Toward high-throughput antibody-antigen modeling

Daron Standley, Osaka University

Brief Bio

- 1998: PhD in computational chemistry, Columbia Univ.
- 1998-2003: Software developer, Schrodinger Inc.
- 2003-2008 : Researcher, Osaka Univ., Inst. for Protein research
- 2008-2014: Associate Prof., Immunology Frontier Research Center
- 2014-2016 : Professor, Kyoto Univ., Inst. for Virus Research
- 2016: Co-founder, KOTAI Biotechnologies
- 2016-present: Professor, Osaka Univ., Research Inst. for Microbial Diseases

Part 1: A personal story

How our lab became interested in antibodies

The lab

About helf of the lab is working on antibod


Daron


Kazutaka


Diego


Shunsuke


Songling


Floris


John


Mara


Arthur


Jan


Ana

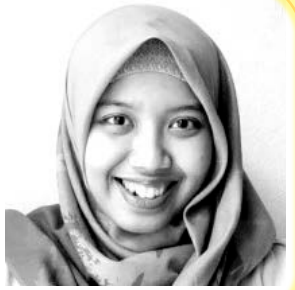

Xu


Aybars


Martin


Xie


Hendra


Nita

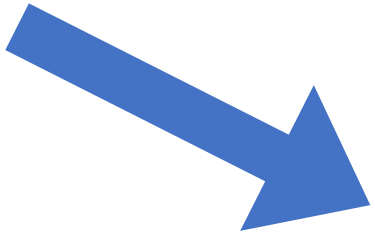

Masako

In 2013 my lab was invited to join a contest

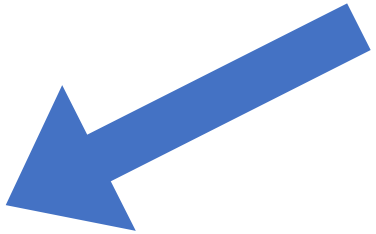
The second antibody modeling assessment:
"To assess the state of the art in antibody 3D modeling"

Organization of the Contest

Organizers send sequences of antibodies to participants



Participants prepare 3D models and send back to organizers



Organizers compare the 3D models to x-ray crystal structures and assess errors

The participants

- Accerlys Inc
- Chemical Computer Group (CCG)
- Schrodinger, Inc.
- Jeff Gray's lab (John Hopkins University)
- Macromoltek
- Astellas Pharma/Osaka University
- Prediction of ImmunoGlobulin Structure (PIGS).

The organizers

Big Pharma

Pfizer

Johnson & Johnson

Academics

Scripps Research

Our team

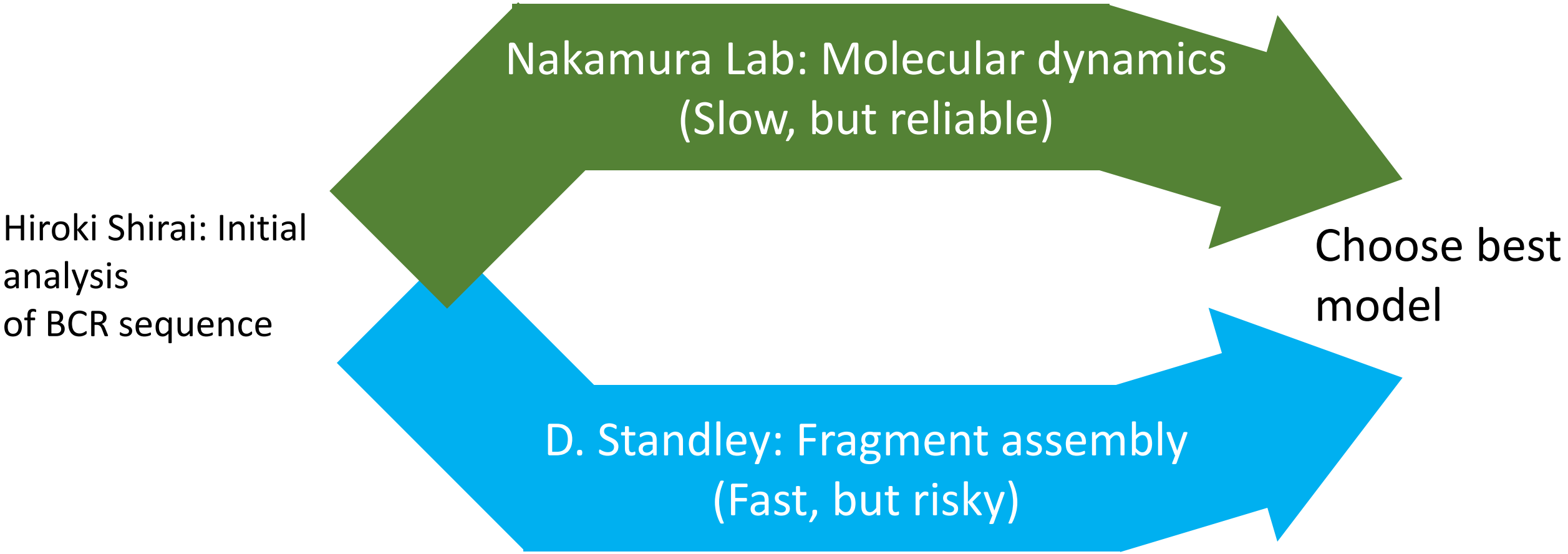
Haruki Nakamura
Inst. Protein Research
Osaka Univ.

Hiroki Shirai
Astellas Pharma

D. Standley
IFReC
Osaka Univ.

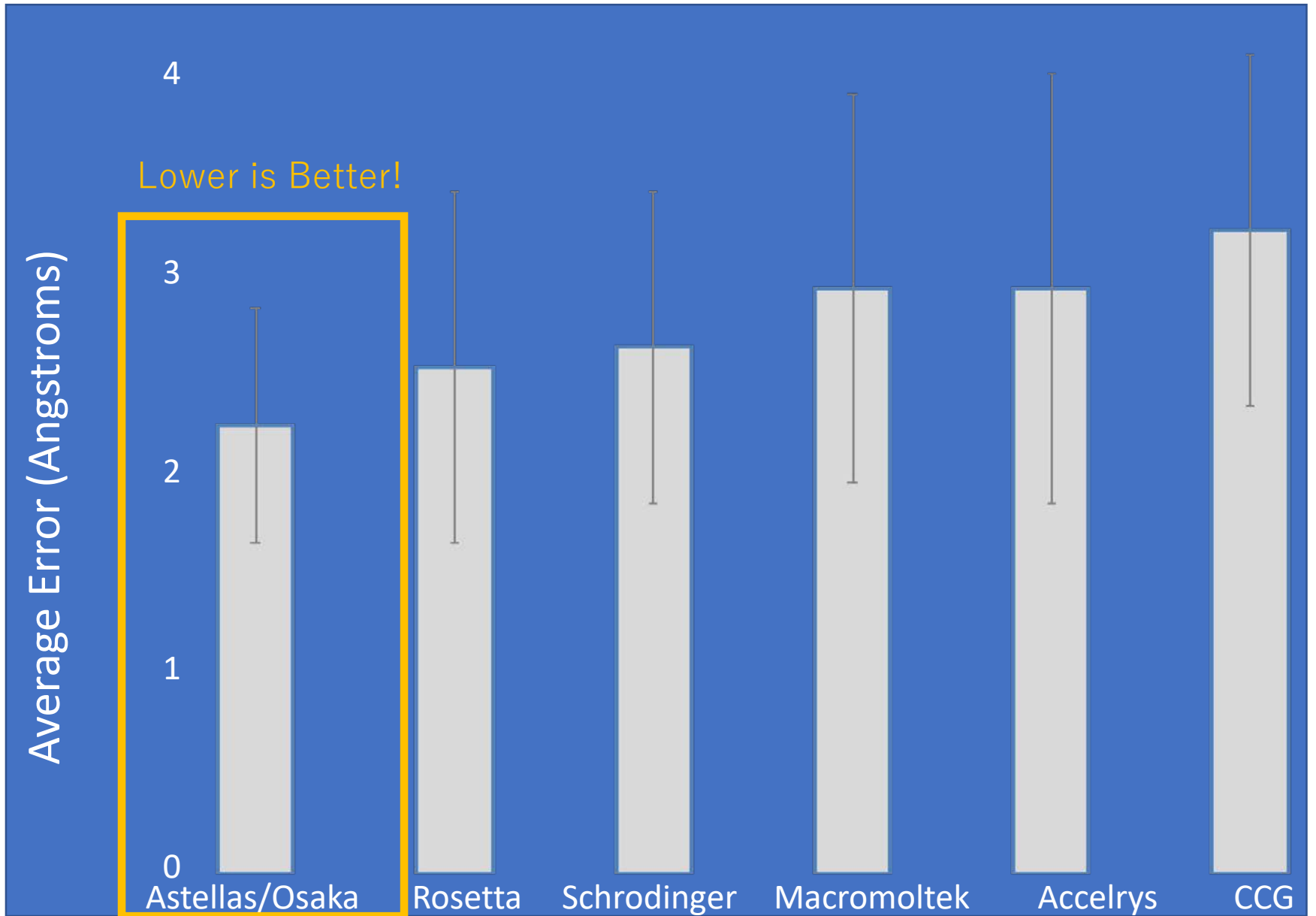
Experienced in antibody modeling

Our strategy



We wrote the software in 1 month!

Results: Our models had the lowest error!



It turned out the fast/risky approach was just as accurate as the slow approach

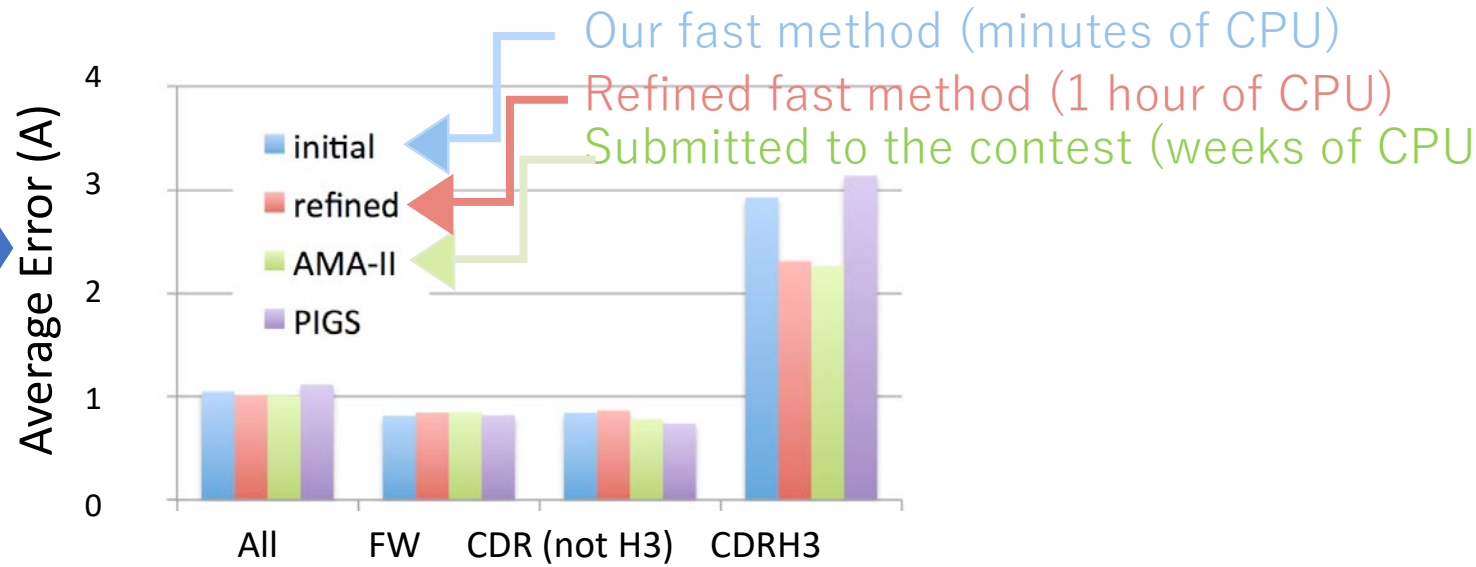
BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 22 2014, pages 3279–3280
doi:10.1093/bioinformatics/btu510

Structural bioinformatics Advance Access publication July 26, 2014

Kotai Antibody Builder: automated high-resolution structural modeling of antibodies

Kazuo Yamashita[1], Kazuyoshi Ikeda[2], Karlou Amada[1], Shide Liang[1], Yuko Tsuchiya[2], Haruki Nakamura[3], Hiroki Shirai[4] and Daron M. Standley[1,*]

AMA-II BCR sequences



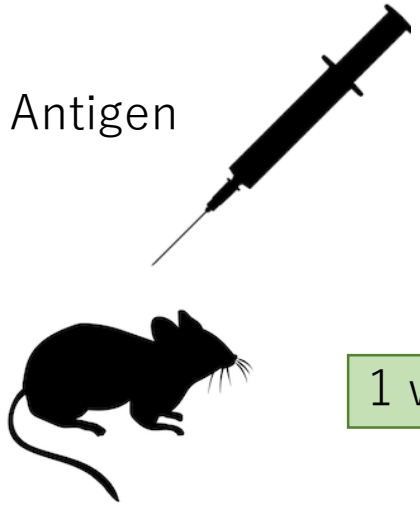
Kazuo Yamashita
CEO KOTAI Bio

Summary of part 1

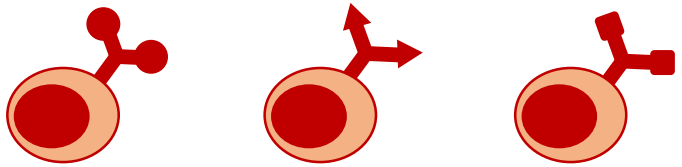
- The AMA-II contest provided an important lesson
- Our lab had no experience working on antibody modeling
- Nevertheless, we were able to make an important breakthrough
- Why did we do so well on antibody modeling?
- Possibly, because were not antibody ‘experts’ we tried something very naïve
- It just happened to work.
- It’s important to try different projects in your career
- Don’t become too attached to one project for your whole life
- Some projects “just happen to work” most projects don’t!

Part 2: Background on antibodies

How does nature make antibodies?

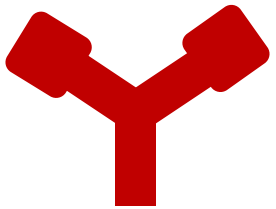


B cells with unique B cell receptors (BCRs)

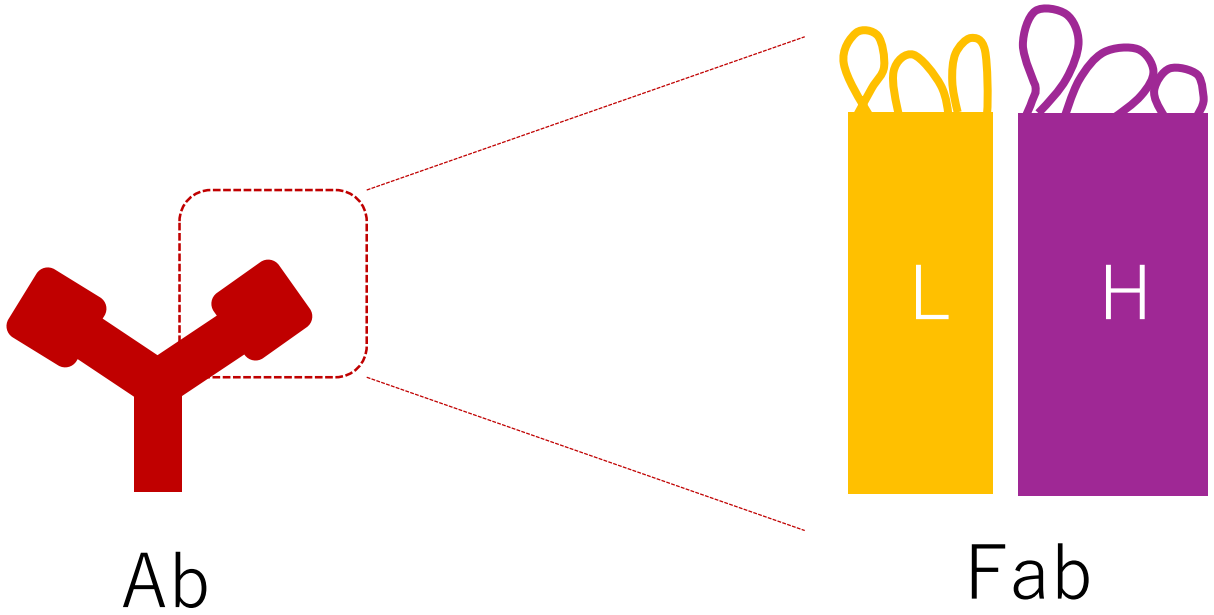


One or more of these BCRs will bind to the antigen of interes

Antibodies are the soluble form of BCRs



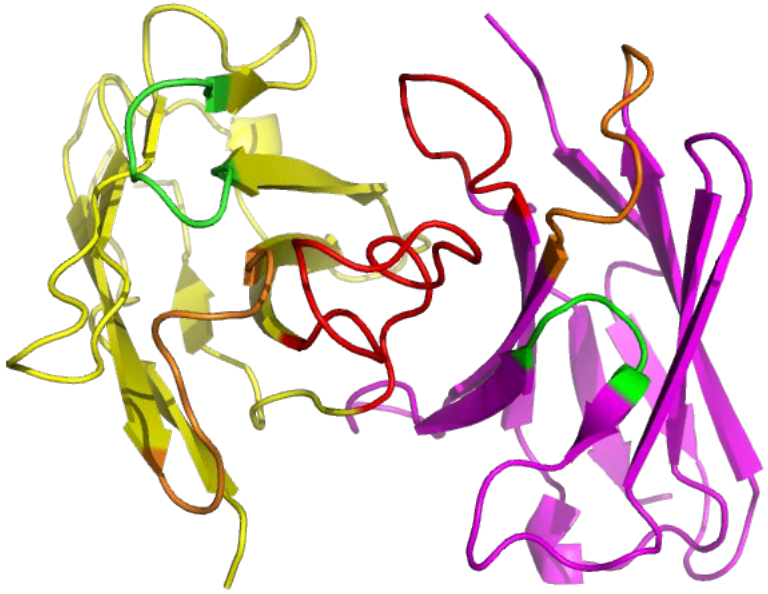
Each antibody variable region (Fab) is composed of two chains (heavy and light)

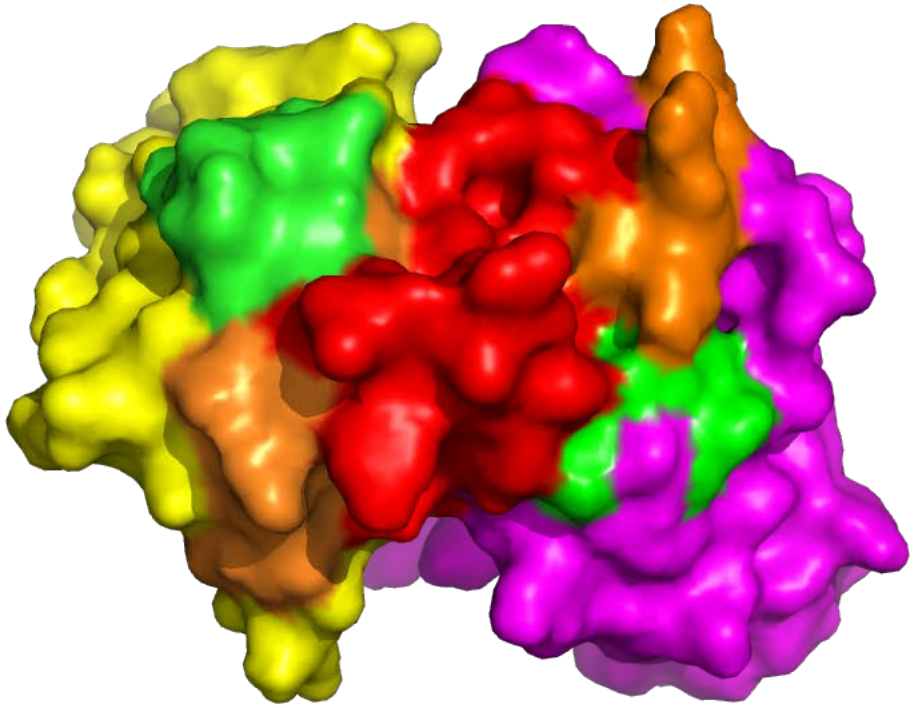


A Fab contains 6 complementarity-determining regions (CDRs)

These six CDR loops give each antibody its unque antigen-binding characteristics

Antigen binding surfaces (paratopes) are defined by CDRs



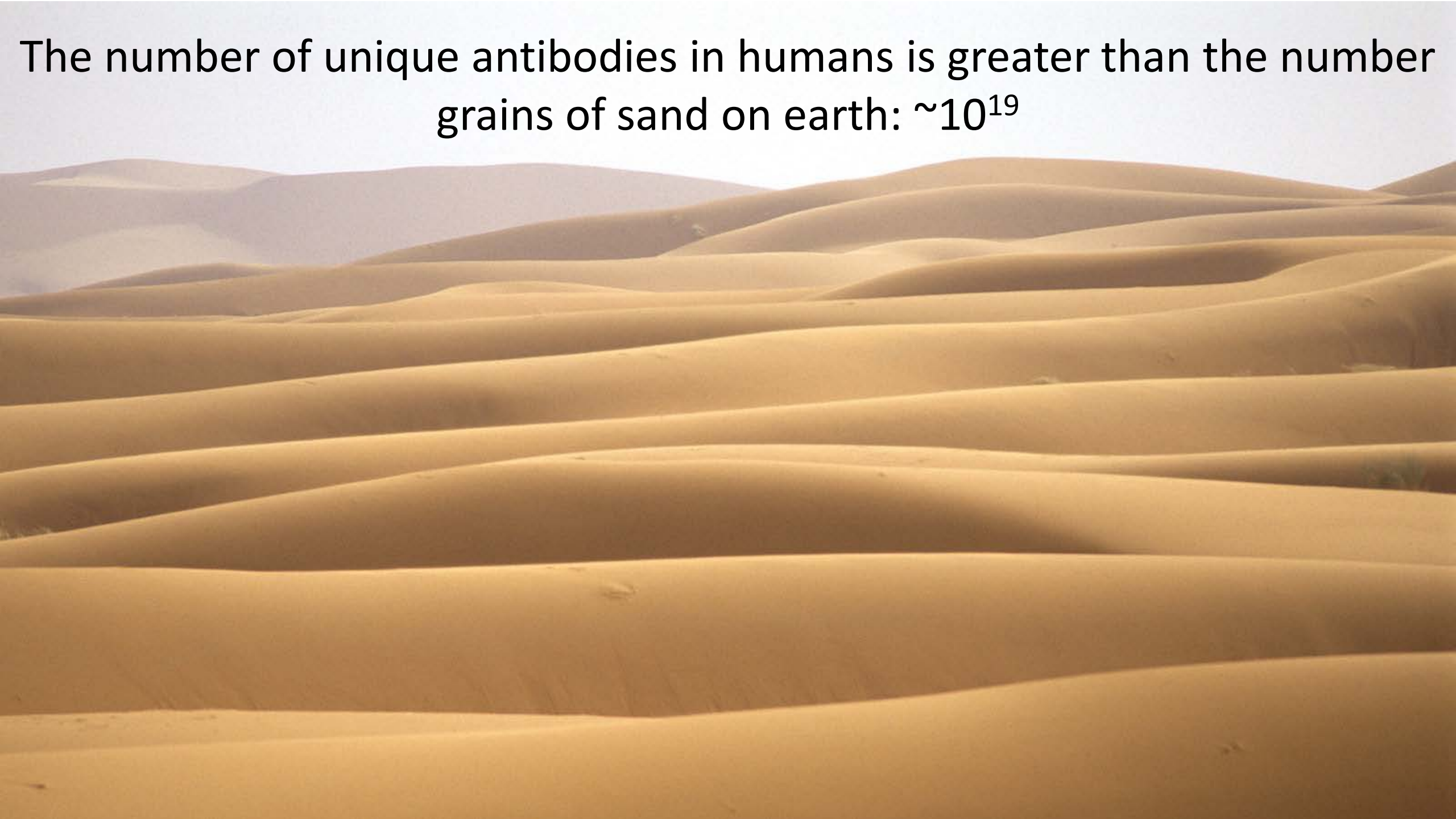


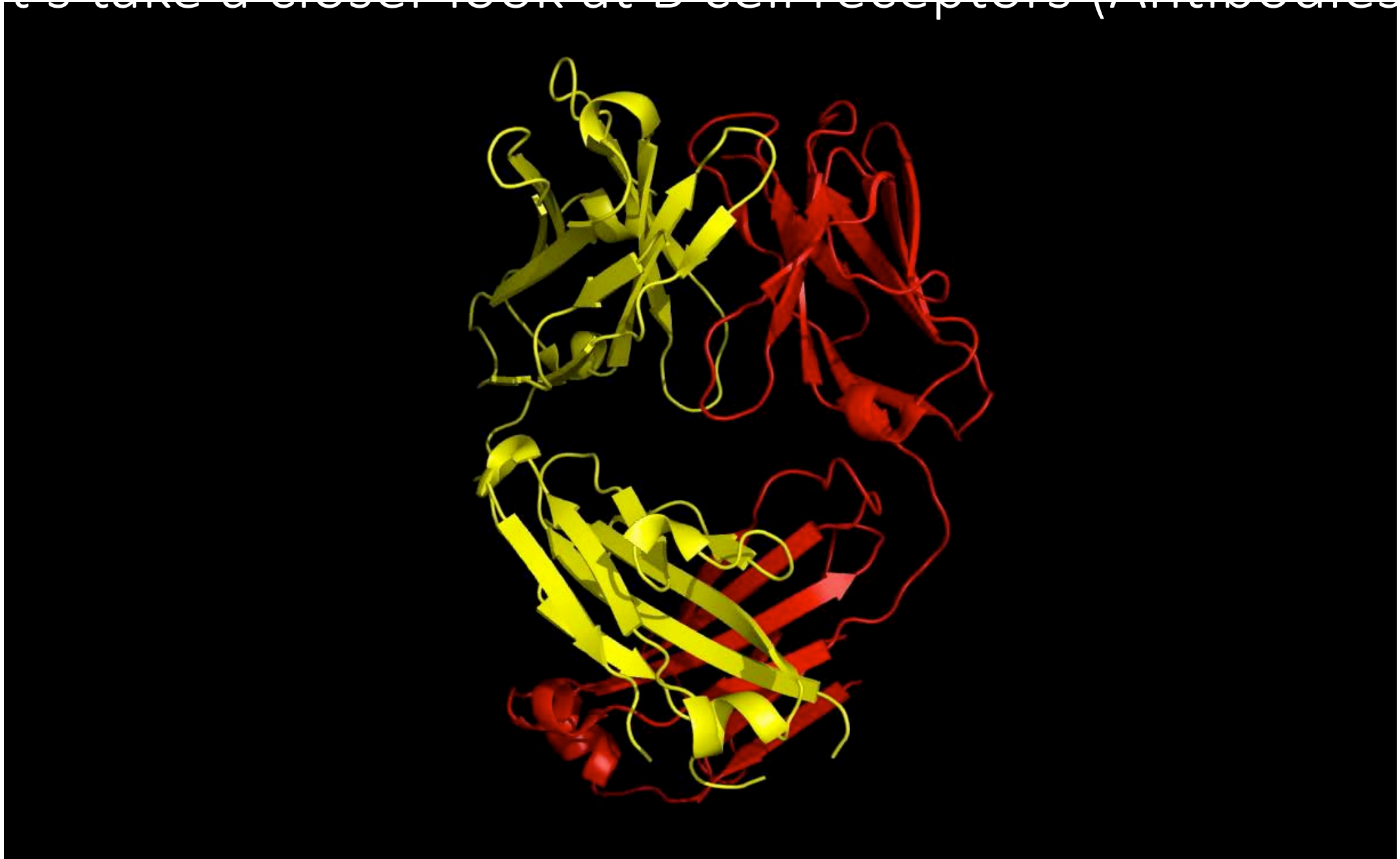
L1 L2 L3 H3 H2 H1

CDR sequences are generated randomly

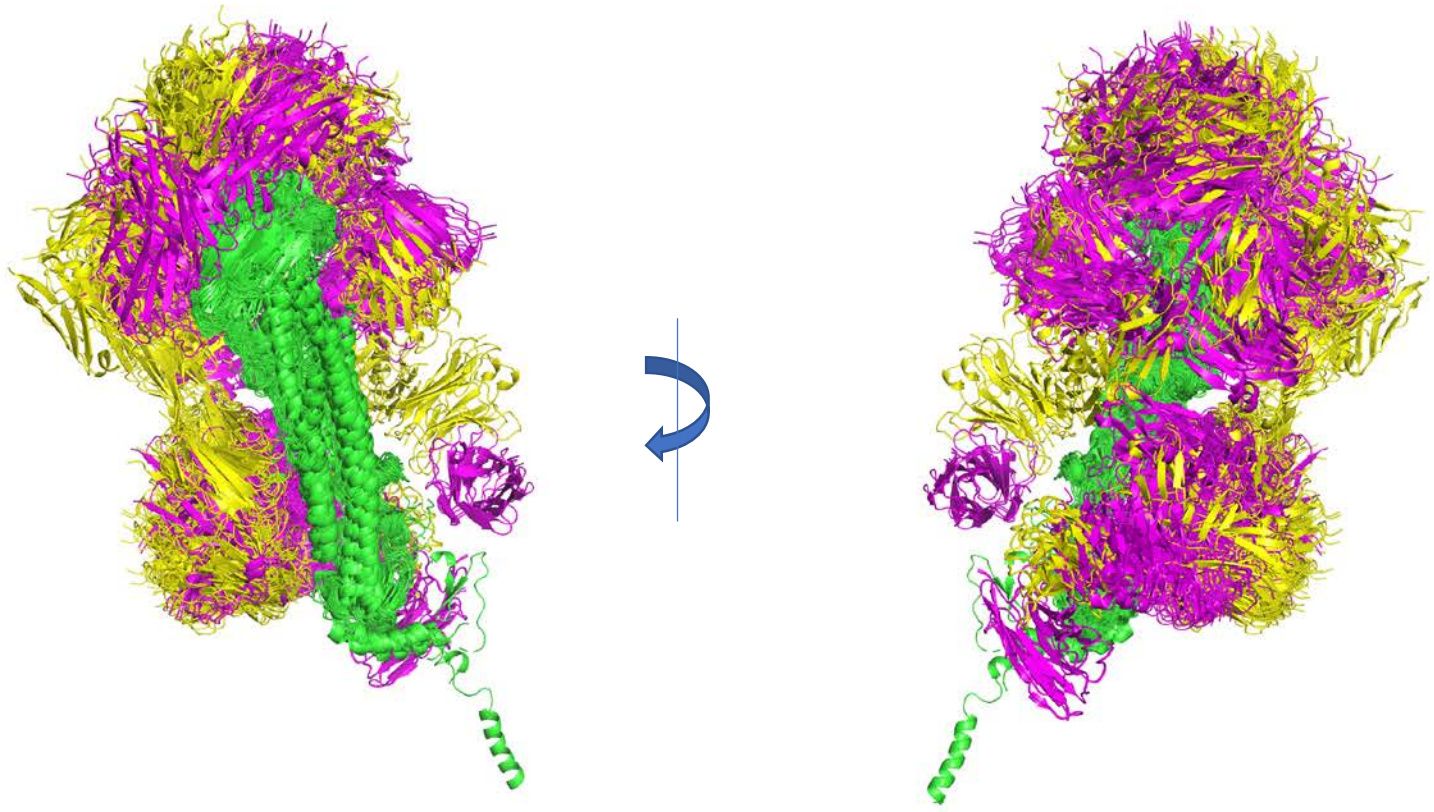
This is a continuous process ocurring throughout our lives

The number of unique antibodies in humans is greater than the number grains of sand on earth: ~10^{19}



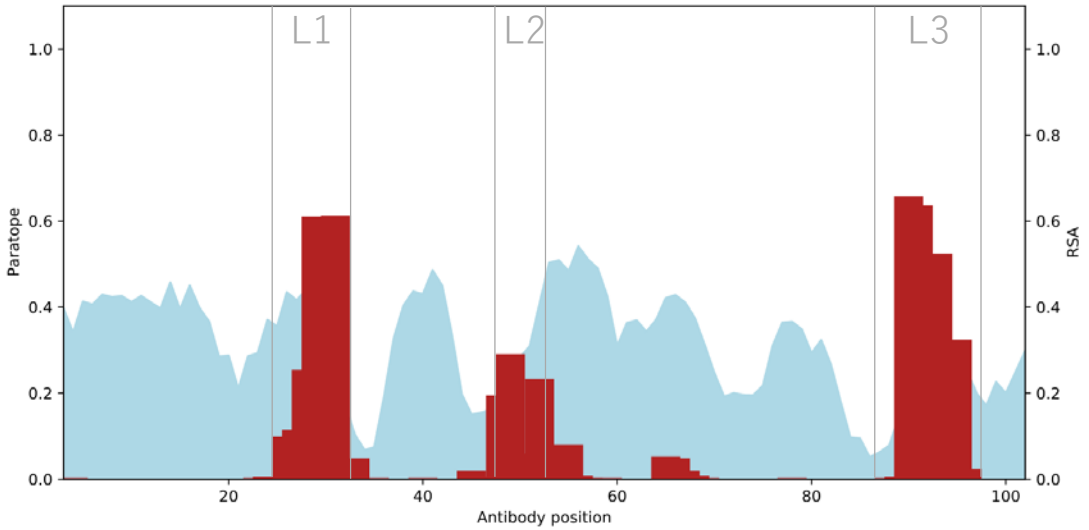


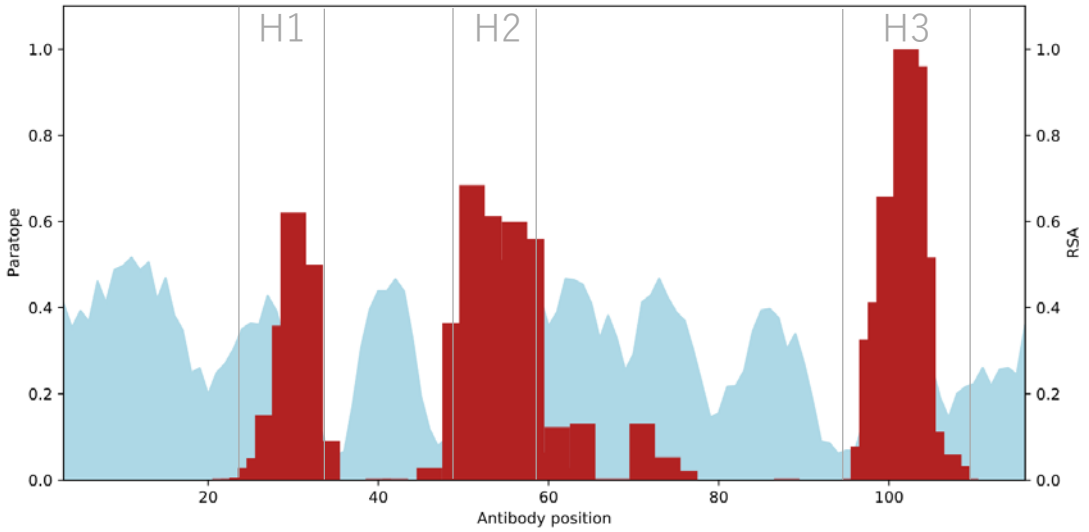
Antibody binding sites (epitopes) can be anywhere on an antigen surface



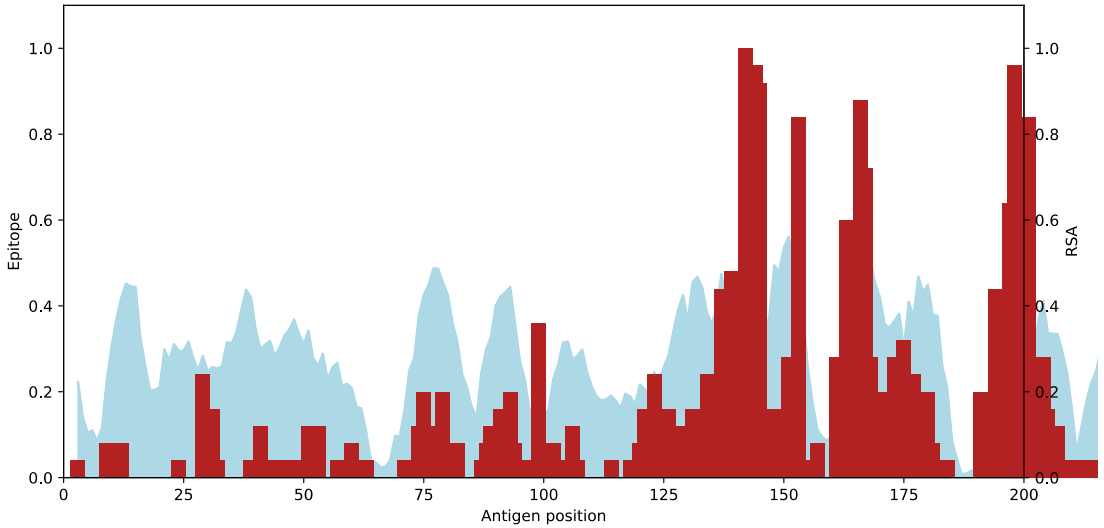
Different antibodies targeting
influenza hemagglutinin

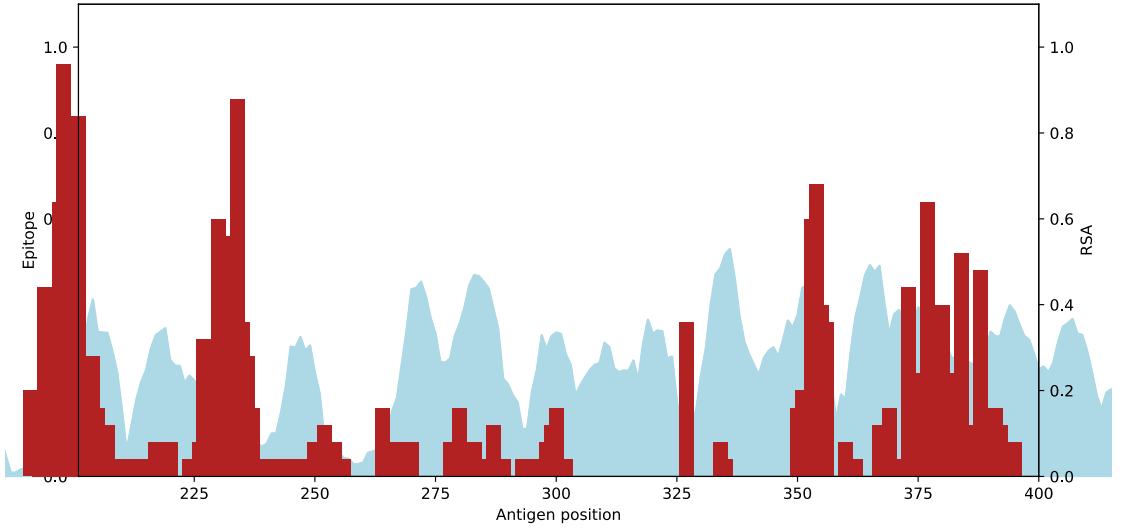
Paratopes correspond to CDRs



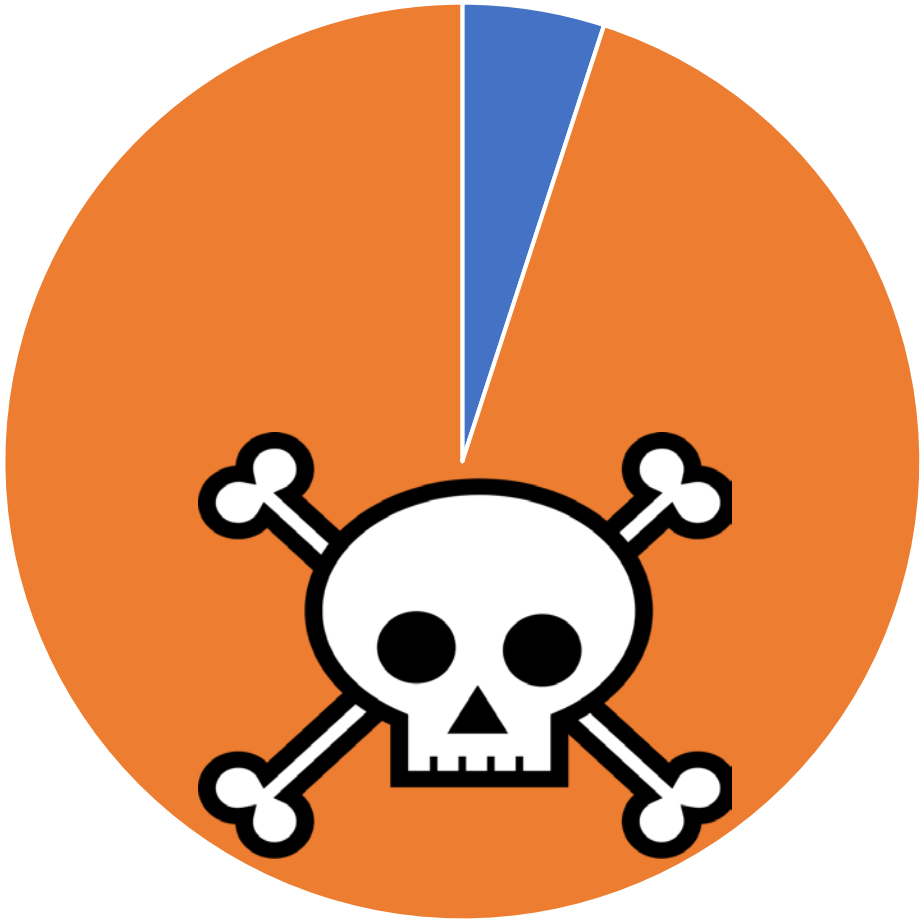


Epitopes can be anywhere (e.g. Hemagglutinin)



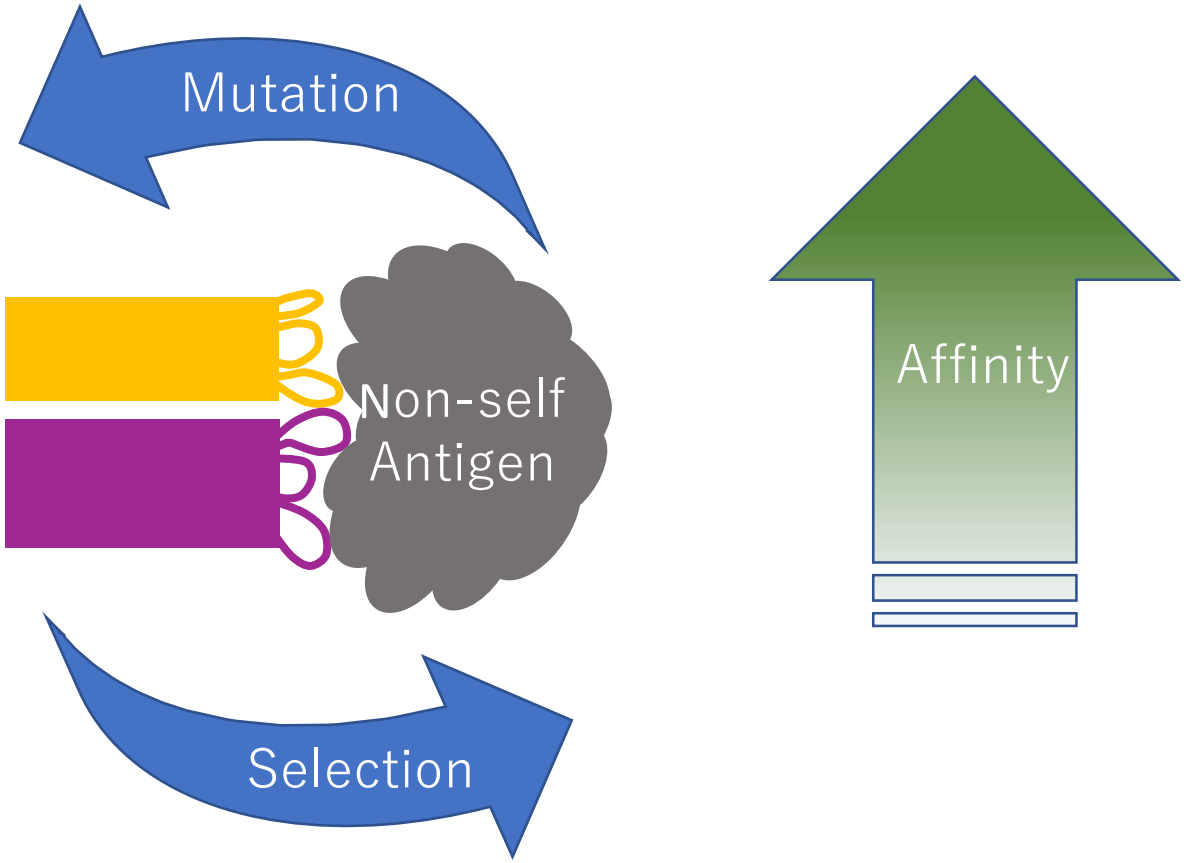


B cells that bind self-molecules are killed

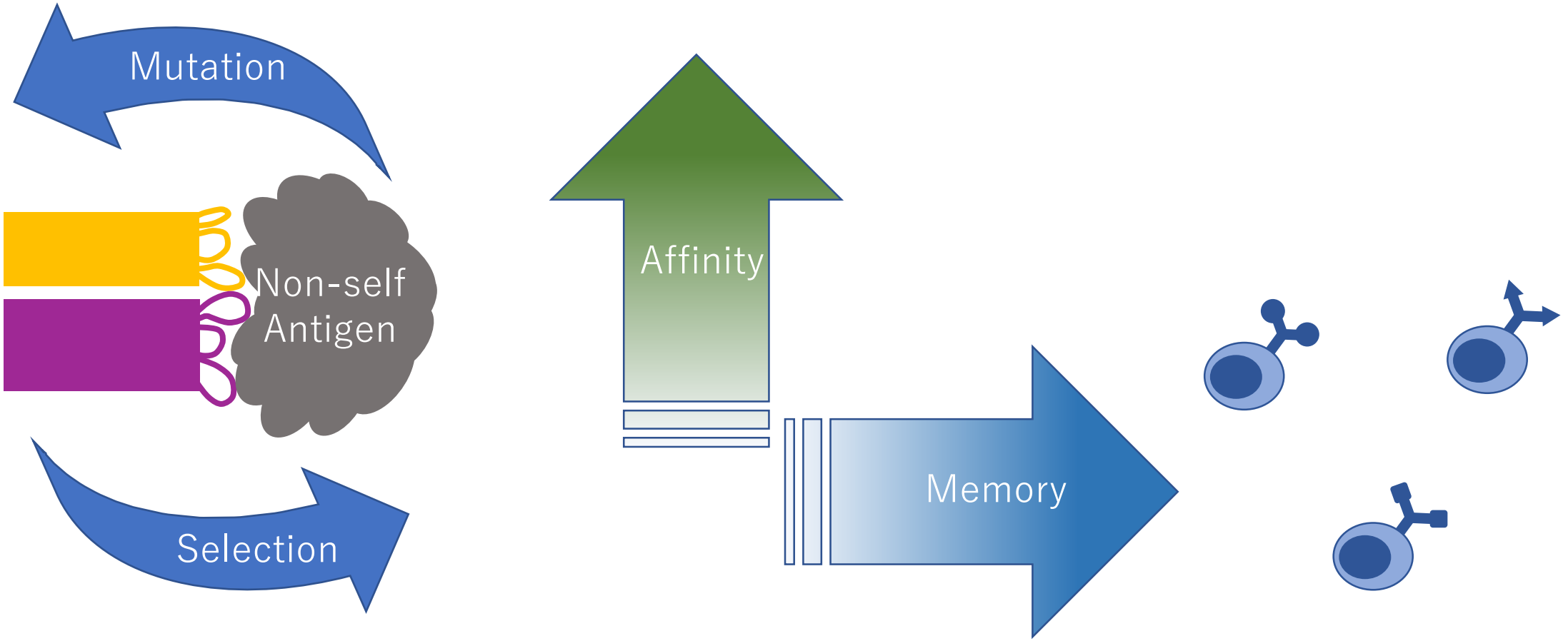


95% of B cells are killed because of self-reactivity

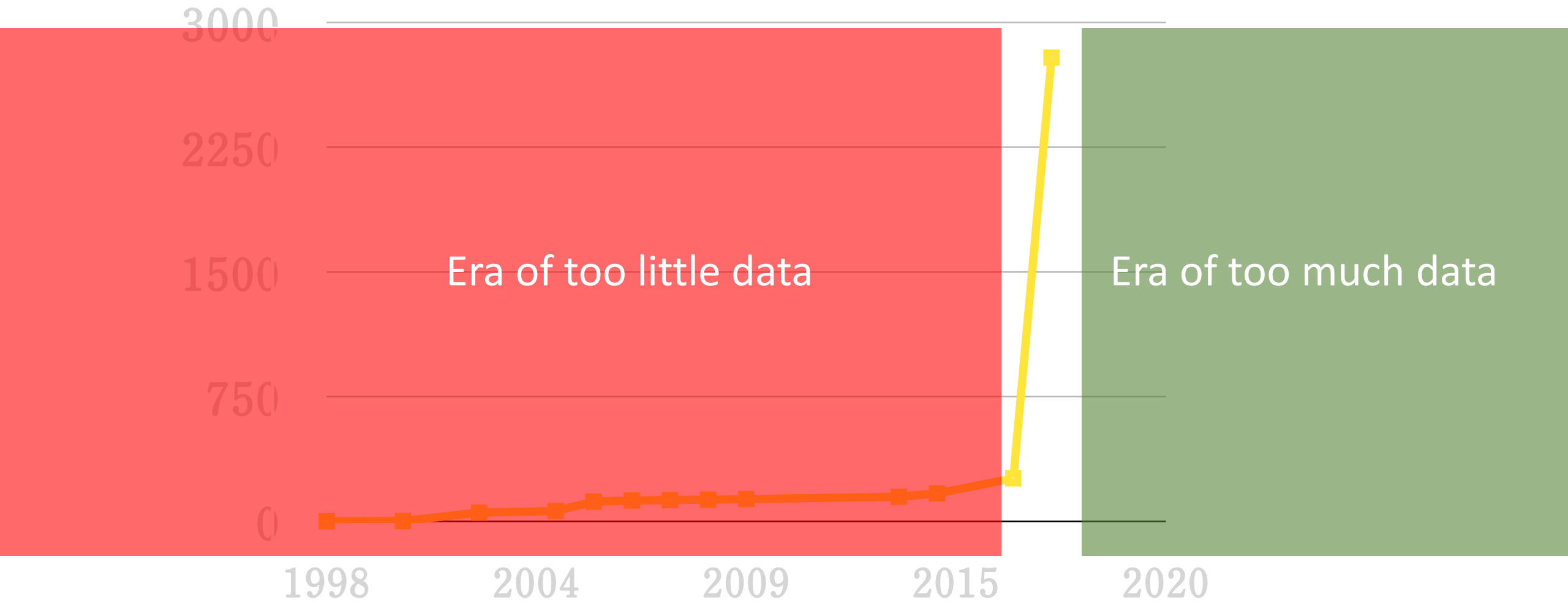
Surviving B cells that bind non-self molecules undergo further optimization



Some cells become long-lived memory B cells



B cell sequencing is an emerging technology


3000
2250
1500
750
0
Era of too little data
Era of too much data
1998
2004
2009
2015
2020

Traditional therapeutic antibody discovery takes time

- Identify target antigen
- Develop antibodies in animals
- Select several monoclonal antibodies
- Humanize (i.e. graft CDRs from animal framework onto human framework)
- Optimize antigen binding etc.
- Evaluate safety
- Evaluate efficacy

Two important examples from Japanese academia

Tasuku Honjo, Kyoto University
Discovered PD-1 cancer checkpoint therapy

Tadamitsu Kishimoto, Osaka University
Discovered IL6-based autoimmune therapy

Invention:
PD-1 discovery (1992)

The EMBO Journal vol.11 no.11 pp.3887 – 3895, 1992

Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death

Yasumasa Ishida, Yasutoshi Agata, Keiichi Shibahara and Tasuku Honjo[1]

Department of Medical Chemistry, Kyoto University Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto 606, Japan

[1]Corresponding author

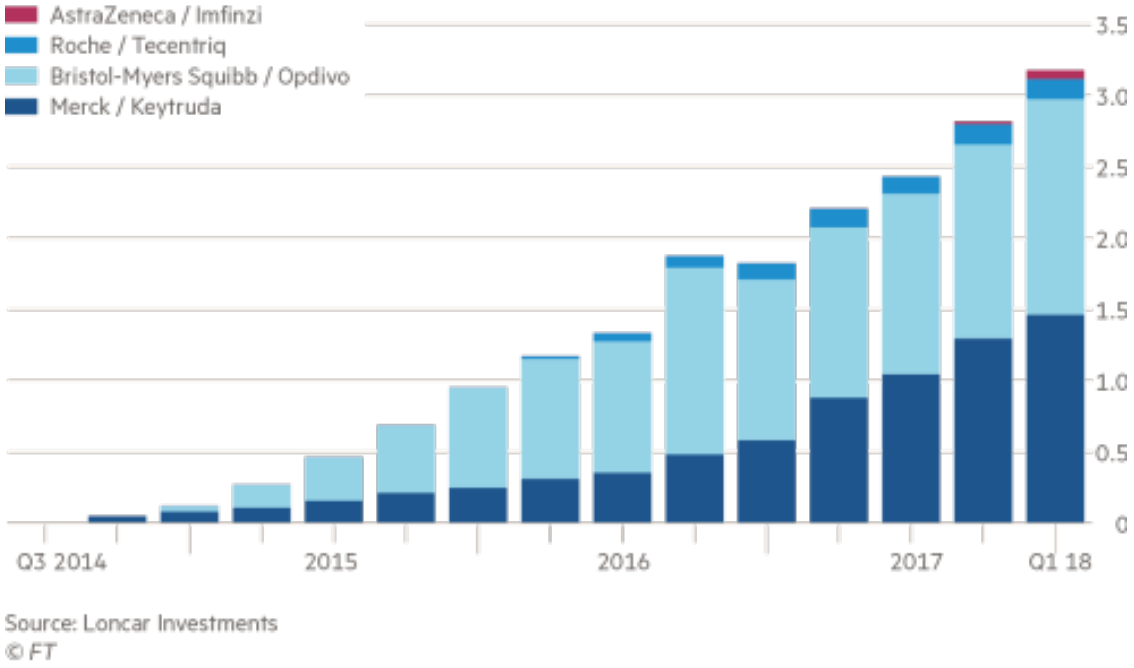
Communicated by J.Tooze

The classical type of programmed cell death is characterized by its dependence on *de novo* RNA and protein synthesis and morphological features of apoptosis. We confirmed that stimulated 2B4.11 (a murine T-cell hybridoma) and interleukin-3 (IL-3)-deprived LyD9 (a murine haematopoietic progenitor cell line) died by the classical type of programmed cell death. Assuming that common biochemical pathways might be involved in the deaths of 2B4.11 and LyD9, we isolated the PD-1 gene,

of actinomycin D or cycloheximide on the death of nerve growth factor (NGF)-deprived rat neurons (Martin *et al.*, 1988) and that of mouse thymocytes induced by glucocorticoids (Cohen and Duke, 1984) or by an endogenous superantigen (MacDonald and Lees, 1990). These facts suggest that at least a few genes, if not specific ones, must be expressed to cause programmed cell death.

The term 'apoptosis', on the other hand, is used to describe the morphological characteristics of a class of cell death (Kerr and Harmon, 1991). In cells dying by apoptosis, the chromatin condenses around the periphery of the nucleus, while the mitochondria and other organelles are unaffected. A unique biochemical feature of apoptotic cells includes fragmentation of DNA into oligonucleosomal pieces. Apoptosis is often associated with programmed cell death, but some of the cells undergoing programmed death apparently do not show the characteristic features of

Commercialization:
Global sales of checkpoint-therapies (2014-2018)



Source: Loncar Investments
© FT

Invention:
IL-6 discovery (1988)

REPORTS

Cloning and expression of the human interleukin-6 (BSF-2/IFN beta 2) receptor

K Yamasaki, T Taga, Y Hirata, H Yawata, Y Kawanishi, B Seed, T Taniguchi, T Hirano, T Kishimoto
+ See all authors and affiliations

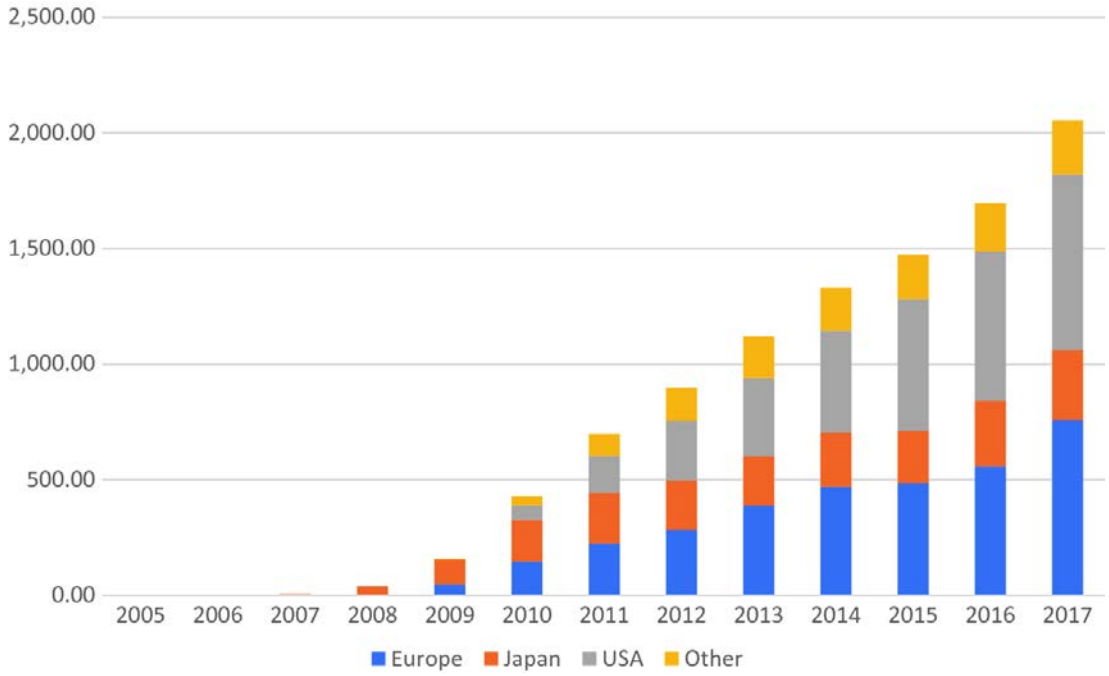
Science 12 Aug 1988:
Vol. 241, Issue 4867, pp. 825-828
DOI: 10.1126/science.3136546

Article | Info & Metrics | eLetters | PDF

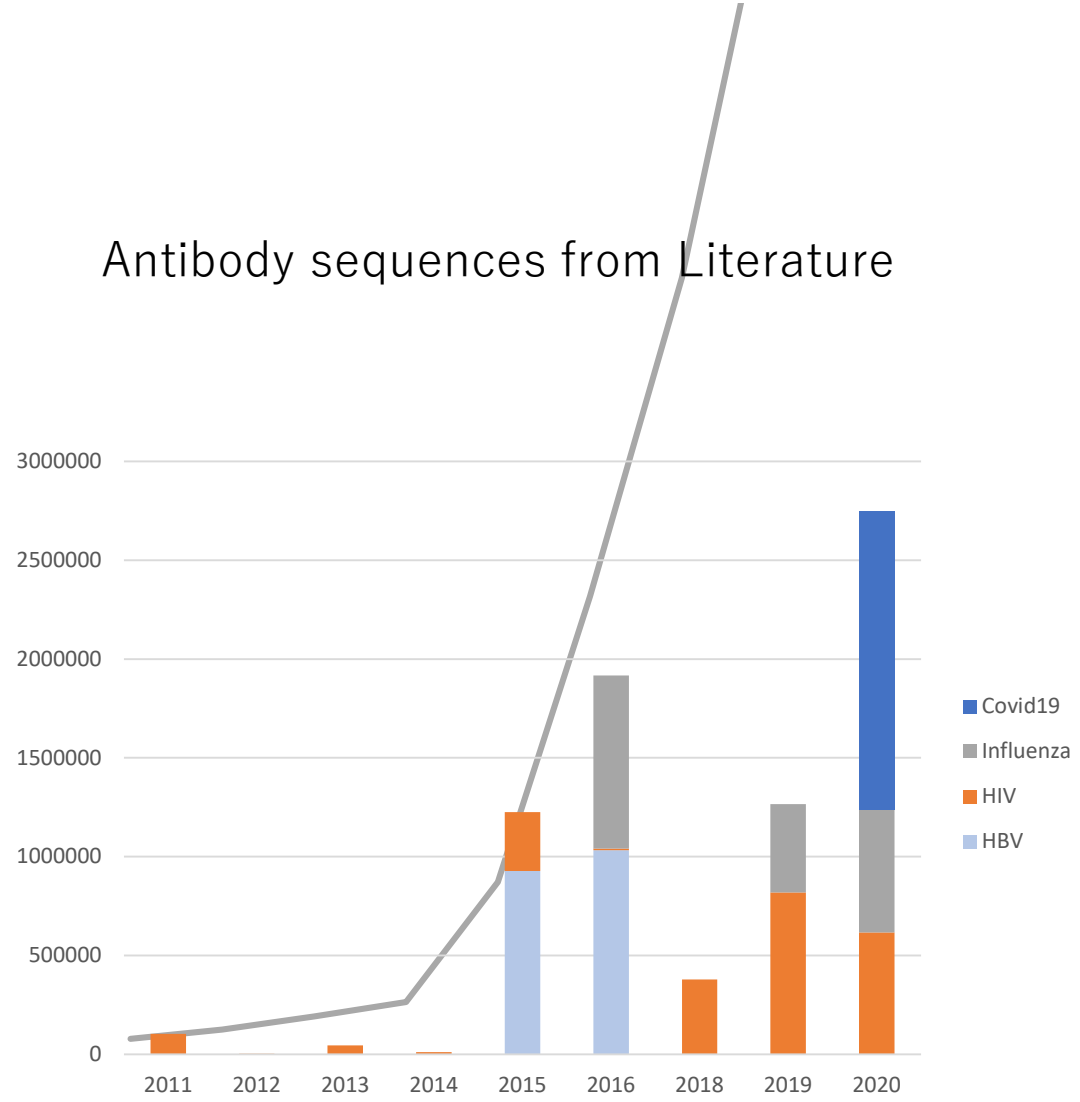
Abstract

Interleukin-6 (IL-6/BSF-2/IFN beta 2) is a multifunctional cytokine that regulates the growth and differentiation of various tissues, and is known particularly for its role in the immune response and acute phase reactions. A complementary DNA encoding the human IL-6 receptor (IL-6-R) has now been isolated. The IL-6-R consists of 468 amino acids, including a signal peptide of approximately 19 amino acids and a domain of

Commercialization:
Global sales of anti-IL6 therapy (2005-2017)



> 20,000,000 antibodies associated with 4 viruses



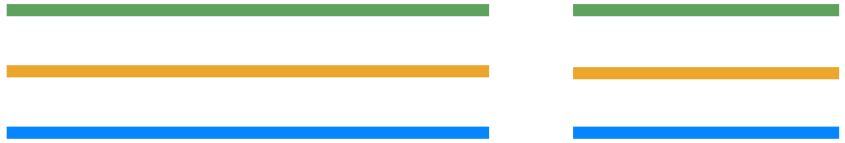
Summary of Part 2

- Collectively, antibodies exhibit a kind of “Immune intelligence”
  - They distinguish between “self” and “non-self”
  - They “learn” to bind to non-self molecules with high affinity and specificity
  - They remember non-self antigens for the future
- High-throughput antibody sequencing marks the beginning of a new era
- There is room for new bioinformatics methods in this exciting field
- Hopefully, we can reduce the time to discover new therapeutic antibodies

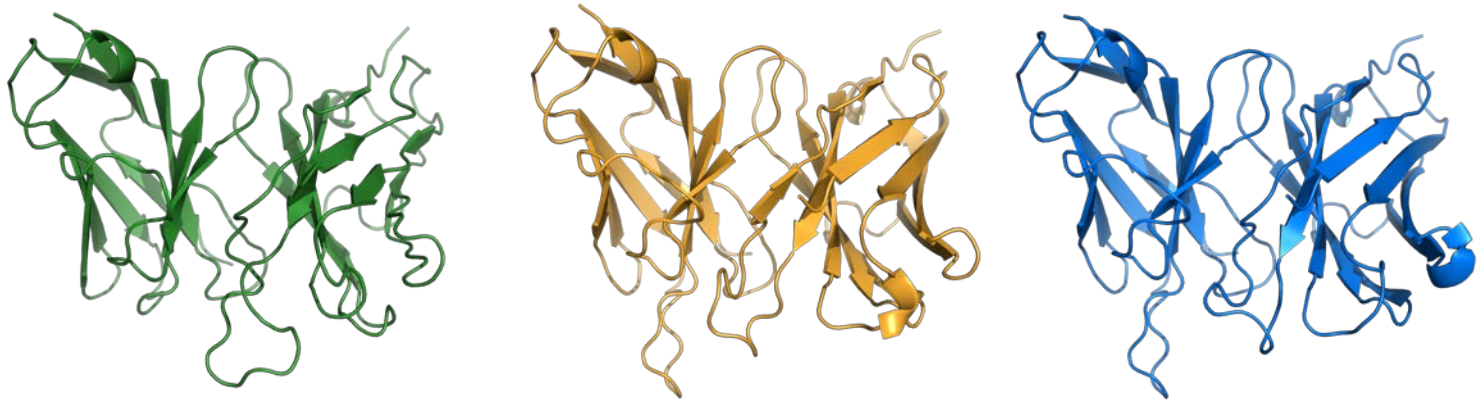
Part3: High-throughput Antibody Modeling

Repertoire Builder: Our latest software

Input:
Up to 10,000 Antibody
or sequences

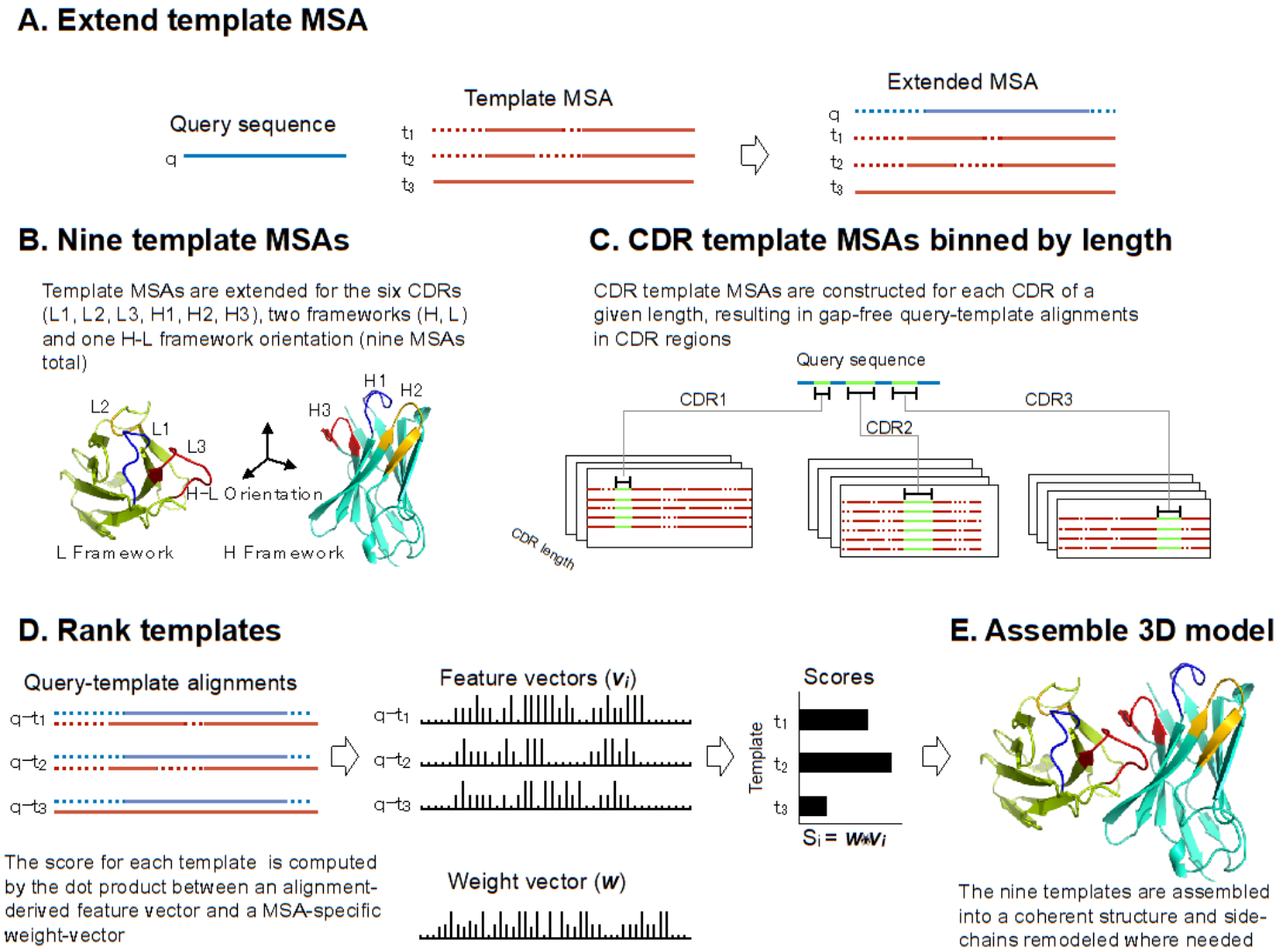


Output:
Up to 10,000
receptor
models

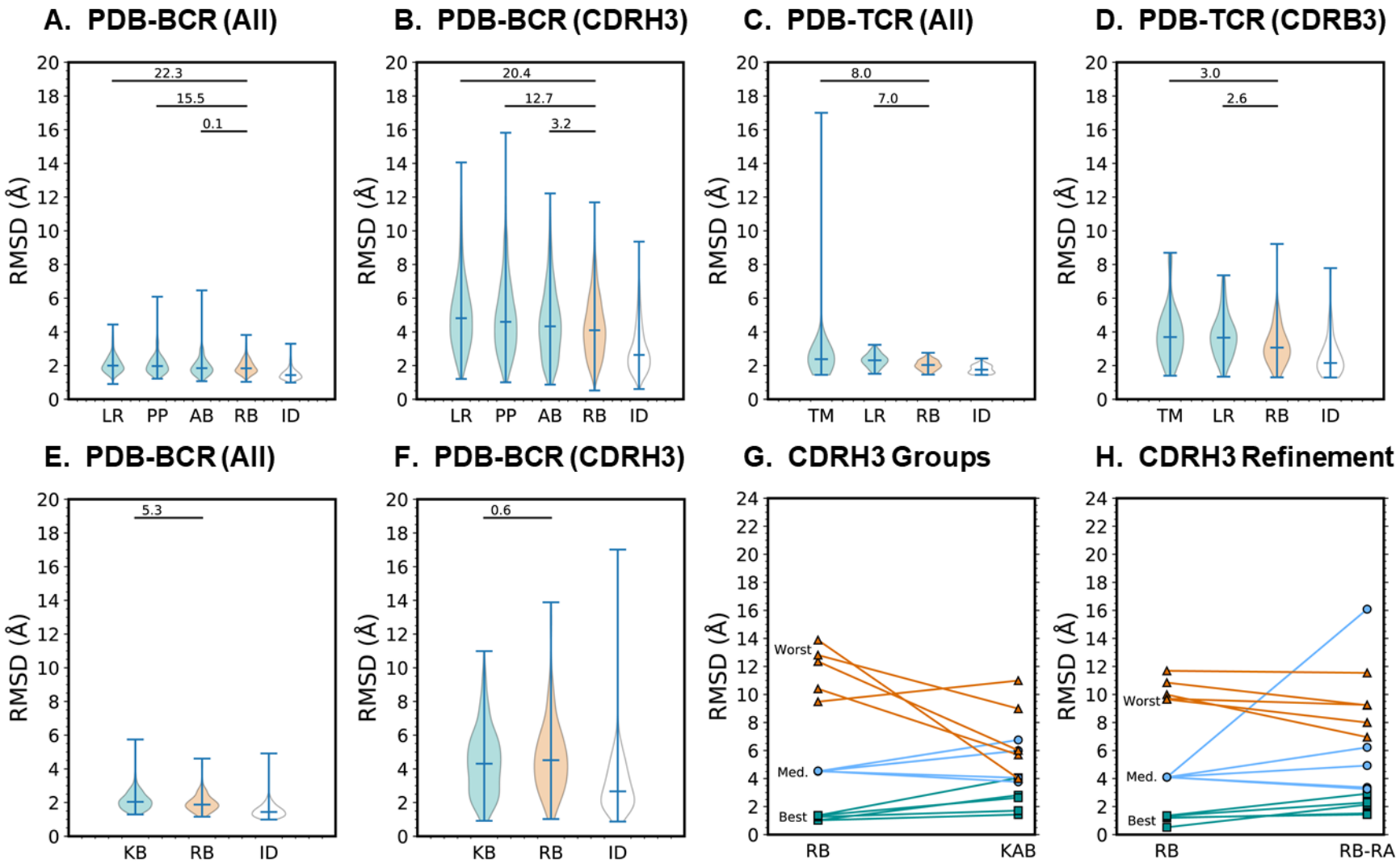


D Schritt, et al. Molecular Systems Design & Engineering 4 (4), (2019)

Repertoire Builder utiluzes MAFFT-ASH-derived feature-vectors



Repertoire Builder Accuracy



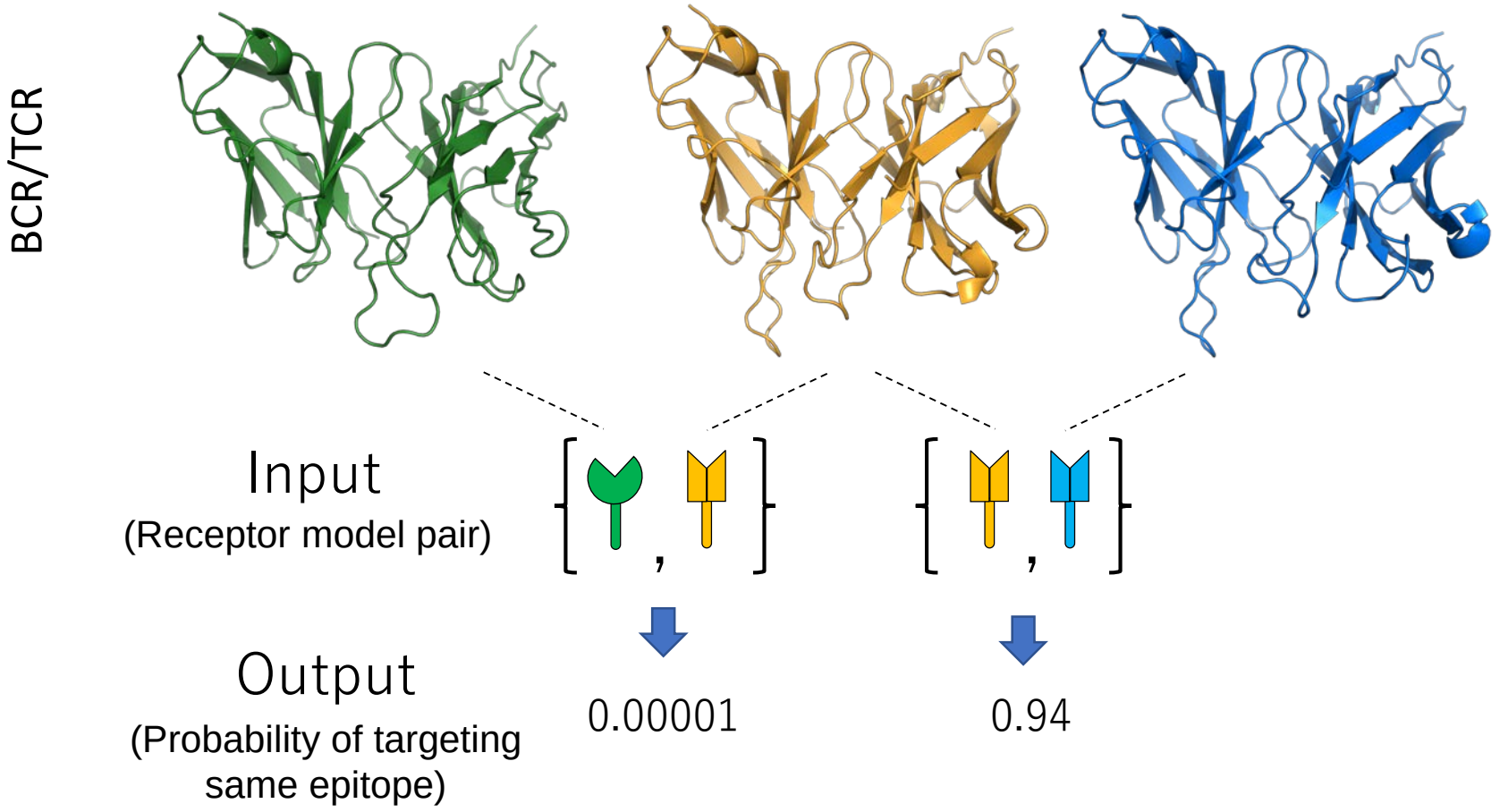
LR: LYRA
PP: PigsPro
AB: ABodyBuilder
RB: Repertoire Builder
ID: Ideal
KB: Kotai Antibody Builder
TM: TCRmodel
RB-A: Repertoire Builder refined by Rosetta Antibody

Summary of Part 3

- Antibody modeling can be very fast (seconds per model)
  - Other very fast modeling methods include Lyra (**Marcatili et al.**) and AbodyBuilder (Deane et al.)
- Repertoire Builder appears to be both fast and acurate
- Sometimes huge computational cost (e.g. MD-based methods) are not better than fast methods (i.e. using structural templates)
- However, making lots of antibody structural models is not enough
- We need to find a way to use this information to understand antibody function
- This means: predicting antigen specificity

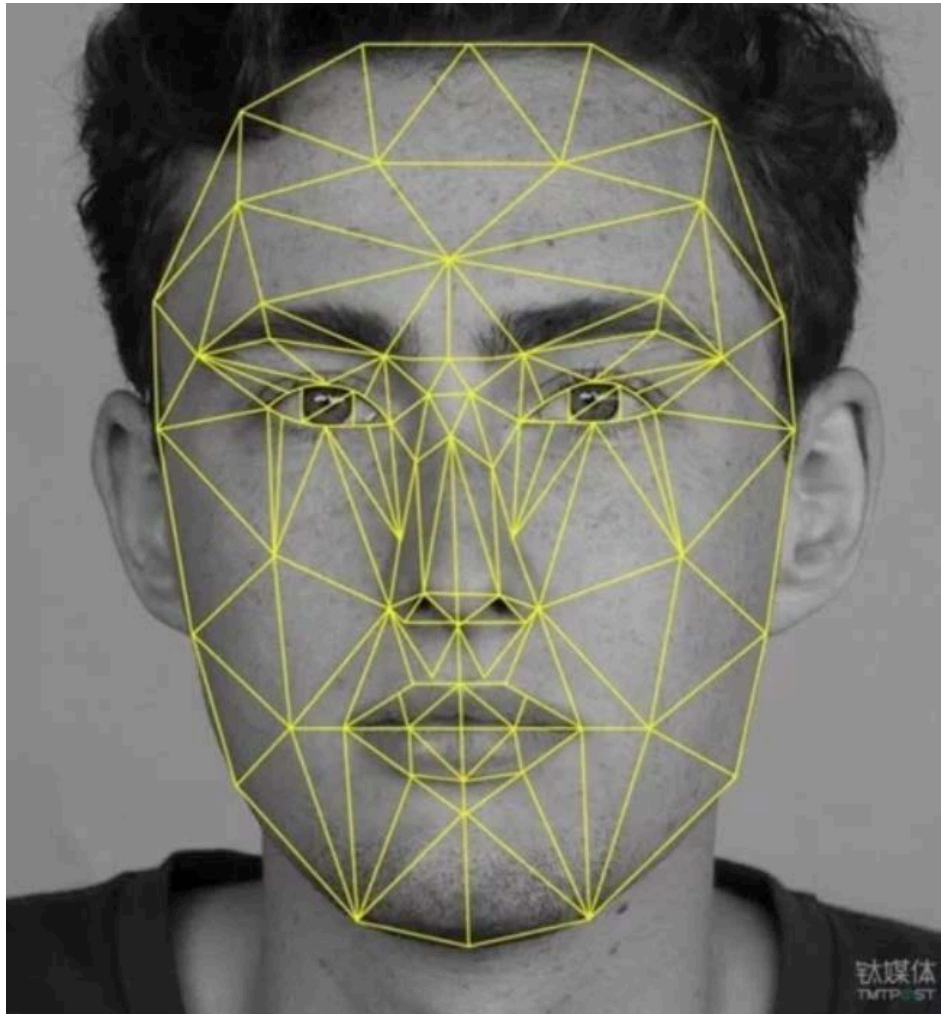
Part 4: Antibody Clustering

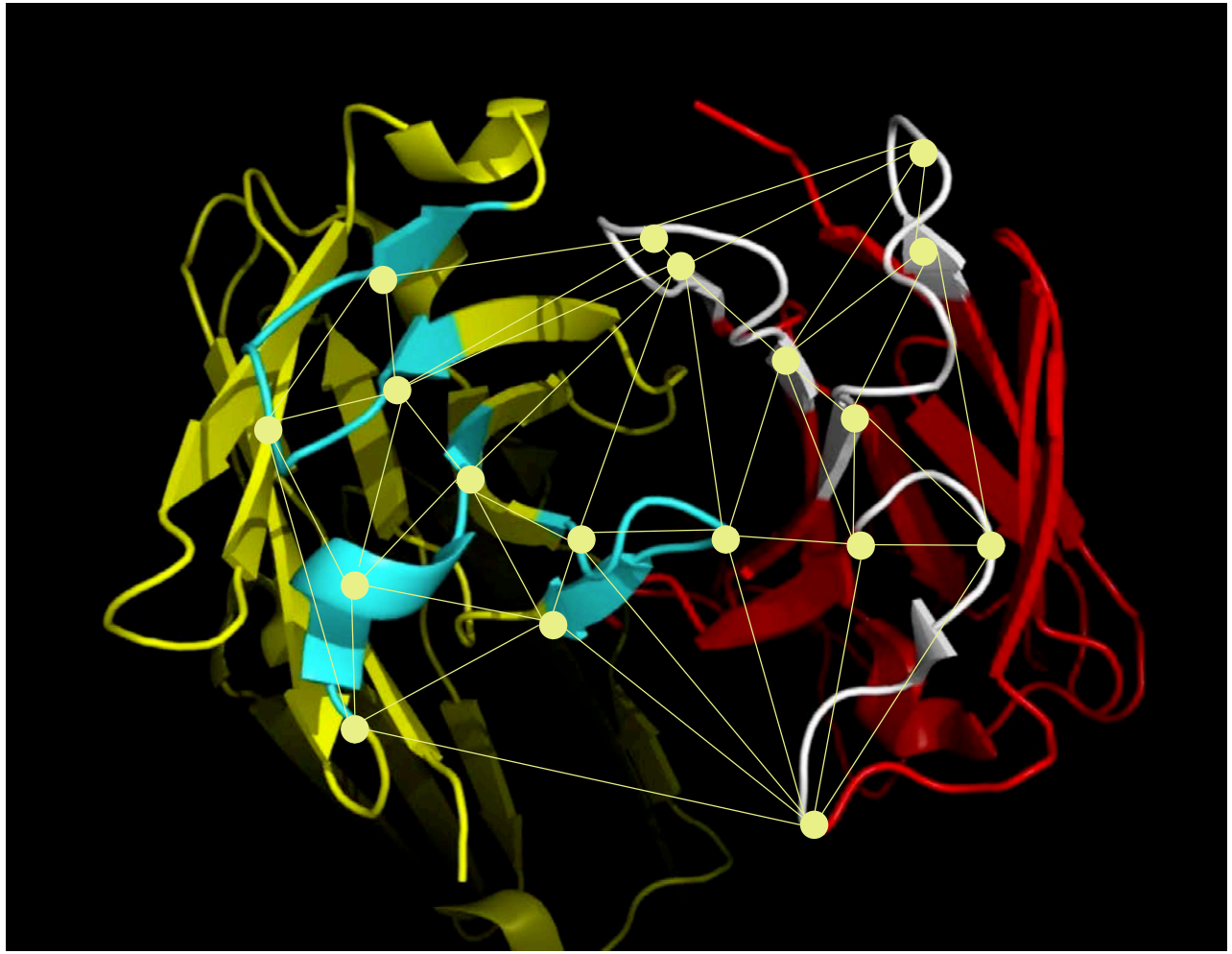
InterClone: Pairwise classifier with hierarchical clustering



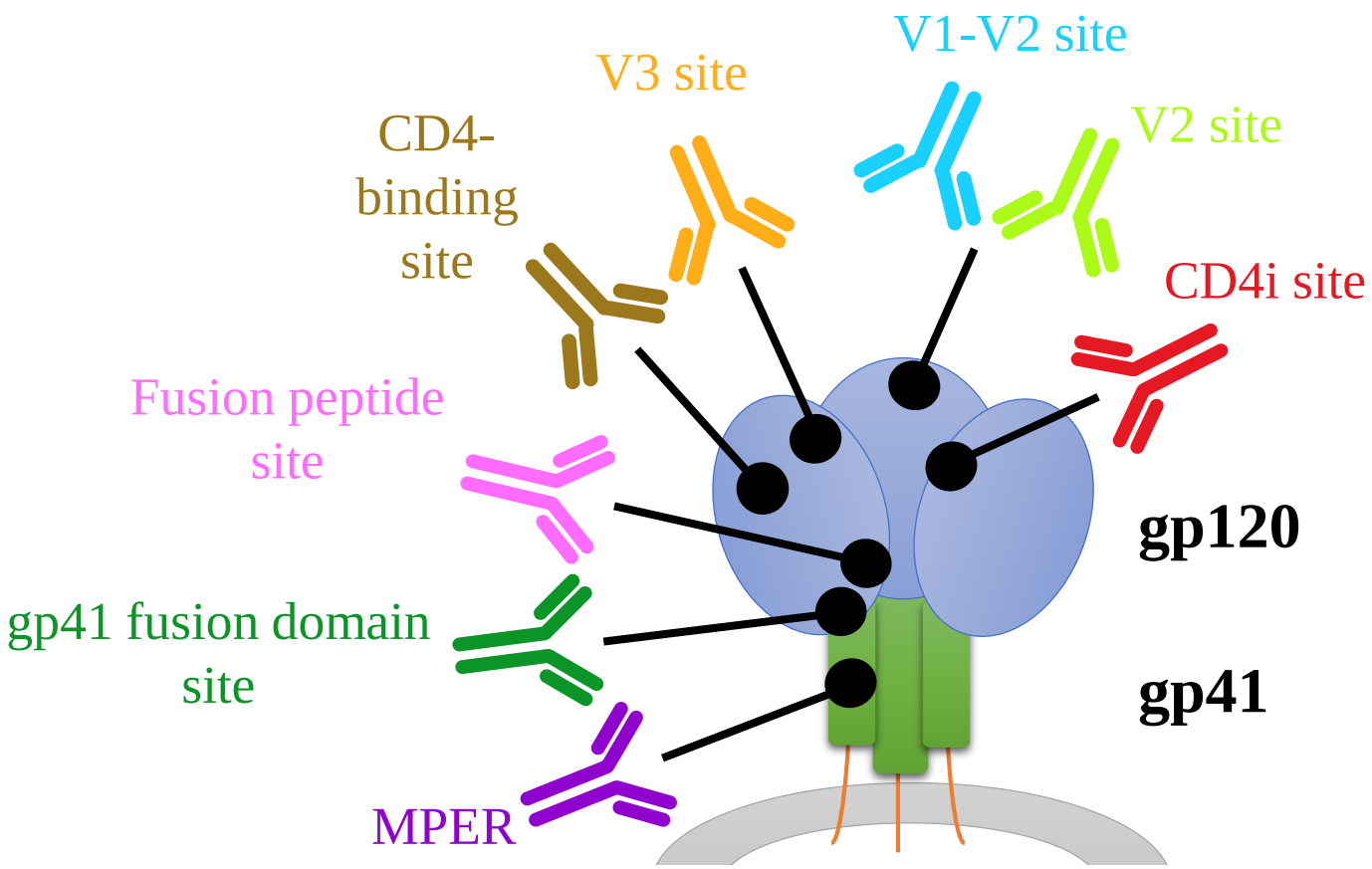
Xu et al Molecular Systems Design & Engineering 4 (4), (2019)

InterClone AI similar to facial recognition





CATNAP : HIV-1 antibody library

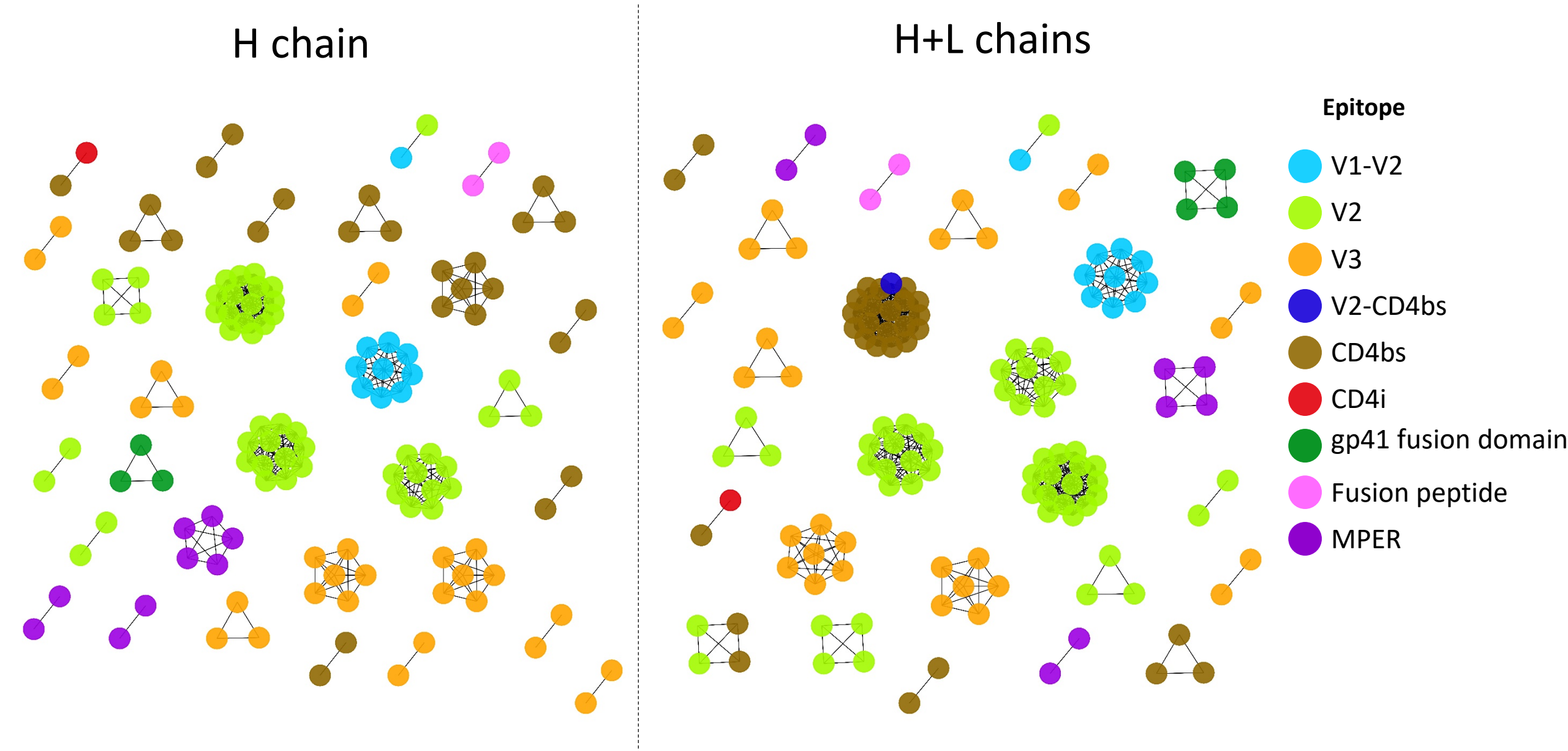


Epitope

- V1-V2
- V2
- V3
- V2-CD4bs
- CD4bs
- CD4i
- gp41 fusion domain
- Fusion peptide
- MPER

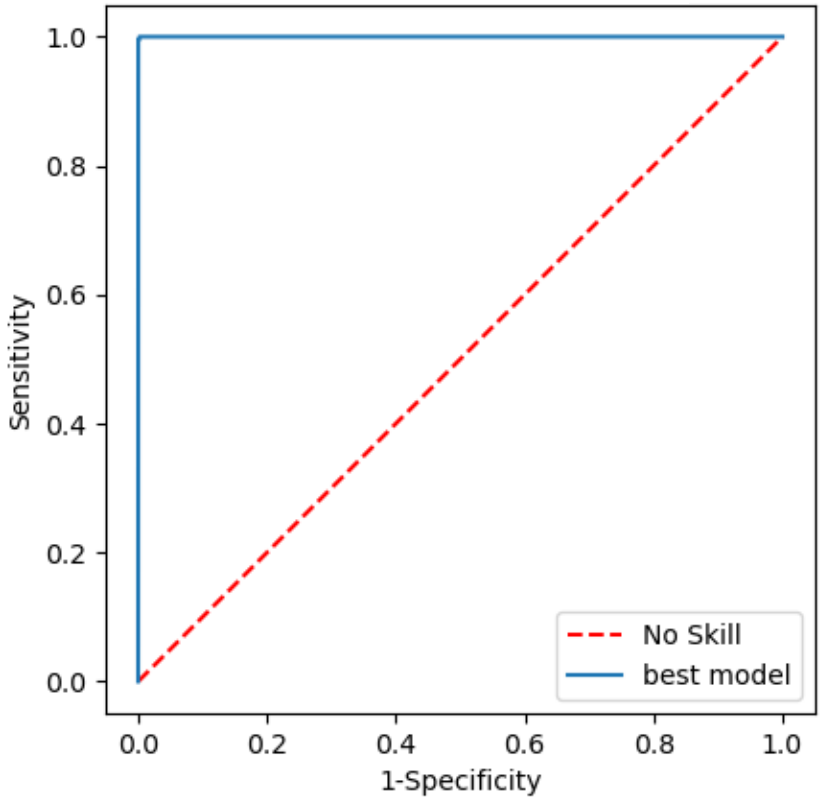
variable loops (V1,V2,V3)
CD4-induced (CD4i)
CD4 binding site (CD4BS)
membrane proximal external region (MPER)

Cocluster CANTAP dataset by InterClone

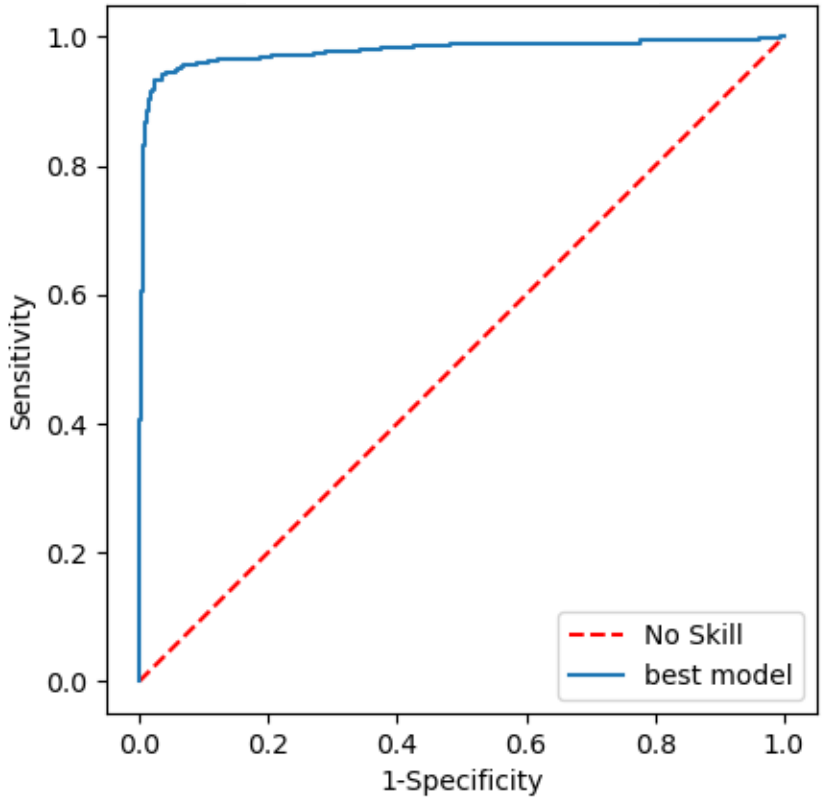


InterClone performance on CATNAP antibodies

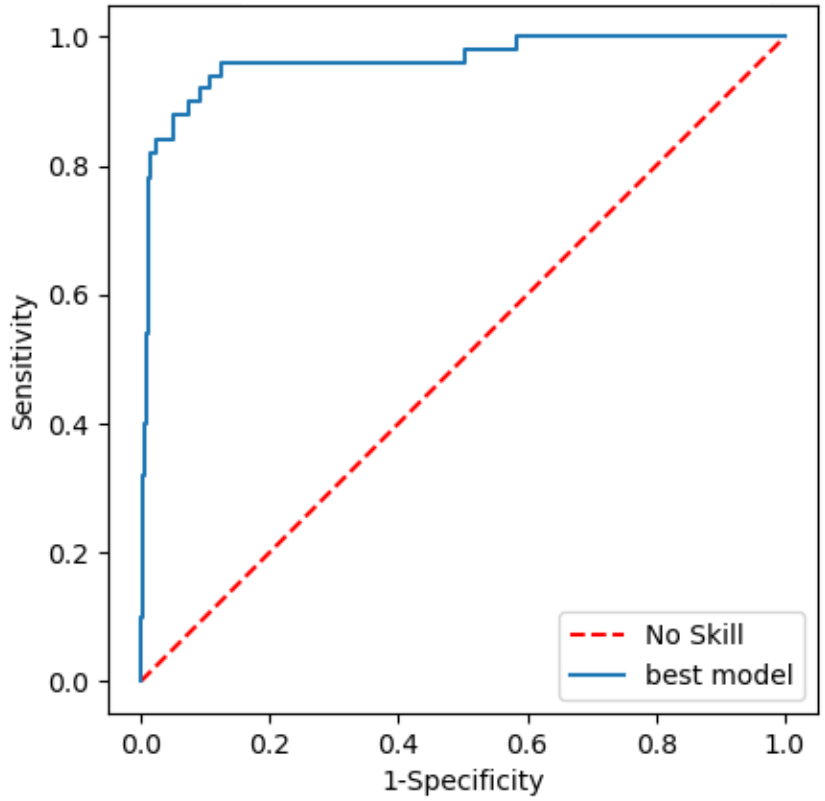
AI has seen both Abs



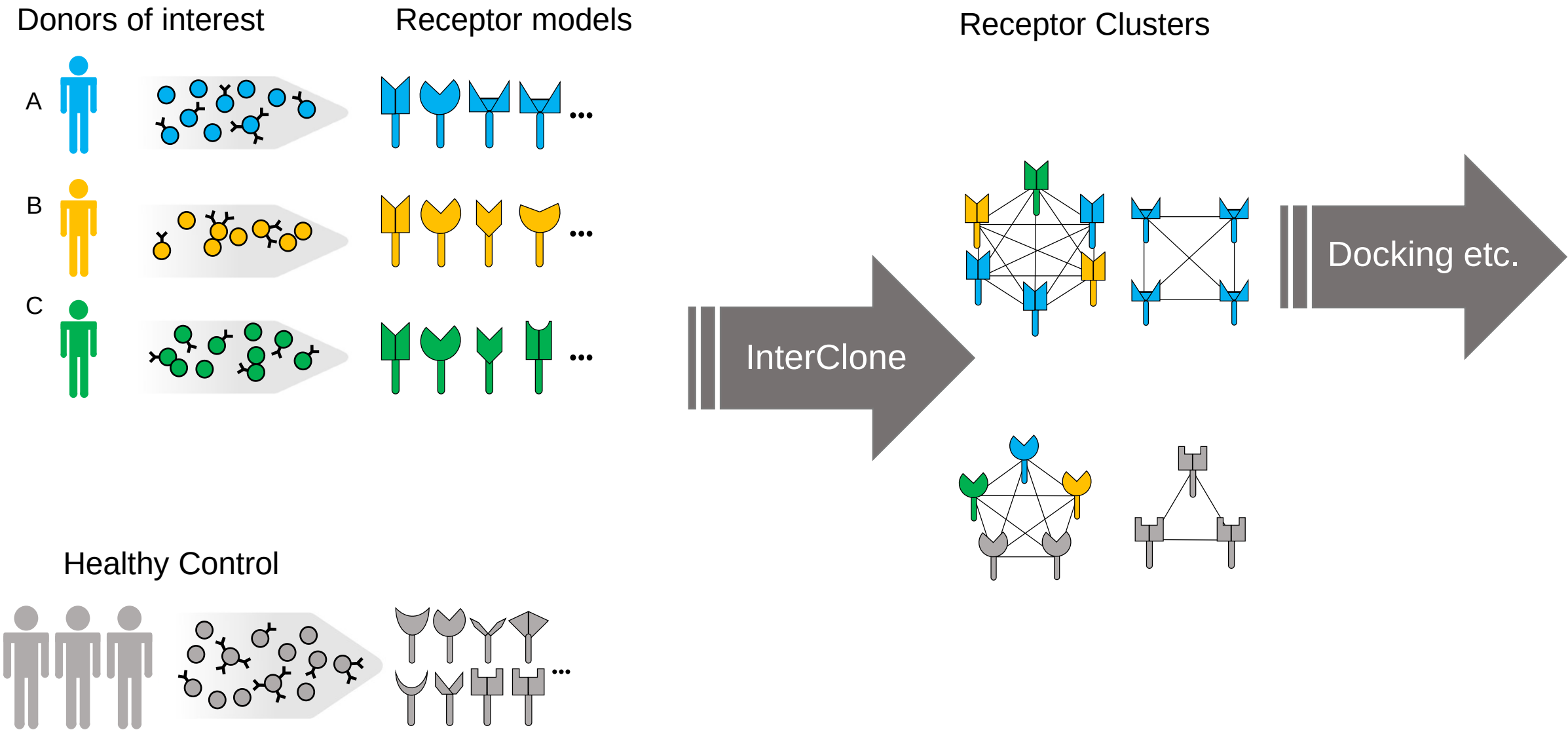
AI has seen one Ab



AI hasn't seen either Ab



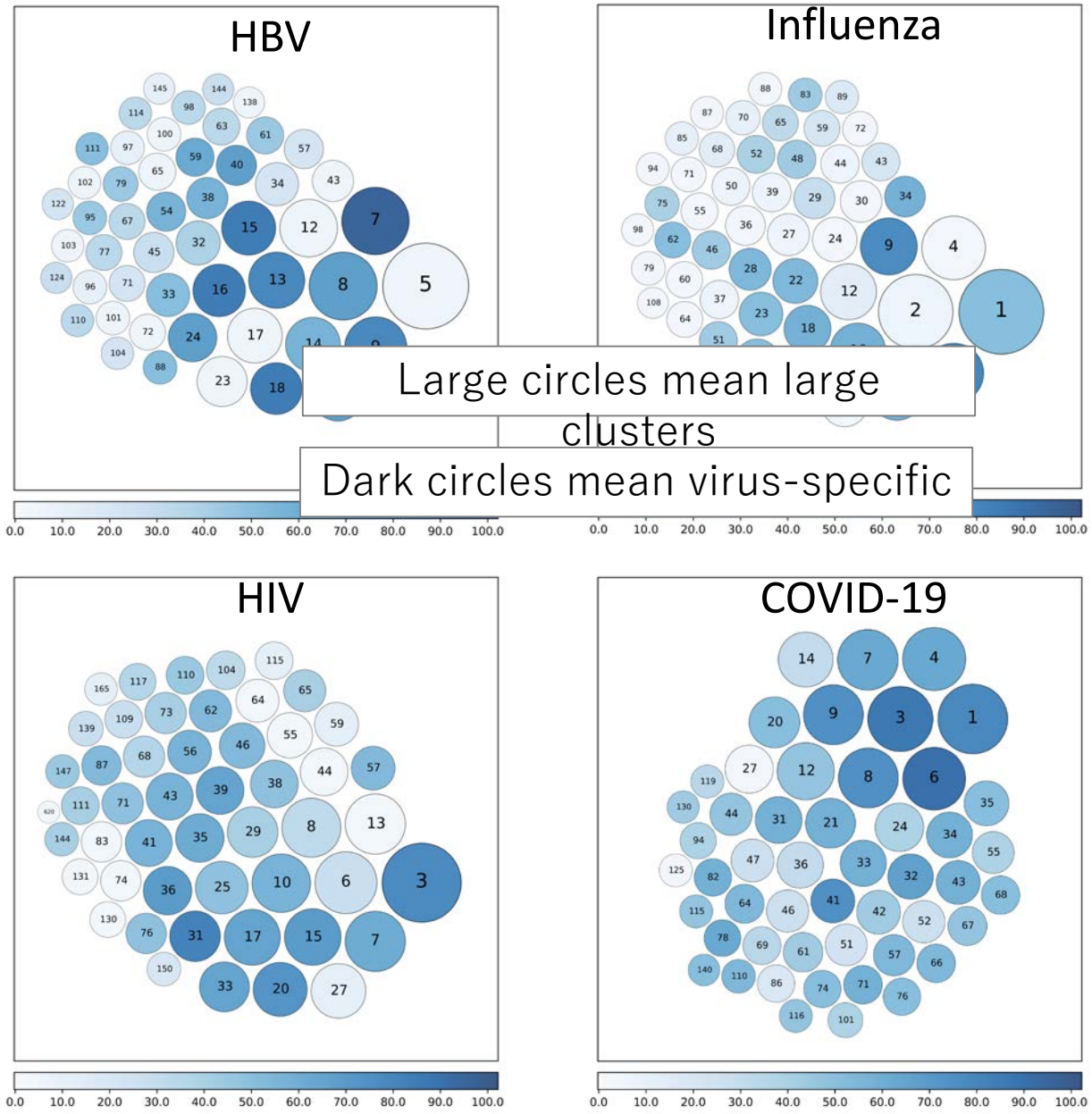
Application: Clustering Ab sequences from different donors



Z. Xu, et al. MSDE 4 (4), (2019); Xu et al (in prep)

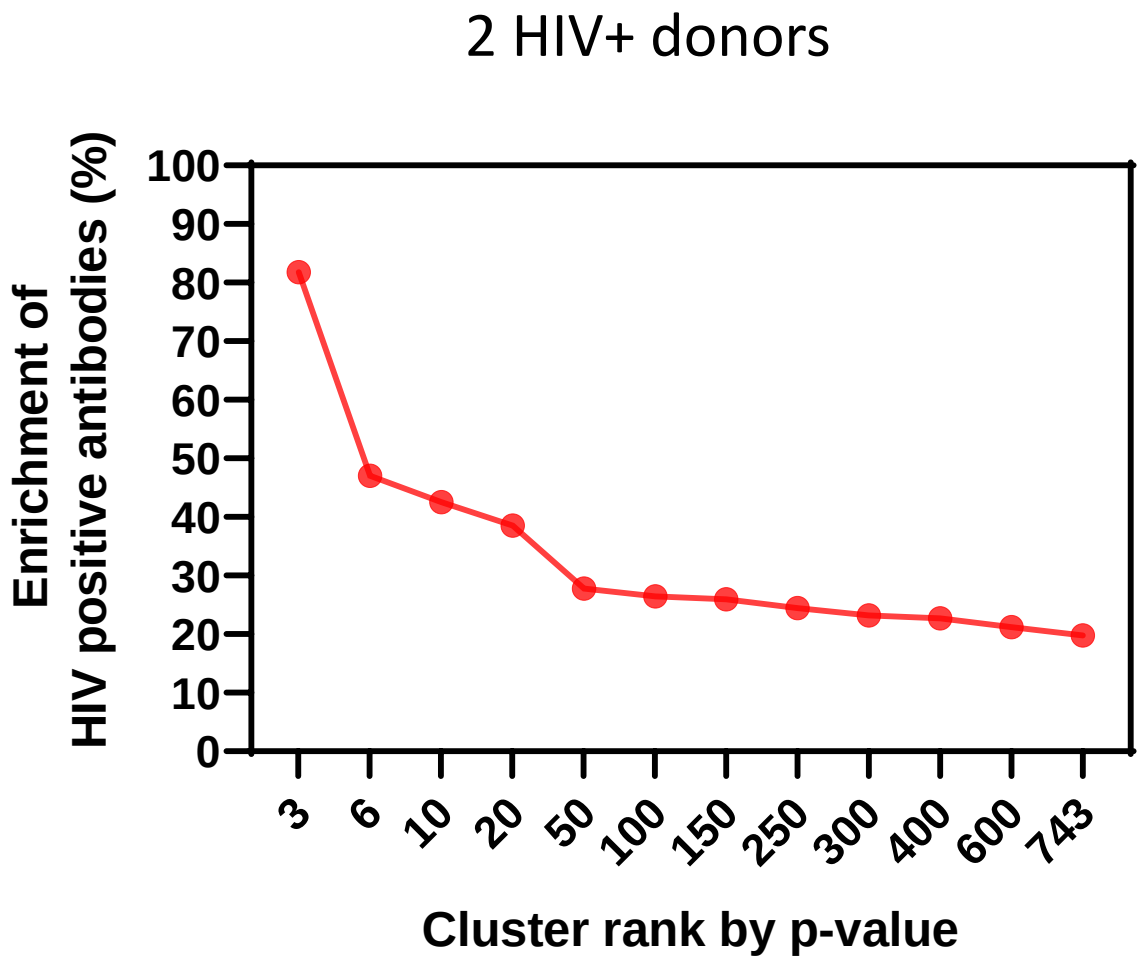
InterClone: Can identify multi-donor antibody clusters for various viruses

Study	Donors	Virus	All sequences	>95% unique	PMID
Galson-2015	9	HBV	2606446	928203	26844287
Galson-2016	9	HBV	2752578	1032531	27312086
Ellebedy-2016	8	Influenza	91994	88734	27525369
Galson-2016	38	Influenza	3429402	786850	27849037
Gupta-2019	3	Influenza	1073957	447555	28179494
Jian-2020	5	Influenza	981706	596052	lead contact
Ellebedy-2020	3	Influenza	26747	25321	32661157
Wu-2011	2	HIV	184672	103342	21835983
Zhu-2012	1	HIV	9442	3661	23024643
Liao-2013	1	HIV	34722	26937	23552890
Zhu-2013	1	HIV	31677	18828	24106303
Schanz-2014	1	HIV	32555	11711	25364977
Wu-2015	1	HIV	354952	212059	25865483
DoriaRose-2015	1	HIV	168151	64601	24590074
Zhou-2015	4	HIV	27201	20800	26004070
Huang-2016	1	HIV	58156	8663	27851912
Setliff-2018	6	HIV	372666	307208	29861170
Waltari-2018	4	HIV	200125	70742	29632541
Armita-2019	7	HIV	992831	819080	31209469
Jia-2020	2	HIV	91026	70716	32315598
Roskin-2020	87	HIV	747891	546069	31959979
Jian-2020	11	Covid19	485034	279924	collaboration
Nielsen-2020	5	Covid19	155297	143813	32941787
Galson-2020	19	Covid19	745666	673499	
Montague-2020	19	Covid19	72148	49050	
Christoph-2020	37	Covid19	264906	248281	32668194
Cervantes-2020	10	Covid19	121728	100335	32669287
Wen-2020	10	Covid19	6207	6092	32377375
Zhang-2020	13	Covid19	8229	7788	32788748

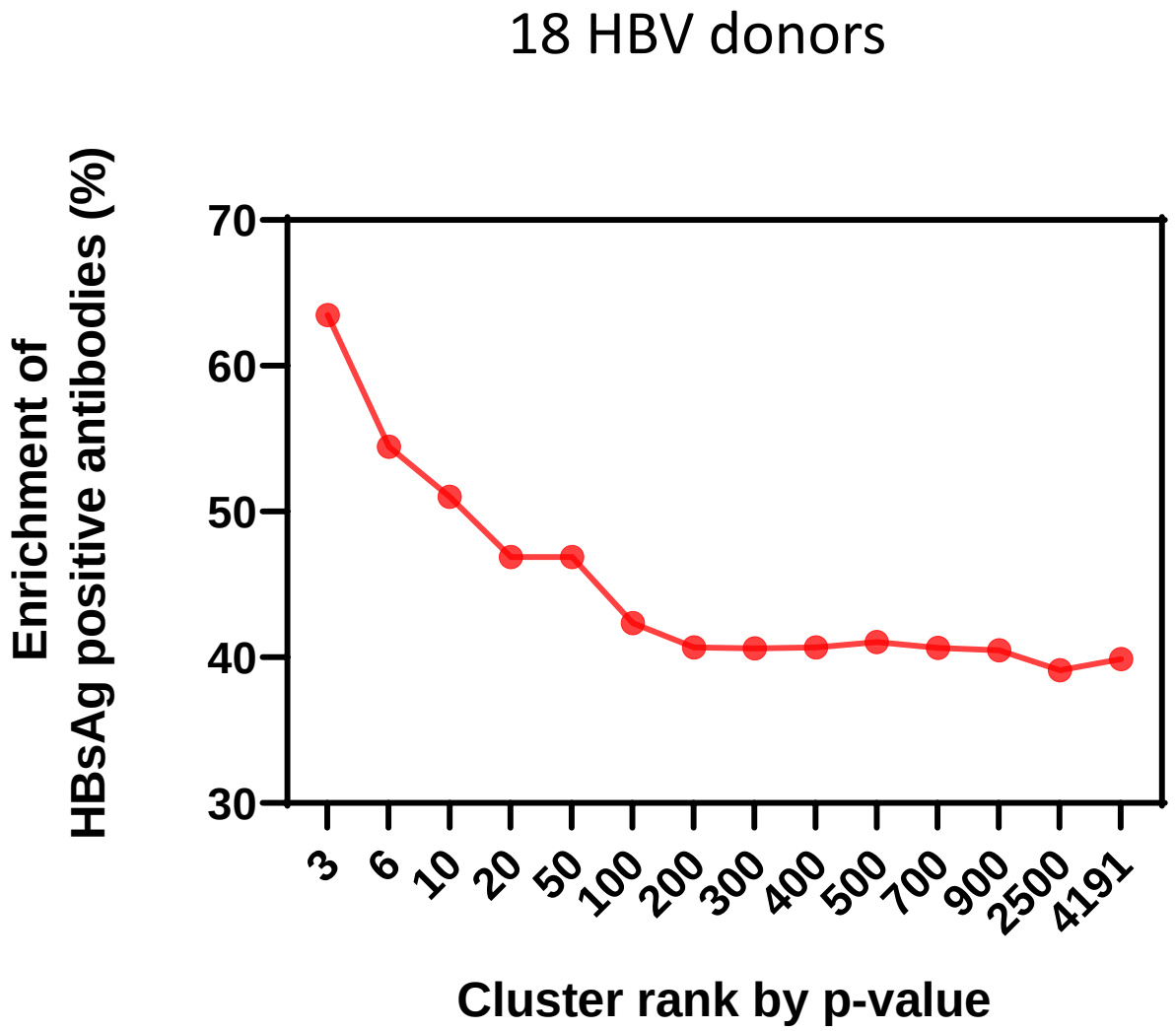


Xu et al (in prep)

Antibodies in HIV-specific clusters are indeed HIV antigen specific

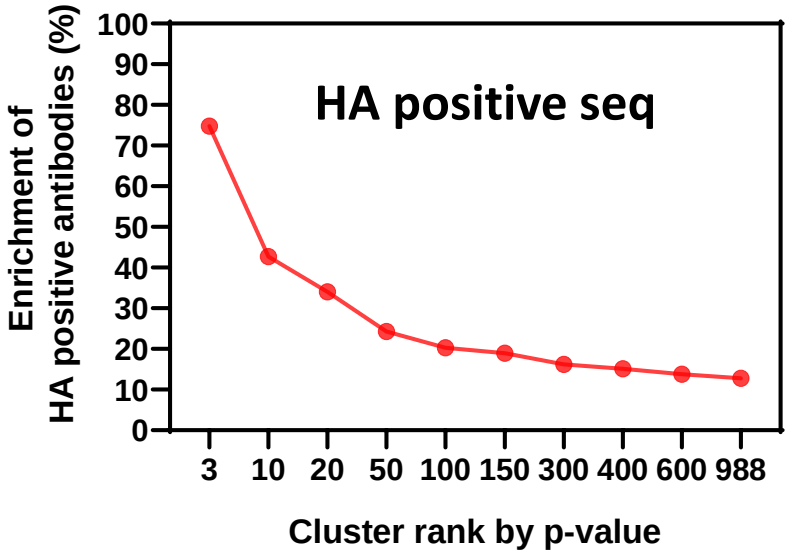


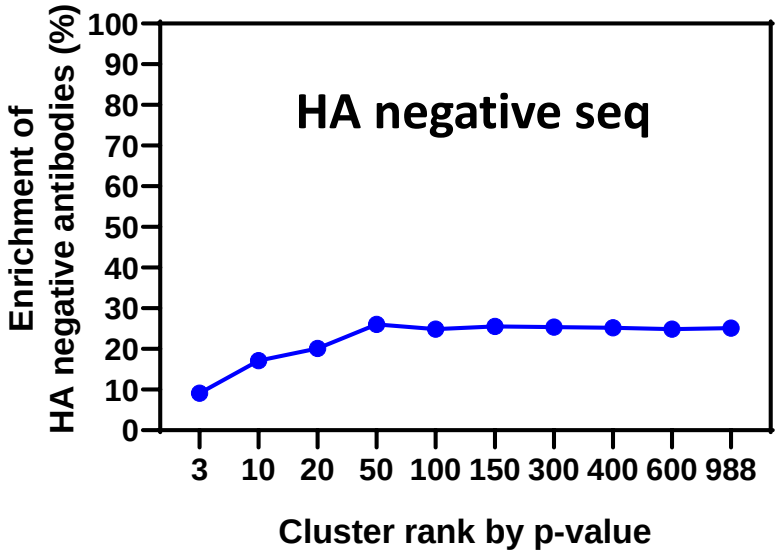
Antibdies in HBV-specific clusters are also HBsAg antigen specific



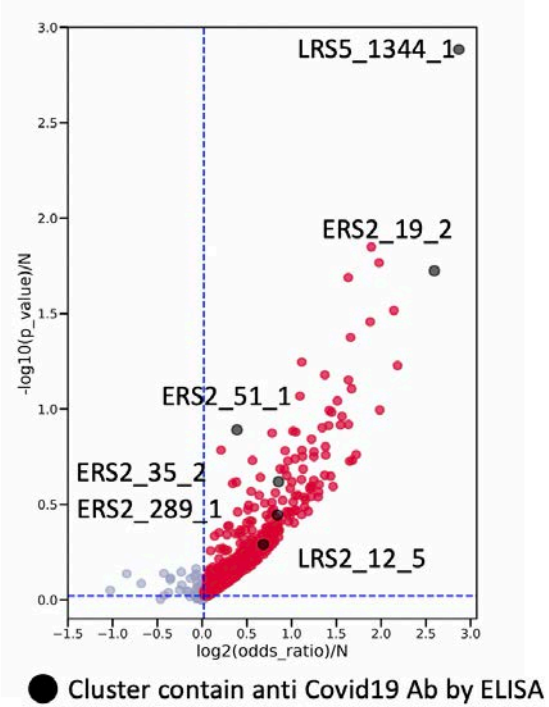
Antibodies in influenza-specific clusters are HA antigen specific

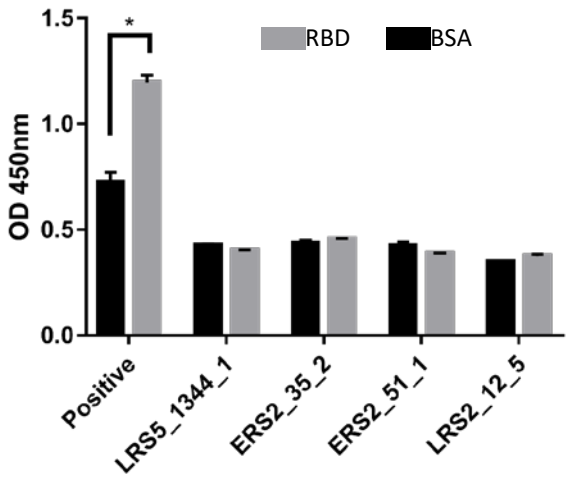
4 Flu donors

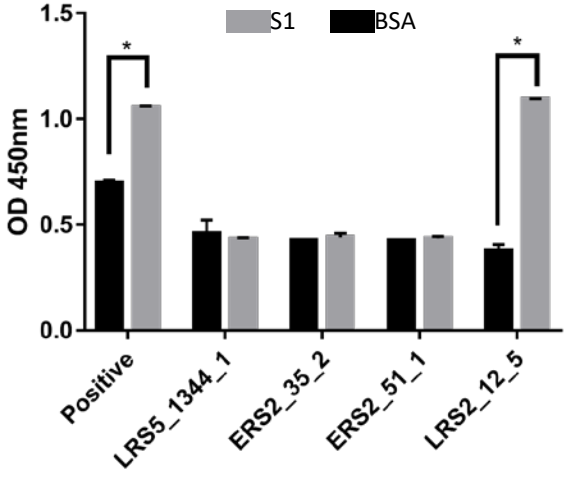


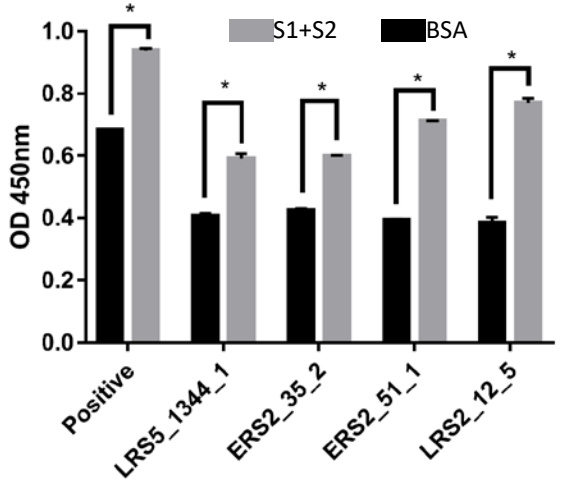


Antibodies in COVID19-specific clusters are Spike-protein antigen specific









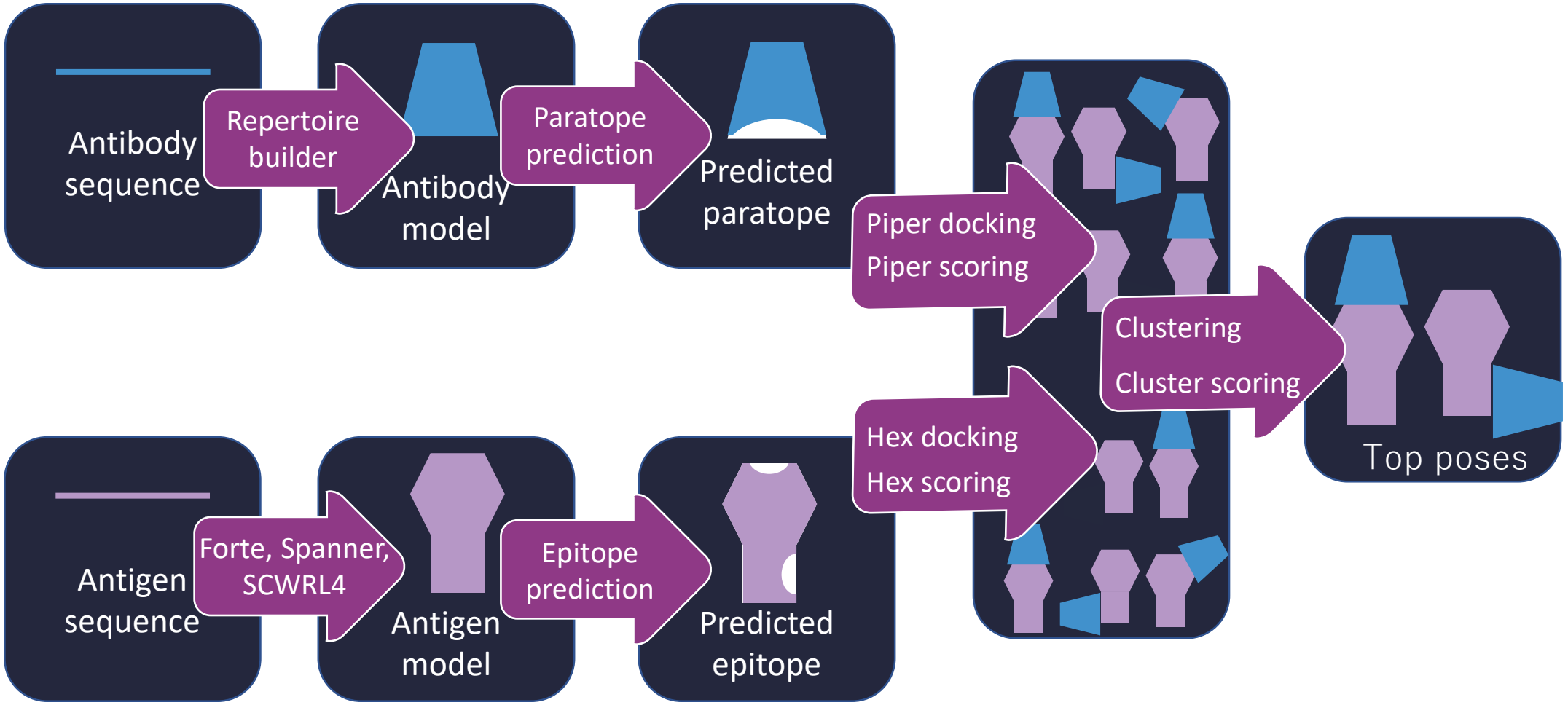
Xu et al (in prep)

Summary of Part 4

- **Antibody models can be clustered using sequence and structural features**
- InterClone was trained on existing 3D structural data
- Antibody clusters enriched in sequences in donors of interest (i.e. virus or vaccine exposed) compared to healthy controls can be identified
- These virus-specific clusters are indeed enriched in antigen-specific antibodies
- InterClone thus appears to perform well on new sequence data
- To our knowledge, InterCone is the only method avilable for antibody clustering
- Web server coming soon!

Part 5: Antibody-antigen docking

Adapt: Docking antibody-antigen pairs from sequenc



Key features of Adapt

- Adapt allows input of sequences and builds structures for docking internally
- Adapt uses epitope and paratope prediction to help docking
- Adapt uses two docking engines (Piper and Hex)
- Adapt uses several customized machine-learning models:
    1. Epitope prediction
    2. Paratope prediction
    3. Piper Scoring
    4. Hex scoring
    5. Scoring clusters of Piper and hex models (poses)

Validating Adapt

567-fold Leave-one-out cross validation

For each Ab-Ag pair:

Remove one pair (test pair)

Train the machine learning models on all other (566) pairs

Test the complete Adapt pipeline on the one test pair (ROC AUC)

Report the average ROC AUC

Holdout test (100 different Ab-Ag pairs)

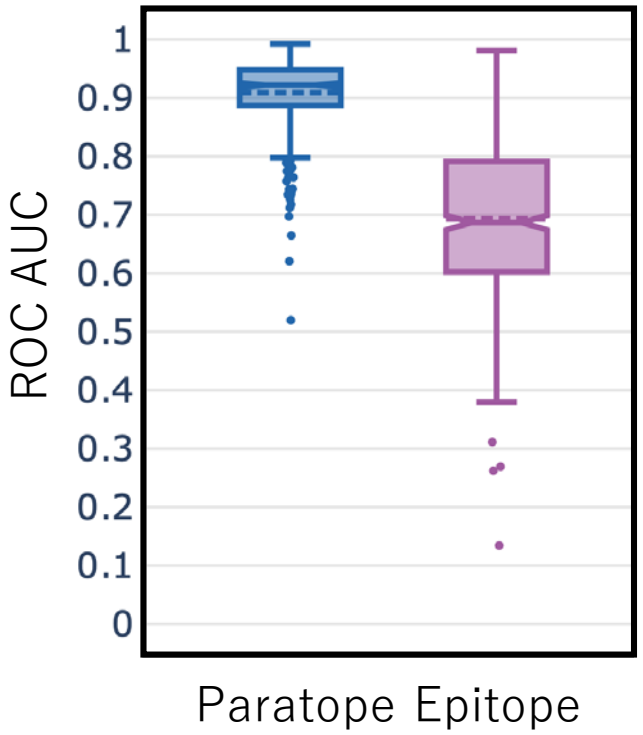
Train the machine learning models on all other (567

Test the complete Adapt pipeline on 100 different p

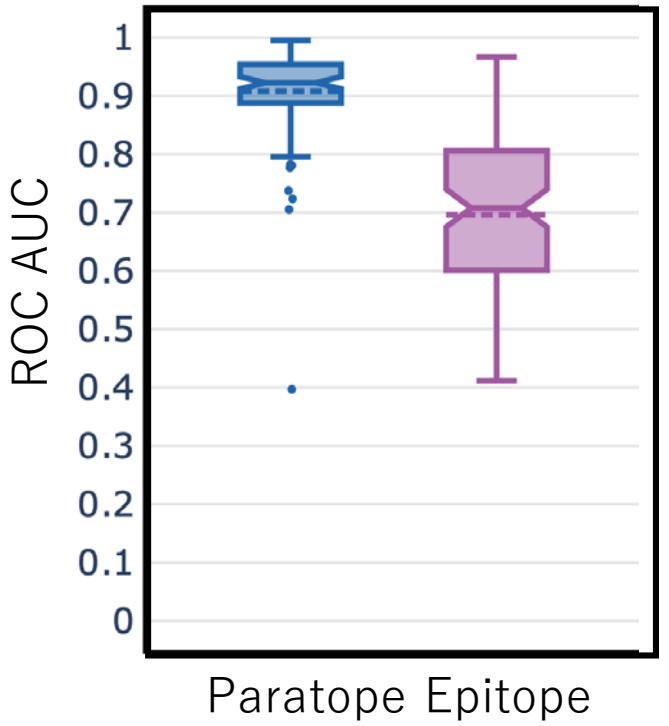
Report the average ROC AUC

Paratope and Epitope Prediction validation

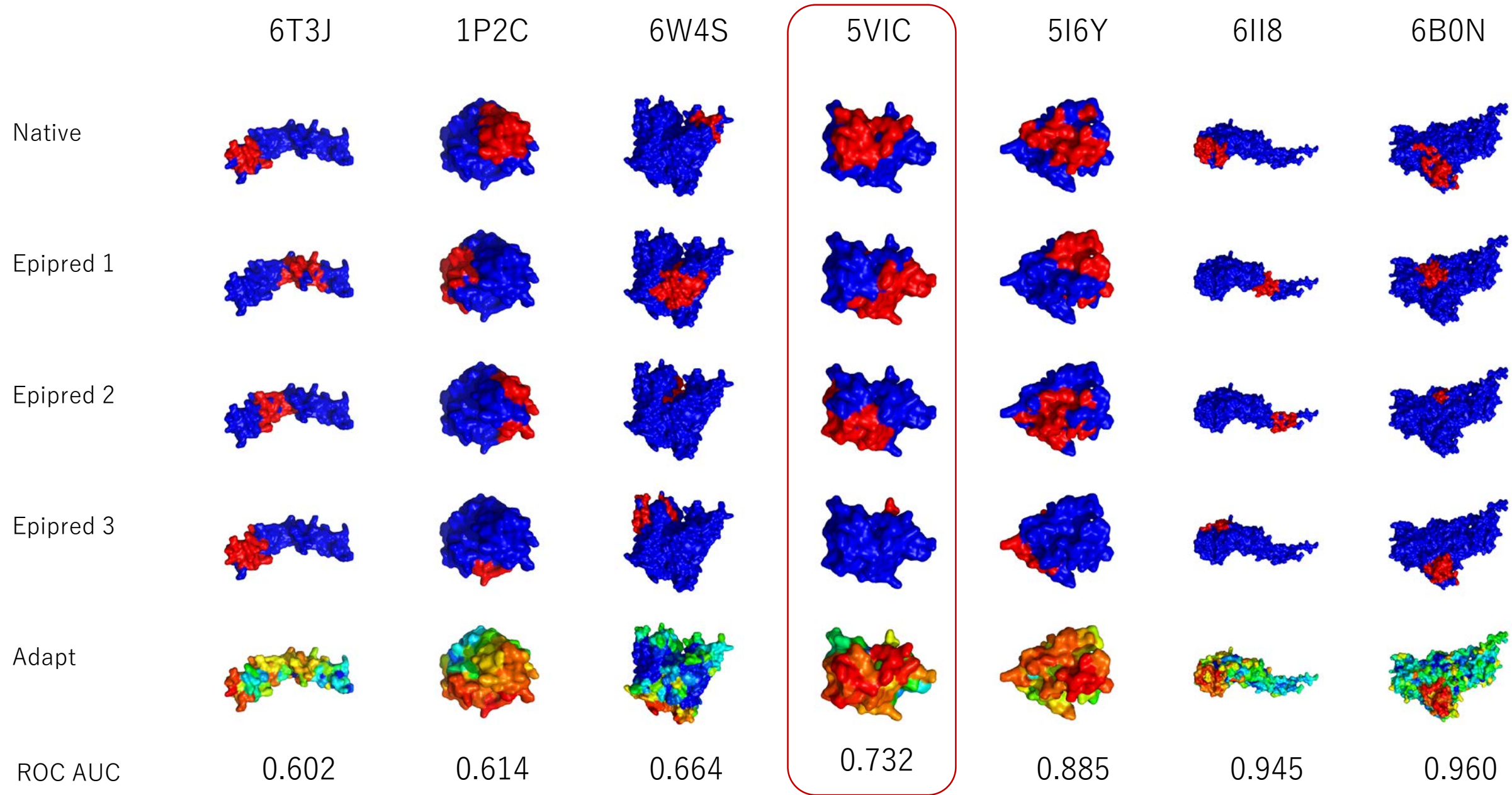
567-fold Leave-one-out cross validation



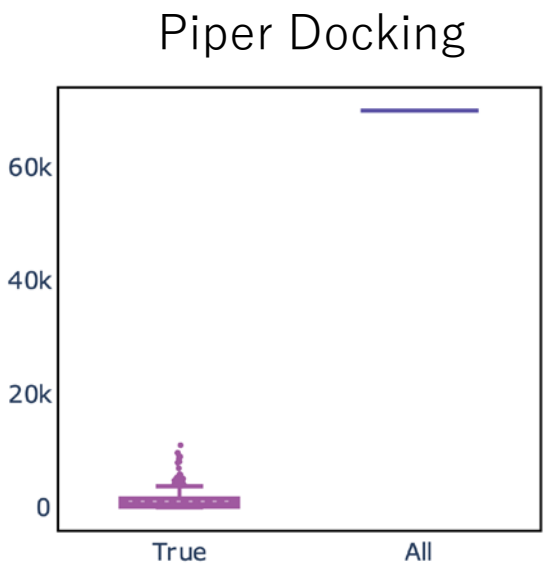
Holdout test (100 different Ab-Ag pairs)

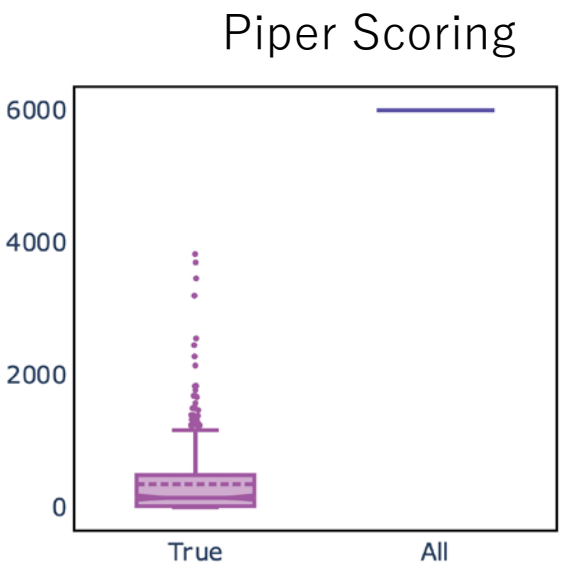


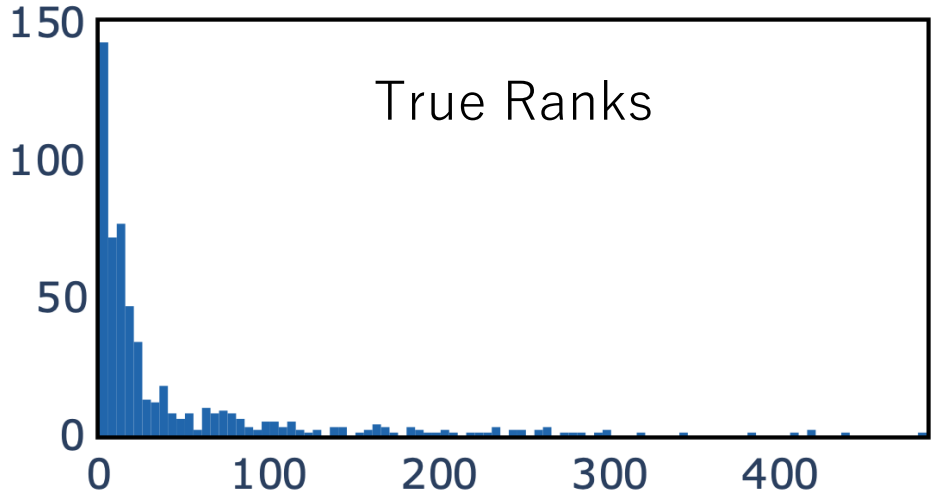
Reperesentative Examples

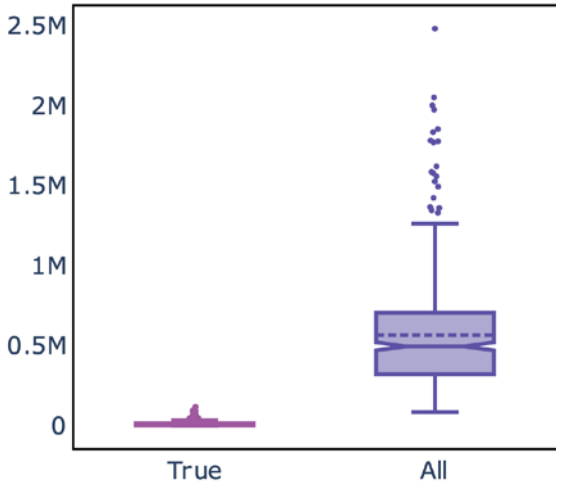


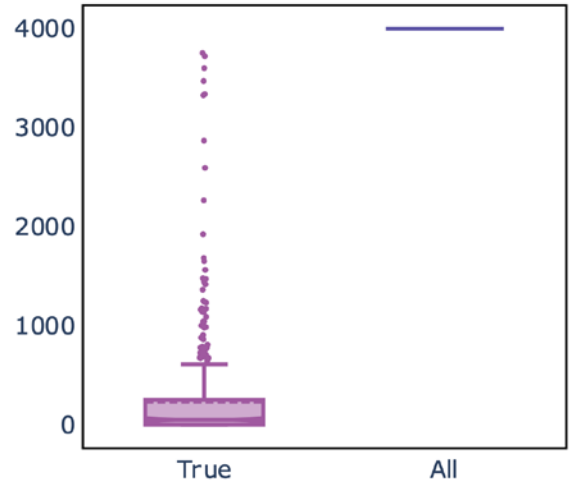
Docking validation: Leave-one-Out Cross Validation

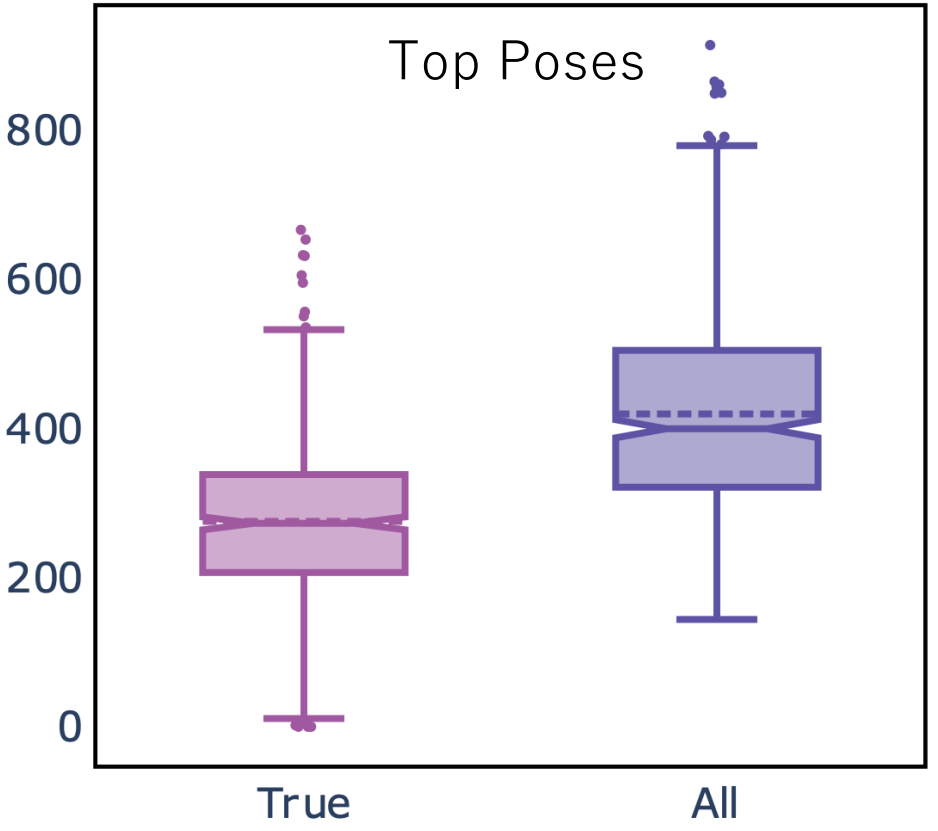




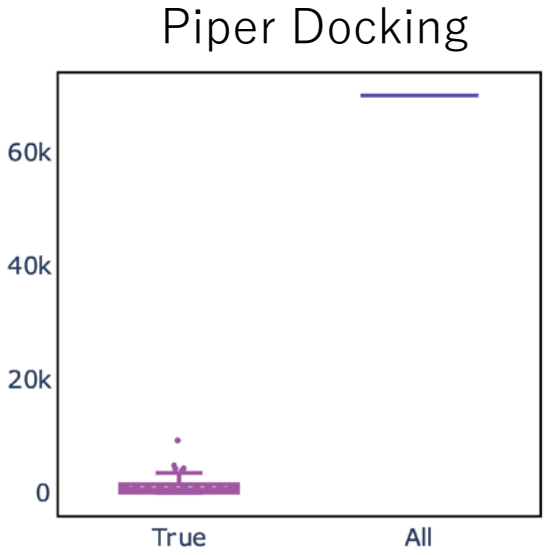


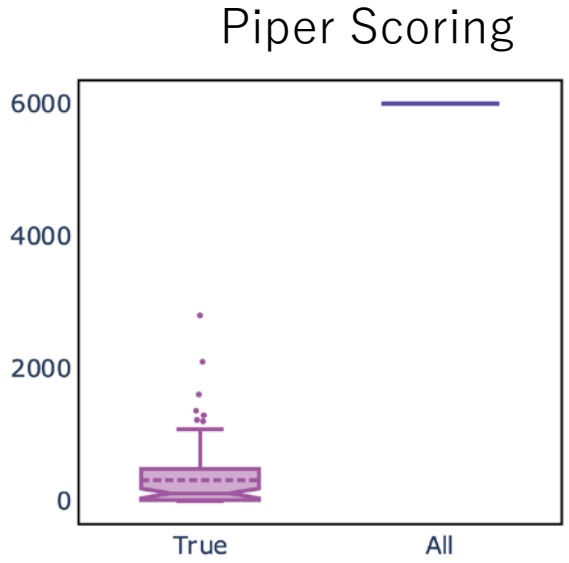


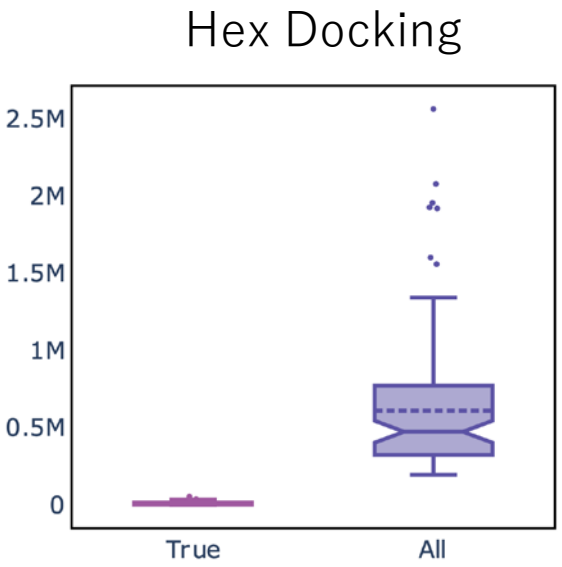


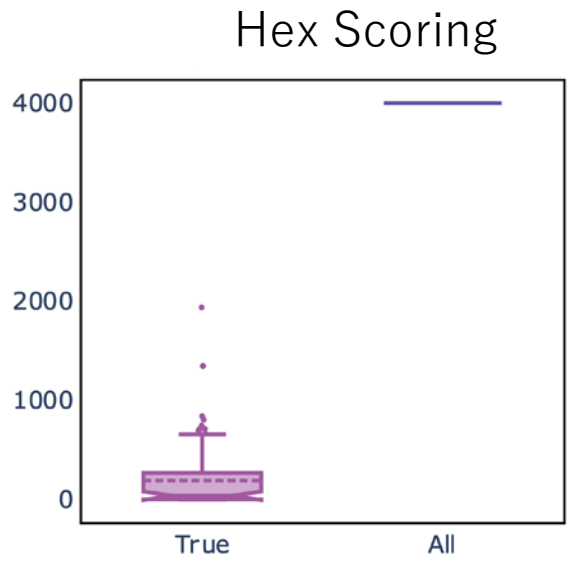


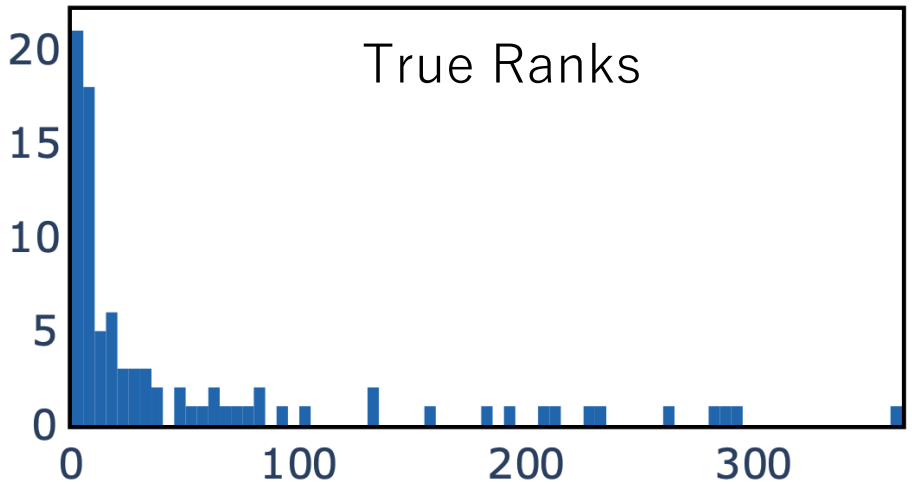
Docking validation: Holdout Set

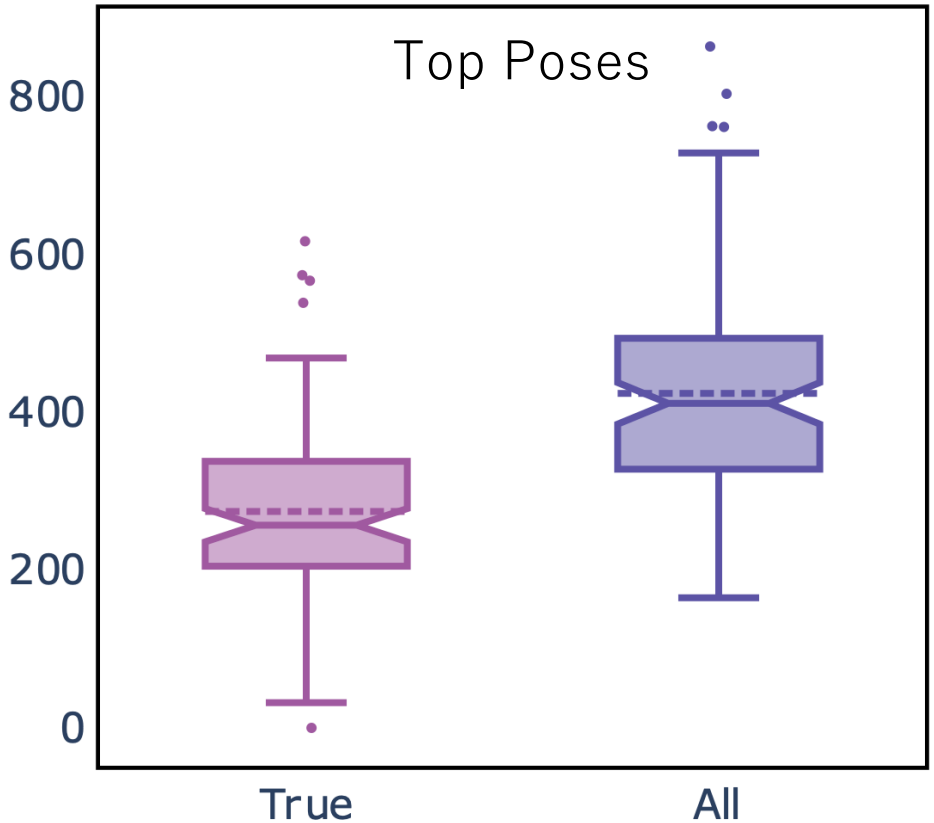












Summary of part 5

- Performance of leave-one-out cross validation was very similar to performance on holdout set (means that we did not over-fit the models)
- The main point of Adapt is to throw away bad models; in this regard, we were successful
- The “success rate” (defined as the ability to produce at least one “True” pose) of using Adapt was 8-9% higher than that of Piper or Hex alone
- Limitations of Adapt: right now it is slow (hours)
- The quality of the final models is “acceptable” but not “high”
- Web server coming soon!