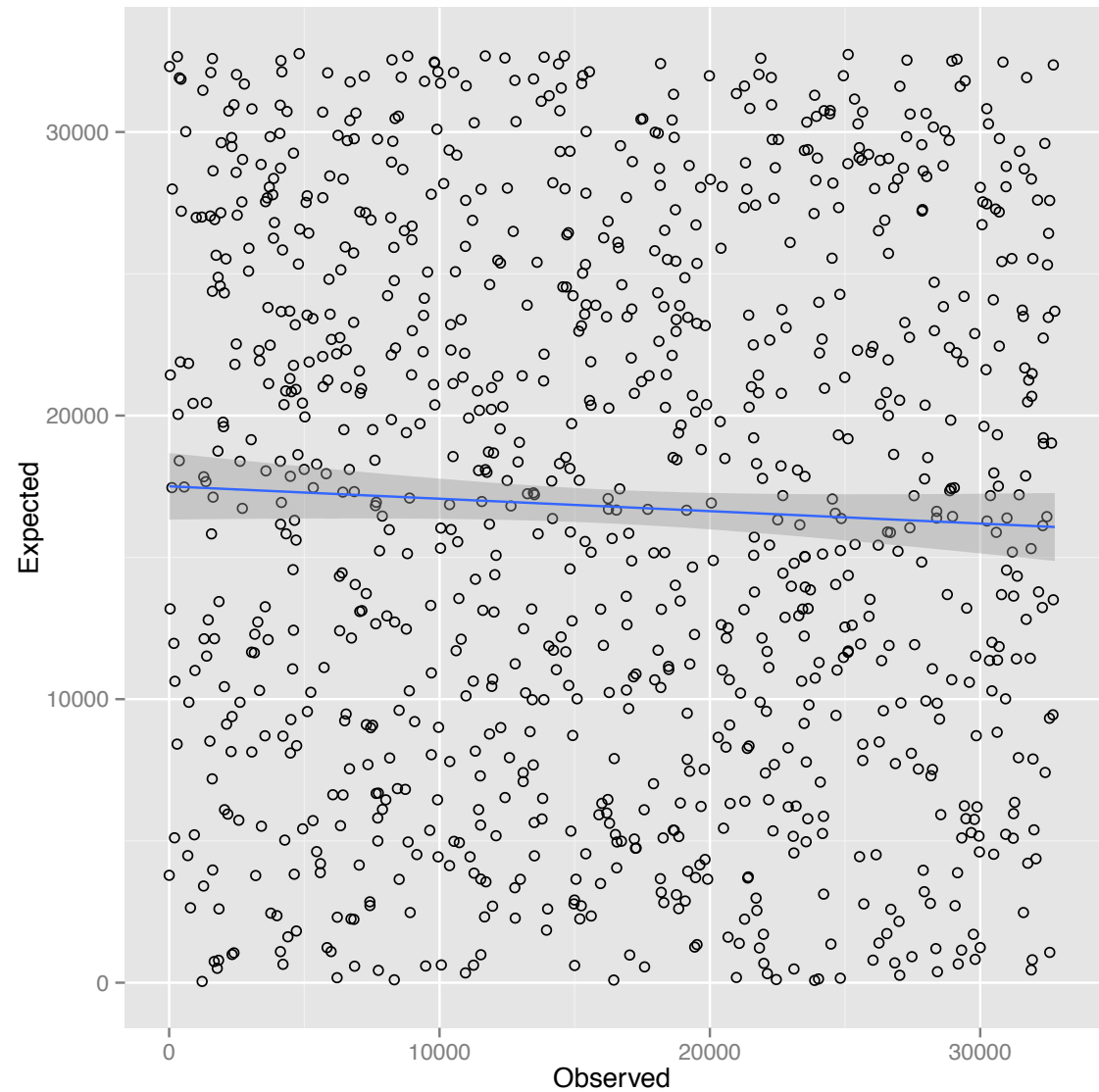
pdf(output_pdf, height=7, width=7)
ggplot(mytable, aes(x=Observed, y=Expected)) +
  geom_point(shape=1) + # Use hollow circles
  geom_smooth(method=lm) # Add linear regression line
                        # (by default includes 95% confidence region)

dev.off()
```

Input File (data.txt)

| Observed | Expected |
|----------|----------|
| 7295 | 9099 |
| 16564 | 4962 |
| 14214 | 11725 |
| 3654 | 23812 |
| 11525 | 3655 |
| 31338 | 11422 |
| 4503 | 9280 |
| 31882 | 15309 |
| 5696 | 21017 |
| 13958 | 1853 |
| 17977 | 29993 |
| 1379 | 20455 |
| 21883 | 32601 |
| 28851 | 29707 |
| 19470 | 3710 |
| 9410 | 23533 |
| 6671 | 7545 |
| 25477 | 30287 |
| 24706 | 11025 |
| 6211 | 177 |
| 31453 | 29323 |
| 31233 | 5963 |

Output File (deluxe-fig.pdf)



Command:

Rscript ggplot-scatter.R data.txt

Questions?