Title: Observed Antibody Space: a resource for data mining next generation sequencing of antibody repertoires.

Aleksandr Kovaltsuk[1], Jinwoo Leem[1], Sebastian Kelm[2], James Snowden[2], Charlotte M. Deane[1,*], Konrad Krawczyk[1,*]

1. University of Oxford, Department of Statistics, Oxford, UK
2. UCB Pharma, Slough, UK

* to whom the correspondence should be addressed,
deane@stats.ox.ac.uk
konrad@proteincontact.org

Abstract. Antibodies are immune system proteins that recognize noxious molecules for elimination. Their sequence diversity and binding versatility have made antibodies the primary class of biopharmaceuticals. Recently it has become possible to query their immense natural diversity using next-generation sequencing of immunoglobulin gene repertoires (Ig-seq). However, Ig-seq outputs are currently fragmented across repositories and tend to be presented as raw nucleotide reads, which means nontrivial effort is required to reuse the data for analysis. To address this issue, we have collected Ig-seq outputs from 53 studies, covering more than half a billion antibody sequences across diverse immune states, organisms and individuals. We have sorted, cleaned, annotated, translated and numbered these sequences and make the data available via our Observed Antibody Space (OAS) resource at antibodymap.org. The data within OAS will be regularly updated with newly released Ig-seq datasets. We believe OAS will facilitate data mining of immune repertoires for improved understanding of the immune system and development of better biotherapeutics.

1. Introduction

Antibodies (or B-cell receptors) are protein products of B-cells and primary actors of adaptive immunity in jawed vertebrates (1). They are highly malleable molecules that can bind to virtually any antigen. An organism holds a great variety of these molecules increasing the probability that an arbitrary antigen can be recognized by an antibody, initiating an immune response (2). Owing to their binding malleability they are the most prominent class of reagents and biotherapeutics (3, 4). Continued successful

33 exploitation of these molecules relies on our ability to discern the functional diversity of antibody
34 repertoires (5–7).
35
36 Next-generation sequencing of immunoglobulin gene repertoires (Ig-seq) has enabled researchers to
37 take snapshots of millions of sequences at a time across individuals, diverse organisms and different
38 immune states (8, 9). The ability to sequence and analyze millions of antibody sequences has the
39 potential to uncover the mechanics of the immune response to any antigen (10, 11) and dysfunctions of
40 the immune system itself (12).
41
42 Many previous studies have addressed the issue of antibody diversity, contributing invaluable evidence
43 to understanding the dynamics of human immune systems (13). Numerous analyses have focused on
44 the frequencies of V(D)J gene usages, which can offer insights into creating biased therapeutic antibody
45 libraries (14–16). Another therapeutic application of antibody repertoire analysis is advancing vaccine
46 design by comparative longitudinal studies of pre- and post-antigen challenge experiments (10, 11, 17–
47 22). Such comparative studies have shown that different individuals can converge on the same antibody
48 sequence against a given vaccine (11, 19). Due to sequencing limitations, these analyses have focused
49 on heavy or light chains separately, whereas one ought to study the paired repertoire to obtain deeper
50 insights of antibody diversity (23).
51
52 Technical advances in sequencing technology have outpaced storage and analysis pipelines (24, 25). This
53 has meant that the outputs of Ig-seq studies are fragmented across repositories making it difficult to
54 perform large-scale data mining of antibody repertoires (25). Metadata such as isotype, age or subject
55 identifiers are not typically standardized, therefore extraction of specific subsets of antibody repertoires
56 for comparative analyses is challenging. Furthermore, the data are typically deposited as raw nucleotide
57 reads. It requires non-trivial *ad hoc* effort to convert such raw reads to amino acid sequences that
58 ultimately dictate the molecular structure and antigen-recognition. Some of these issues are addressed
59 by services that provide Ig-seq-specific data deposition and analysis pipelines such as Immport
60 (http://immport.org) (26, 27), ImmunoseqAnalyzer (http://clients.adaptivebiotech.com/), IReceptor
61 (http://ireceptor.irmacs.sfu.ca/) or VDJServer (http://vdjserver.org) (28). The IReceptor and the
62 VDJServer are the main resources that fall under the umbrella of the organized effort of the Adaptive
63 Immune Receptor Repertoire (AIRR) Community to provide standardized deposition and analysis
64 pipelines for the Ig-seq outputs (24). These services chiefly focus on facilitating bulk deposition of raw

data to perform standardized sequencing analyses. Ultimately, because immunoinformatics is not the chief focus of such services, bulk data download from such websites is limited and converting the raw nucleotide data obtained into a format suitable for analysis still requires non-trivial effort.

To address these issues, we have created the Observed Antibody Space (OAS) resource that allows large-scale data mining of antibody repertoires. We have so far collected the raw outputs of 53 Ig-seq experiments covering over half a billion sequences. We have organized the sequences by metadata such as organism, isotype, B-cell type and source, and the immune status of B-cell donors to facilitate bulk retrieval of specific subsets for comparative analyses. We have converted all of the Ig-seq sequences to amino acids and numbered them using the IMGT scheme. The data is available for querying or bulk download at http://antibodymap.org. We believe that OAS will facilitate data mining antibody repertoires for improved understanding of the dynamics of the immune system and thus better engineering of biotherapeutics.

2. Materials and Methods

A list of study accession codes of publically available Ig-seq datasets were obtained via a literature review. The majority of raw reads were downloaded from the European Nucleotide Archive (ENA) (29) and the National Center for Biotechnology Information (NCBI) websites (30). In a small number of cases, another public Ig-seq repository was specified e.g. (14, 31–33). Metadata were manually extracted from the deposited datasets and arranged in a reproducible format.

The downloaded FASTQ files were processed depending on the sequencing platform. Paired raw Illumina reads were assembled with FLASH (34). The assembled antibody sequences were converted to the FASTA format using FASTX-toolkit (35). As raw reads from Roche 454 are not paired, these FASTQ files were directly converted to the FASTA format with the FASTX-toolkit.

The heavy chain sequences were automatically annotated with isotype information unless such data was given in the corresponding publication. Automatic isotype annotation was performed by aligning the constant heavy domain 1 (CH1) of any given antibody sequence against the IMGT isotype reference (36) of the respective species using the Smith-Waterman algorithm (37). We assigned a score of two for a nucleotide match, and a score of minus one for a nucleotide mismatch or a gap. The IMGT isotype

97 references comprised 21 nucleotide-long fragments of the CH1 domain of the antibody isotypes. To
98 ensure a high confidence of correct isotype identification, we employed a conservative threshold of 30
99 in the Smith-Waterman algorithm scoring function. Sequences whose Smith-Waterman algorithm score
100 was below the threshold for all isotypes were assigned as 'bulk'. The robustness of this protocol was
101 confirmed on the author-annotated Ig-seq datasets (38–40) where it resulted in 99% accurate
102 annotations. Around 1% of the Ig-seq data had a very short (or missing) of CH1 domain sequence. Such
103 sequences were also assigned as 'bulk'.
104
105 IgBlastn (41) was used to convert the FASTA files of antibody nucleotide sequences to amino acids. The
106 amino acid sequences were then numbered with ANARCI (42) using the IMGT scheme (43). ANARCI does
107 not number a sequence if it does not align to a suitable Hidden Markov Model (44). ANARCI therefore
108 ensures that the antibody sequences do not harbor unusual indels or stop codons in the antibody
109 regions, that the V and J genes align to the respective species amino acid IMGT germlines (36), and that
110 the length of CDR-H3 is not greater than 37 residues in human, mouse, rat, rabbit, alpaca and rhesus
111 antibodies. Due to technical limitations of sequencing platforms, certain reads were missing significant
112 portions of the variable region (e.g. portions of CDR1), sequences that did not have all three CDRs were
113 discarded as incomplete. The V and J genes are identified during the ANARCI numbering step.
114
115 Using the protocol above we annotated Ig-seq results of 53 independent studies. In order to streamline
116 updating OAS with new data, we have generated a procedure to automatically identify Ig-seq datasets
117 from raw sequence read archives. We apply our antibody annotation protocol to each raw nucleotide
118 dataset deposited in the NCBI/ENA repositories, if we find more than 10,000 antibody sequences in any
119 given dataset, it is set aside for manual inspection. Manual inspection is still necessary to efficiently
120 assign metadata as these are currently deposited in a non-standardized manner. This procedure allows
121 for automatic identification of new Ig-seq datasets and semi automatically updating of OAS.
122

123 **3. Results**
124
125 We have so far collected raw sequencing outputs from 53 Ig-seq studies. All raw nucleotide reads were
126 converted into amino acids using IgBlastn (41). The full amino acid sequences were then IMGT
127 numbered using ANARCI (42). As well as providing IMGT and gene annotations, ANARCI acts as a broad-
128 brush filter for antibody sequences that are likely to be erroneous (see Materials and Methods).

Applying the same retrieval, amino acid conversion, gene annotation and numbering protocol to all sequences assures the same point of reference across the 53 heterogeneous Ig-seq datasets (45). This protocol produces the full IMGT-numbered sequences together with gene annotations for each of the datasets.

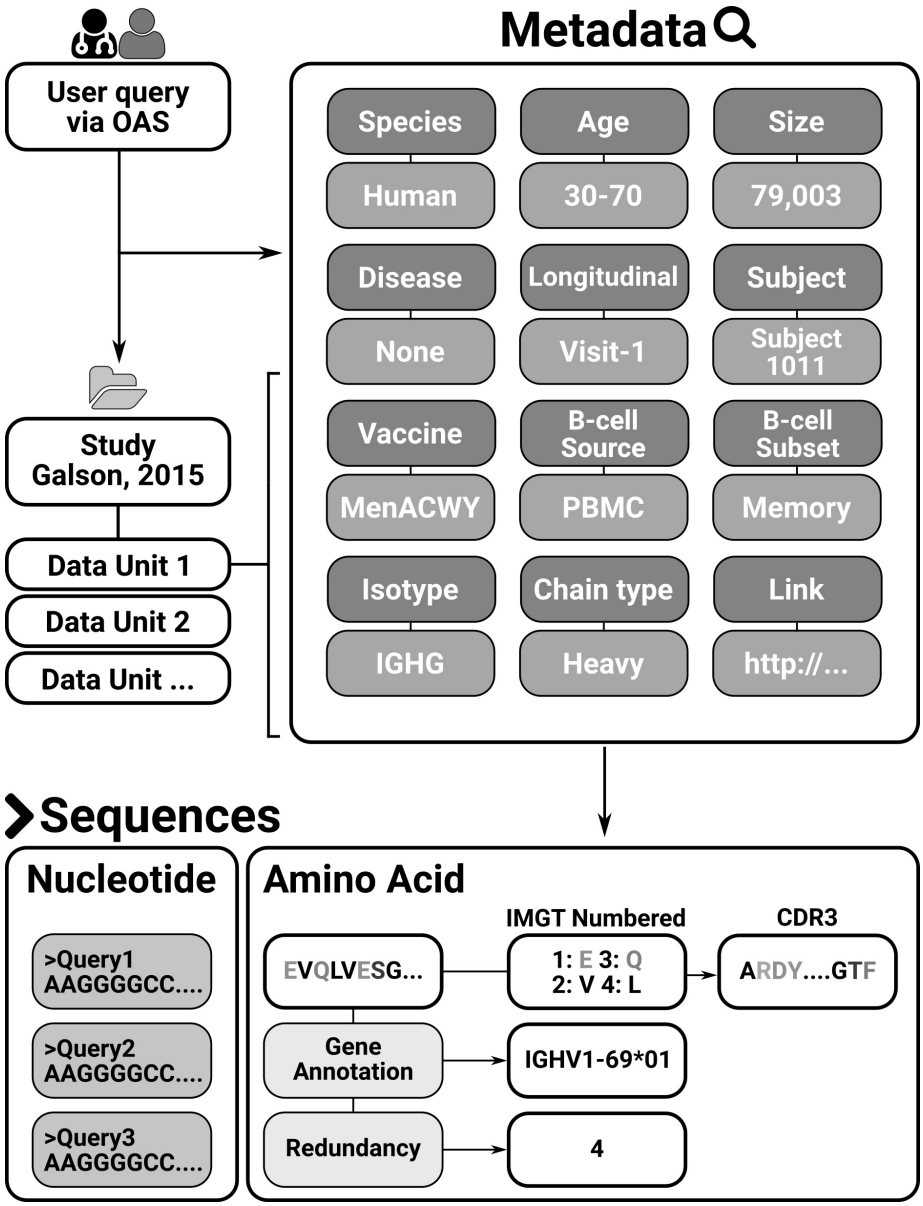

Figure 1. T**he Observed Antibody Space database**. The data from 53 studies is sorted into Data Units. Each Data Unit is a set of antibody sequences that share the same set of meta-data. Each sequence in a Data Unit is further associated with sequence-specific annotations.

The numbered amino acid sequences in each dataset are sorted by metadata e.g. individuals, age, vaccination regime, B-cell type and source *etc*. (Figure 1). Deposition of such metadata is currently not standardized and requires *ad hoc* manual curation for each dataset. In an effort to organize the antibody sequences using such metadata, we have grouped the sequences within each dataset into Data Units. Each Data Unit represents a group of sequences within a given dataset with a unique combination of metadata values. The metadata values are summarized in Table 1.

Metadata name	**Metadata description**
Chain	Heavy/light chain annotation.
Isotype	Identified or deposited isotype information.
Age	Information on the age of the human B-cell donors.
Disease	Indication of whether the donor was sick at the time of B-cell extraction.
Vaccine	Indication if the B-cell donor was purposely immunized prior to B-cell extraction.
B-cell subset	Indication if a particular B-cell subset was sorted for Ig-seq
Species	Organism of the B-cell donor
B-cell source	Which organ/tissue the B-cells were extracted from.
Subject	Indication of a particular B-cell donor the B-cells were sourced from.
Longitudinal	If the study was longitudinal, an indicator of the time point.
Size	Number of non-redundant sequences in the dataset.
Link	Link to the source publication.

Table 1. Metadata descriptors of each Data Unit in OAS. Each Data Unit is uniquely identified by the study and a collection of the metadata values.

157 As of April 29th 2018, 53 Ig-seq studies are included in OAS totaling 608,651,423 sequences
158 (552,824,460 VH and 55,826,963 VL sequences). The majority of these sequences are murine
159 (~50.4%) and human (~47.4%). Twenty-two of the Ig-seq studies interrogate the immune system of
160 diseased individuals, the most common ailment being HIV (13 studies). The database also contains 22 Ig-
161 seq studies of naive antibody gene repertoires (the collection of B-cells from donors who are healthy
162 and not purposefully vaccinated). The main source of B-cells in the OAS database is peripheral blood
163 (~231m of sequences) followed by spleen/splenocytes (~198m) and bone marrow (~124m). The
164 database holds isotype information for each individual heavy sequence and the two most common
165 isotypes are IgM (~312m) and IgG (~139m). For ~65m sequences we were not able to assign isotypes
166 with high confidence. The median total redundant size of the Ig-seq studies in the OSA database is
167 2,006,196 sequences, while the largest Ig-seq study was that by Greiff et al., (246,449,189 redundant
168 sequences) (14). Detailed statistics on each dataset are given in Table 2. All the data may be bulk
169 downloaded or individual Data Units queried, at http://antibodymap.org.
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187

Study	Species	Disease	Vaccine	B-cell source	B-cell subset	Total ANARCI parsed sequences
Banerjee et al., (46)	Rabbit	None	HIV	PBMC	Unsorted	4,334,088 (2,926,727)
Bashford-Rogers et al., (47)	Human	CLL/None	None	PBMC	Unsorted	129,013 (86,166)
Bhiman et al., (48)	Human	HIV	None	PBMC	Unsorted	785,751 (187,067)
Bonsignori et al., (49)	Human	HIV/None	None	PBMC	Memory/ Unsorted	210,377 (57,374)
Collins et al., (50)	Mouse	None	None	Splenocytes	Unsorted	812,439 (194,752)
Corcoran et al., (51)	Human/ Mouse/ Rhesus	None	None	PBMC	Unsorted	5,307,880 (2,840,877)
Cui et al., (52)	Mouse	None	NP-CGG/None	Splenocytes	Memory	5,513,816 (935,646)
Doria-Rose et al., (13)	Human	HIV	None	PBMC	Unsorted	2,164,901 (549,544)
Fisher et al.,(53)	Mouse	None	Plasmodium	Spleen	Unsorted	175,015 (113,594)
Galson et al., (38)	Human	None	Hepatitis-B	PBMC	Unsorted/ Plasma cells/ HepB-specific	21,755,739 (10,442,291)
Galson et al., (39)	Human	None	Hepatitis-B	PBMC	Unsorted/ Plasma cells/ HepB-specific	26,687,394 (14,343,236)
Galson et al., (21)	Human	None	Meningitis	PBMC	Naïve/ Plasma cells/ Memory/ Marginal zone	7,918,197 (3,282,907)
Galson et al., (17)	Human	None	Flu	PBMC	Plasma cells	13,685,210 (5,065,786)
Greiff et al., (40)	Mouse	None	NP-CGG	Bone marrow/ Spleen	Plasma cells/ Plasmablasts	7,955,739 (2,891,649)
Greiff et al., (33)	Mouse	None	NP-CGG	Spleen	ASCs/Plasma cells/ Naive	788,787 (523,716)
Greiff et al., (14)	Mouse	None	OVA/Hepatitis-B/ NP-HEL/None	Spleen/ Bone marrow	Plasma cells/ Pre-B-cells/ Naive	246,449,189 (129,417,638)
Gupta et al., (54)	Human	None	Flu/Hepatitis-A/ Hepatitis-B	PBMC	Unsorted	25,134,322 (9,966,175)
Halliley et al., (55)	Human	None	Flu/Tetanus	Bone marrow	Plasma cells	2,348,164 (1,208,616)
Huang et al., (56)	Human	HIV	None	PBMC	Memory	11,693,783 (5,701,433)
Jiang et al., (57)	Human	None	Flu	PBMC	Naïve/ Plasmablasts	3,199,271 (1,809,306)
Joyce et al., (58)	Human	None	None	PBMC	Unsorted	2,747,688 (1,463,421)
Khan et al., (59)	Mouse	None	OVA	Spleen	Unsorted	24,175,033 (7,113,411)
Levin et al., (60)	Human	Allergy	None	PBMC/ Nasal biopsy	Unsorted	528,173 (370,465)

Levin et al., (61)	Human	Allergy	None	PBMC/ Bone marrow	Unsorted	29,643,305 (9,557,586)
Li et al., (62)	Camel	None	None	PBMC	Unsorted	1,152,359 (1,127,651)
Liao et al., (63)	Human	HIV	None	PBMC	Unsorted	1,420,314 (619,492)
Lindner et al., (64)	Mouse	None	E.Coli/ Clostridia/ Lactobacillus	Biopsy of small intestine	Unsorted	1,686,350 (544,061)
Meng et al., (65)	Human	CMV/ EBV/None	None	PBMC/Lung/ Spleen/Bone marrow/Colon/ Jejunum/Lymph node/Ileum	Unsorted	45,576,606 (21,738,501)
Menzel et al., (66)	Mouse	None	NP-CGG	Spleen/Bone marrow	ASCs	14,355,151 (6,058,480)
Mroczek et al., (67)	Human	None	None	PBMC	Immature/ Transitional/ Mature/ Plasmacytes/ Memory	104,154 (85,525)
Ota et al., (68)	Mouse	None	None	Spleen/Lymph	Unsorted	21,505 (9,619)
Palanichamy et al., (69)	Human	MS	None	CSF/PBMC	Unsorted	776,895 (292,801)
Parameswaran et al., (11)	Human	Dengue/None / Non-dengue febrile illness/	None	PBMC	Unsorted	26,584 (23,606)
Prohaska et al., (70)	Mouse	None	None	Spleen/ Peritoneum	B-1/ B-2/ Marginal Zone/ Follicular	336,723 (198,983)
Rettig et al., (32)	Mouse	None	None	Spleen/ Splenocytes	Unsorted	41,908 (24,908)
Rubelt et al., (71)	Human	None	None	PBMC	Naïve/Memory	2,320,947 (1,719,507)
Schanz et al., (31)	Human	HIV/None	None	PBMC	Unsorted	12,734,958 (5,412,549)
Zhu et al., (72)	Human	HIV	None	PBMC	Unsorted	1,962,643 (532,350)
Stern et al., (73)	Human	MS	None	Cervical lymph node/ White matter lesion/ Pia mater/ Choroid plexus/ Cortex/ Spleen	Unsorted	8,550,247 (3,321,530)
Sundling et al., (74)	Rhesus	None	HIV	PBMC	Unsorted	40,960 (26,298)
Tipton et al., (75)	Human	SLE/None	Flu/Tetanus	PBMC	Unsorted	28,204,742 (13,301,396)
Turchaninova et al., (76)	Human	None	None	PBMC	Memory/Plasma cells/Naive	183,949 (176,441)
Vander Heiden et al., (77)	Human	MG/None	None	PBMC	Memory/Naïve/ Unsorted	13,939,166 (5,170,299)
VanDuijn et al., (78)	Rat	None	DNP/HuD	Splenocytes	Unsorted	6,359,396 (4,234,597)

Vergani et al., (79)	Human	None	None	PBMC	Unsorted	14,161,949 (5,987,086)
Wasemann et al., (80)	Mouse	None	NP-CGG	Lamina propria/ Bone marrow/ Spleen	Unsorted	146,370 (40,132)
Wu et al., (81)	Human	HIV	None	PBMC	Unsorted	394,144 (198,468)
Wu et al., (82)	Human	HIV	None	PBMC	Unsorted	5,545,910 (1,370,109)
Wu et al., (83)	Human	Allergy/ None	None	PBMC/ Nasal biopsy	Unsorted	35,034 (23,923)
Zhou et al., (22)	Human	HIV	None	PBMC	Unsorted	1,541,645 (458,227)
Zhou et al., (84)	Human	HIV	None	PBMC	Unsorted	722,112 (291,670)
Zhu et al., (85)	Human	HIV	None	PBMC	Unsorted	874,930 (174,435)
Zhu et al., (86)	Human	HIV	None	PBMC	Unsorted	1,290,499 (699,828)

188 Table 2. **Summary of Ig-seq studies that are incorporated into the Observed Antibody Space database.**
189 The datasets are organized into studies related to a given Ig-seq experiment. Each study in the OAS
190 database is subdivided into Data Units. Each Data Unit is a collection of IMGT-numbered amino acid
191 sequences uniquely identified by the metadata descriptors given in Table 1, five of which (species,
192 disease, vaccine, B-cell source and B-cell type) are given in this Table. The 'total ANARCI parsed
193 sequences' field indicates the total number of redundant sequences in our database, with the non-
194 redundant numbers in parentheses. Abbreviations: PBMC, peripheral blood mononuclear cell; CLL,
195 chronic lymphocytic leukemia; NP-CGG, chicken gamma globulin; ASC, antibody secreting cell; OVA,
196 ovalbumin; NP-HEL, hen egg lysozyme; CSF, cerebrospinal fluid; MS, multiple sclerosis; SLE, systemic
197 lupus erythematosus; MG, myasthenia gravis; DNP, dinitrophenyl; HuD, paraneoplastic
198 encephalomyelitis antigen.
199
200
201
202
203
204
205
206
207
208

209
210

211 **4. Discussion**
212

213 Here, we describe the Observed Antibody Space (OAS) database, a unified repository to facilitate large-
214 scale data mining of antibody repertoires in both their amino acid and nucleotide forms. Absence of well
215 established repositories in Ig-seq deposition space required us to perform a combination of literature
216 search and manual curation of the datasets in order to organize the data into OAS. The current lack of
217 widely-adopted deposition standards hampers automatic updating of OAS, as datasets where we find
218 large number of antibodies still require manual curation to perform metadata annotation correctly.
219 Hopefully, efforts such as that by the AIRR community will result in standardization of Ig-seq outputs
220 and will further streamline deposition procedures facilitating large-scale data mining of antibody
221 repertoires (24). Devising unified antibody repertoire repositories is challenging due to both the size of
222 the datasets as well as the diverse data descriptors and analytical pipelines desired by bioinformaticians,
223 wetlab scientists and clinicians (33).
224

225 OAS is the first organized collection of a large body of Ig-seq outputs. In order to allow comparative
226 bioinformatics analyses across different subsets of antibody repertoires, we have annotated the
227 datasets by commonly used metadata descriptions such as organism, isotype, B-cell type and source,
228 and the immune state of B-cell donors. To facilitate research about particular antibody sequences or
229 regions, we make full IMGT-numbered high-quality amino acid sequences available together with gene
230 annotations, as well as raw nucleotide data.
231

232 This data should aid in-depth comparative analyses across different studies to discern the commonalities
233 observed between independent samples as well as directing Ig-seq experiments on not yet interrogated
234 antibody repertoires. Revealing shared preferences can be invaluable in identifying the portions of the
235 theoretically allowed antibody space that are strategically used to start immune responses (6).
236 Furthermore, such comparative studies can offer a way of deconvoluting the various degrees of freedom
237 of immune repertoires such as differences between diversity of isotypes (67) or organisms (87). Charting
238 the differences between repertoires of human/mouse is of particular interest for engineering better
239 humanized biotherapeutics (88).
240

241 Beyond identifying broad commonalities across repertoires, data mining Ig-seq outputs provides novel
242 avenues for designing better antibody-based therapeutics. The plethora of currently available Ig-seq
243 data offers a glimpse at a set of sequences that should be able to fold and function in an organism.
244 Aligning therapeutic candidates to sequences in Ig-seq repertoires can reveal mutational choices that
245 might be naturally acceptable hence providing shortcuts for antibody engineering such as humanization
246 (89). Furthermore, contrasting the naturally observed antibodies with therapeutic ones can offer insight
247 as to the naturally favored biophysical properties of these molecules (4). All such future applications rely
248 on the availability of well-structured datasets that can offer a unified point of reference for
249 bioinformatics analyses. We hope that OAS will aid data mining antibody repertoires, help identify
250 strategic preferences of our immune systems and will ultimately improve how we engineer antibodies
251 into better therapeutics.
252

253 **Acknowledgment:**
254 We would like to thank all members of Oxford Protein Informatics Group for testing our OAS resource.
255 In particular, we are grateful to Garret M. Morris and Matthew Raybould for their comments, which
256 significantly improved the quality of our work.
257

258 **Funding:**
259 This work was supported by funding from Biotechnology and Biological Sciences Research Council
260 (BBSRC) [BB/M011224/1] and UCB Pharma Ltd awarded to AK.

261 **References:**
262
263 1. Kindt, T. J., R. A. Goldsby, B. A. Osborne, and J. Kuby. 2007. *Kuby immunology*,. WH Freeman &
264 Company.
265 2. Glanville, J., W. Zhai, J. Berka, D. Telman, G. Huerta, G. R. Mehta, I. Ni, L. Mei, P. D. Sundar, G. M. R.
266 Day, D. Cox, A. Rajpal, and J. Pons. 2009. Precise determination of the diversity of a combinatorial
267 antibody library gives insight into the human immunoglobulin repertoire. *Proc. Natl. Acad. Sci. U. S. A.*
268 106: 20216–20221.
269 3. Kaplon, H., and J. M. Reichert. 2018. Antibodies to watch in 2018. *MAbs* 1–21.
270 4. Jain, T., T. Sun, S. Durand, A. Hall, N. R. Houston, J. H. Nett, B. Sharkey, B. Bobrowicz, I. Caffry, Y. Yu, Y.
271 Cao, H. Lynaugh, M. Brown, H. Baruah, L. T. Gray, E. M. Krauland, Y. Xu, M. Vásquez, and K. D. Wittrup.

272 2017. Biophysical properties of the clinical-stage antibody landscape. *Proc. Natl. Acad. Sci.* 114.
273 5. Miho, E., A. Yermanos, C. R. Weber, C. T. Berger, S. T. Reddy, and V. Greiff. 2018. Computational
274 Strategies for Dissecting the High-Dimensional Complexity of Adaptive Immune Repertoires. *Front.*
275 *Immunol.* 9.
276 6. Greiff, V., C. R. Weber, J. Palme, U. Bodenhofer, E. Miho, U. Menzel, and S. T. Reddy. 2017. Learning
277 the High-Dimensional Immunogenomic Features That Predict Public and Private Antibody Repertoires. *J.*
278 *Immunol.* ji1700594.
279 7. Kovaltsuk, A., K. Krawczyk, J. D. Galson, D. F. Kelly, C. M. Deane, and J. Trück. 2017. How B-Cell
280 Receptor Repertoire Sequencing Can Be Enriched with Structural Antibody Data. *Front. Immunol.* 8:
281 1753.
282 8. Georgiou, G., G. C. Ippolito, J. Beausang, C. E. Busse, H. Wardemann, and S. R. Quake. 2014. The
283 promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* 32:
284 158–68.
285 9. Friedensohn, S., T. A. Khan, and S. T. Reddy. 2017. Advanced Methodologies in High-Throughput
286 Sequencing of Immune Repertoires. *Trends Biotechnol.* 35: 203–214.
287 10. Galson, J., A. J. Pollard, J. Trück, and D. F. Kelly. 2014. Studying the antibody repertoire after
288 vaccination: Practical applications. *Trends Immunol.* 35: 319–331.
289 11. Parameswaran, P., Y. Liu, K. M. Roskin, K. K. L. Jackson, V. P. Dixit, J. Y. Lee, K. L. Artiles, S. Zompi, M.
290 J. Vargas, B. B. Simen, B. Hanczaruk, K. R. McGowan, M. A. Tariq, N. Pourmand, D. Koller, A. Balmaseda,
291 S. D. Boyd, E. Harris, and A. Z. Fire. 2013. Convergent antibody signatures in human dengue. *Cell Host*
292 *Microbe* 13: 691–700.
293 12. Ghraichy, M., J. D. Galson, D. F. Kelly, and J. Trück. 2017. B-cell receptor repertoire sequencing in
294 patients with primary immunodeficiency: A review. *Immunology* 145–160.
295 13. Doria-Rose, N. A., C. A. Schramm, J. Gorman, P. L. Moore, J. N. Bhiman, B. J. DeKosky, M. J. Ernandes,
296 I. S. Georgiev, H. J. Kim, M. Pancera, R. P. Staupe, H. R. Altae-Tran, R. T. Bailer, E. T. Crooks, A. Cupo, A.
297 Druz, N. J. Garrett, K. H. Hoi, R. Kong, M. K. Louder, N. S. Longo, K. McKee, M. Nonyane, S. O'Dell, R. S.
298 Roark, R. S. Rudicell, S. D. Schmidt, D. J. Sheward, C. Soto, C. K. Wibmer, Y. Yang, Z. Zhang, J. C. Mullikin,
299 J. M. Binley, R. W. Sanders, I. A. Wilson, J. P. Moore, A. B. Ward, G. Georgiou, C. Williamson, S. S. A.
300 Karim, L. Morris, P. D. Kwong, L. Shapiro, and J. R. Mascola. 2014. Developmental pathway for potent
301 V1V2-directed HIV-neutralizing antibodies. *Nature* 508: 55–62.
302 14. Greiff, V., U. Menzel, E. Miho, C. Weber, R. Riedel, S. Cook, A. Valai, T. Lopes, A. Radbruch, T. H.
303 Winkler, and S. T. Reddy. 2017. Systems Analysis Reveals High Genetic and Antigen-Driven

304 Predetermination of Antibody Repertoires throughout B Cell Development. *Cell Rep.* 19: 1467–1478.
305 15. Hoi, K. H., and G. C. Ippolito. 2013. Intrinsic bias and public rearrangements in the human
306 immunoglobulin Vλ light chain repertoire. *Genes Immun.* 1–6.
307 16. Dekosky, B. J., T. Kojima, A. Rodin, W. Charab, G. C. Ippolito, A. D. Ellington, and G. Georgiou. 2014.
308 In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire.
309 *Nat. Med.* 21: 1–8.
310 17. Galson, J., J. Trück, D. F. Kelly, and R. van der Most. 2016. Investigating the effect of AS03 adjuvant
311 on the plasma cell repertoire following pH1N1 influenza vaccination. *Sci. Rep.* 6: 37229.
312 18. Galson, J., J. Trück, E. A. Clutterbuck, A. Fowler, V. Cerundolo, A. J. Pollard, G. Lunter, and D. F. Kelly.
313 2016. B cell repertoire dynamics after sequential Hepatitis B vaccination, and evidence for cross-reactive
314 B cell activation. *Submitt. Manuscr.* 1–13.
315 19. Jackson, K. J. L., Y. Liu, K. M. Roskin, J. Glanville, R. A. Hoh, K. Seo, E. L. Marshall, T. C. Gurley, M. A.
316 Moody, B. F. Haynes, E. B. Walter, H. X. Liao, R. A. Albrecht, A. García-Sastre, J. Chaparro-Riggers, A.
317 Rajpal, J. Pons, B. B. Simen, B. Hanczaruk, C. L. Dekker, J. Laserson, D. Koller, M. M. Davis, A. Z. Fire, and
318 S. D. Boyd. 2014. Human responses to influenza vaccination show seroconversion signatures and
319 convergent antibody rearrangements. *Cell Host Microbe* 16: 105–114.
320 20. Lee, J., D. R. Boutz, V. Chromikova, M. G. Joyce, C. Vollmers, K. Leung, A. P. Horton, B. J. DeKosky, C.-
321 H. Lee, J. J. Lavinder, E. M. Murrin, C. Chrysostomou, K. H. Hoi, Y. Tsybovsky, P. V Thomas, A. Druz, B.
322 Zhang, Y. Zhang, L. Wang, W.-P. Kong, D. Park, L. I. Popova, C. L. Dekker, M. M. Davis, C. E. Carter, T. M.
323 Ross, A. D. Ellington, P. C. Wilson, E. M. Marcotte, J. R. Mascola, G. C. Ippolito, F. Krammer, S. R. Quake,
324 P. D. Kwong, and G. Georgiou. 2016. Molecular-level analysis of the serum antibody repertoire in young
325 adults before and after seasonal influenza vaccination. *Nat Med* 22: 1456–1464.
326 21. Galson, J., E. A. Clutterbuck, J. Trück, M. N. Ramasamy, M. Münz, A. Fowler, V. Cerundolo, A. J.
327 Pollard, G. Lunter, and D. F. Kelly. 2015. BCR repertoire sequencing: Different patterns of B-cell
328 activation after two Meningococcal vaccines. *Immunol. Cell Biol.* 93: 885–895.
329 22. Zhou, T., J. Zhu, X. Wu, S. Moquin, B. Zhang, P. Acharya, I. S. Georgiev, H. R. Altae-Tran, G. Y. Chuang,
330 M. G. Joyce, Y. DoKwon, N. S. Longo, M. K. Louder, T. Luongo, K. McKee, C. A. Schramm, J. Skinner, Y.
331 Yang, Z. Yang, Z. Zhang, A. Zheng, M. Bonsignori, B. F. Haynes, J. F. Scheid, M. C. Nussenzweig, M. Simek,
332 D. R. Burton, W. C. Koff, J. C. Mullikin, M. Connors, L. Shapiro, G. J. Nabel, J. R. Mascola, and P. D. Kwong.
333 2013. Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway
334 for HIV-1 neutralization by VRC01-class antibodies. *Immunity* 39: 245–258.
335 23. DeKosky, B. J., G. C. Ippolito, R. P. Deschner, J. J. Lavinder, Y. Wine, B. M. Rawlings, N. Varadarajan, C.

336 Giesecke, T. Dörner, S. F. Andrews, P. C. Wilson, S. P. Hunicke-Smith, C. G. Willson, A. D. Ellington, and G.
337 Georgiou. 2013. High-throughput sequencing of the paired human immunoglobulin heavy and light
338 chain repertoire. *Nat. Biotechnol.* 31: 166–9.
339 24. Rubelt, F., C. E. Busse, S. A. C. Bukhari, J. P. Bürckert, E. Mariotti-Ferrandiz, L. G. Cowell, C. T. Watson,
340 N. Marthandan, W. J. Faison, U. Hershberg, U. Laserson, B. D. Corrie, M. M. Davis, B. Peters, M. P.
341 Lefranc, J. K. Scott, F. Breden, E. T. Luning Prak, and S. H. Kleinstein. 2017. Adaptive Immune Receptor
342 Repertoire Community recommendations for sharing immune-repertoire sequencing data. *Nat.*
343 *Immunol.* 18: 1274–1278.
344 25. Breden, F., E. T. Luning Prak, B. Peters, F. Rubelt, C. A. Schramm, C. E. Busse, J. A. Vander Heiden, S.
345 Christley, S. A. C. Bukhari, A. Thorogood, F. A. Matsen IV, Y. Wine, U. Laserson, D. Klatzmann, D. C.
346 Douek, M.-P. Lefranc, A. M. Collins, T. Bubela, S. H. Kleinstein, C. T. Watson, L. G. Cowell, J. K. Scott, and
347 T. B. Kepler. 2017. Reproducibility and Reuse of Adaptive Immune Receptor Repertoire Data. *Front.*
348 *Immunol.* 8: 1418.
349 26. Bhattacharya, S., S. Andorf, L. Gomes, P. Dunn, H. Schaefer, J. Pontius, P. Berger, V. Desborough, T.
350 Smith, J. Campbell, E. Thomson, R. Monteiro, P. Guimaraes, B. Walters, J. Wiser, and A. J. Butte. 2014.
351 ImmPort: Disseminating data to the public for the future of immunology. *Immunol. Res.* 58: 234–239.
352 27. Bhattacharya, S., P. Dunn, C. G. Thomas, B. Smith, H. Schaefer, J. Chen, Z. Hu, K. A. Zalocusky, R. D.
353 Shankar, S. S. Shen-Orr, E. Thomson, J. Wiser, and A. J. Butte. 2018. ImmPort, toward repurposing of
354 open access immunological assay data for translational and clinical research. *Sci. Data* 5.
355 28. Cowell, L. G. 2018. VDJServer: A Cloud-Based Analysis Portal and Data Commons for Immune
356 Repertoire Sequences and Rearrangements. *Front. Immunol.* 9: 976.
357 29. Leinonen, R., R. Akhtar, E. Birney, L. Bower, A. Cerdeno-Tárraga, Y. Cheng, I. Cleland, N. Faruque, N.
358 Goodgame, R. Gibson, G. Hoad, M. Jang, N. Pakseresht, S. Plaister, R. Radhakrishnan, K. Reddy, S.
359 Sobhany, P. Ten Hoopen, R. Vaughan, V. Zalunin, and G. Cochrane. 2011. The European nucleotide
360 archive. *Nucleic Acids Res.* 39.
361 30. NCBI Resource Coordinators. 2017. Database Resources of the National Center for Biotechnology
362 Information. *Nucleic Acids Res.* 45: D12–D17.
363 31. Schanz, M., T. Liechti, O. Zagordi, E. Miho, S. T. Reddy, H. F. Günthard, A. Trkola, and M. Huber. 2014.
364 High-throughput sequencing of human immunoglobulin variable regions with subtype identification.
365 *PLoS One* 9.
366 32. Rettig, T. A., C. Ward, B. A. Bye, M. J. Pecaut, and S. K. Chapes. 2018. Characterization of the naive
367 murine antibody repertoire using unamplified high-throughput sequencing. *PLoS One* 13.

368 33. Greiff, V., P. Bhat, S. C. Cook, U. Menzel, W. Kang, and S. T. Reddy. 2015. A bioinformatic framework
369 for immune repertoire diversity profiling enables detection of immunological status. *Genome Med.* 7:
370 49.
371 34. Magoč, T., and S. L. Salzberg. 2011. FLASH: Fast length adjustment of short reads to improve genome
372 assemblies. *Bioinformatics* 27: 2957–2963.
373 35. HannonLab. 2014. FASTX toolkit. *Cold Spring Harb. Lab. Cold Spring Harb. NY* .
374 36. Giudicelli, V., D. Chaume, and M.-P. Lefranc. 2005. IMGT/GENE-DB: a comprehensive database for
375 human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res.* 33: D256-61.
376 37. Smith, T. F., and M. S. Waterman. 1981. Identification of common molecular subsequences. *J. Mol.*
377 *Biol.* 147: 195–197.
378 38. Galson, J. D., J. Trück, E. A. Clutterbuck, A. Fowler, V. Cerundolo, A. J. Pollard, G. Lunter, and D. F.
379 Kelly. 2016. B-cell repertoire dynamics after sequential hepatitis B vaccination and evidence for cross-
380 reactive B-cell activation. *Genome Med.* 8: 68.
381 39. Galson, J., J. Trück, A. Fowler, E. A. Clutterbuck, M. Münz, V. Cerundolo, C. Reinhard, R. van der Most,
382 A. J. Pollard, G. Lunter, and D. F. Kelly. 2015. Analysis of B Cell Repertoire Dynamics Following Hepatitis B
383 Vaccination in Humans, and Enrichment of Vaccine-specific Antibody Sequences. *EBioMedicine* 2: 2070–
384 2079.
385 40. Greiff, V., U. Menzel, U. Haessler, S. C. Cook, S. Friedensohn, T. A. Khan, M. Pogson, I. Hellmann, and
386 S. T. Reddy. 2014. Quantitative assessment of the robustness of next-generation sequencing of antibody
387 variable gene repertoires from immunized mice. *BMC Immunol.* 15: 40.
388 41. Ye, J., N. Ma, T. L. Madden, and J. M. Ostell. 2013. IgBLAST: an immunoglobulin variable domain
389 sequence analysis tool. *Nucleic Acids Res.* 41.
390 42. Dunbar, J., and C. M. Deane. 2015. ANARCI: Antigen receptor numbering and receptor classification.
391 *Bioinformatics* btv552.
392 43. Lefranc, M.-P., C. Pommié, M. Ruiz, V. Giuducelli, E. Foulquier, L. Truong, V. Thouvenin-Contet, and
393 G. Lefranc. 2003. IMGT unique numbering fro immunoglobulin and T cell receptor variable domains and
394 Ig superfamily V-like domains. *Dev. Comp. Immunol.* 27: 55–77.
395 44. Eddy, S. R. 1995. Multiple alignment using hidden Markov models. *Proc. Int. Conf. Intell. Syst. Mol.*
396 *Biol.* 3: 114–120.
397 45. Shugay, M., O. V Britanova, E. M. Merzlyak, M. a Turchaninova, I. Z. Mamedov, T. R. Tuganbaev, D. a
398 Bolotin, D. B. Staroverov, E. V Putintseva, K. Plevova, C. Linnemann, D. Shagin, S. Pospisilova, S.
399 Lukyanov, T. N. Schumacher, and D. M. Chudakov. 2014. Towards error-free profiling of immune

400 repertoires. *Nat. Methods* 11: 653–5.
401 46. Banerjee, S., H. Shi, M. Banasik, H. Moon, W. Lees, Y. Qin, A. Harley, A. Shepherd, and M. W. Cho.
402 2017. Evaluation of a novel multi-immunogen vaccine strategy for targeting 4E10/10E8 neutralizing
403 epitopes on HIV-1 gp41 membrane proximal external region. *Virology* 505: 113–126.
404 47. Bashford-Rogers, R. J. M., A. L. Palser, B. J. Huntly, R. Rance, G. S. Vassiliou, G. A. Follows, and P.
405 Kellam. 2013. Network properties derived from deep sequencing of human B-cell receptor repertoires
406 delineate B-cell populations. *Genome Res.* 23: 1874–1884.
407 48. Bhiman, J. N., C. Anthony, N. A. Doria-Rose, O. Karimanzira, C. A. Schramm, T. Khoza, D. Kitchin, G.
408 Botha, J. Gorman, N. J. Garrett, S. S. A. Karim, L. Shapiro, C. Williamson, P. D. Kwong, J. R. Mascola, L.
409 Morris, and P. L. Moore. 2015. Viral variants that initiate and drive maturation of V1V2-directed HIV-1
410 broadly neutralizing antibodies. *Nat. Med.* 21: 1332–1336.
411 49. Bonsignori, M., T. Zhou, Z. Sheng, L. Chen, F. Gao, M. G. Joyce, G. Ozorowski, G. Y. Chuang, C. A.
412 Schramm, K. Wiehe, S. M. Alam, T. Bradley, M. A. Gladden, K. K. Hwang, S. Iyengar, A. Kumar, X. Lu, K.
413 Luo, M. C. Mangiapani, R. J. Parks, H. Song, P. Acharya, R. T. Bailer, A. Cao, A. Druz, I. S. Georgiev, Y. D.
414 Kwon, M. K. Louder, B. Zhang, A. Zheng, B. J. Hill, R. Kong, C. Soto, J. C. Mullikin, D. C. Douek, D. C.
415 Montefiori, M. A. Moody, G. M. Shaw, B. H. Hahn, G. Kelsoe, P. T. Hraber, B. T. Korber, S. D. Boyd, A. Z.
416 Fire, T. B. Kepler, L. Shapiro, A. B. Ward, J. R. Mascola, H. X. Liao, P. D. Kwong, and B. F. Haynes. 2016.
417 Maturation Pathway from Germline to Broad HIV-1 Neutralizer of a CD4-Mimic Antibody. *Cell* 165: 449–
418 463.
419 50. Collins, A. M., Y. Wang, K. M. Roskin, C. P. Marquis, K. J. L. Jackson, H. Schroeder, L. Cavacini, S.
420 Tonegawa, F. Alt, D. Baltimore, K. Jackson, M. Kidd, Y. Wang, A. Collins, M. Atkinson, M. Cowan, A.
421 Feeney, D. Granoff, P. Shackelford, S. Holmes, A. Lucas, B. Nadel, L. Liu, A. Lucas, L. Pappas, Y. Wang, K.
422 Jackson, B. Gaeta, W. Pomat, P. Siba, W. Sewell, A. Collins, C. Watson, C. Scheepers, M. Kidd, E. Houpt, K.
423 Breitbach, P. Wongprompitak, I. Steinmetz, G. Peltz, E. Marino, S. Grey, M. Potter, R. Riblet, C. Johnston,
424 M. Lefranc, I. Retter, K. Jackson, Y. Wang, A. Collins, J. Ye, N. Ma, T. Madden, J. Ostell, R. Edgar, V. Greiff,
425 U. Menzel, U. Haessler, S. Cook, S. Friedensohn, T. Khan, M. Pogson, I. Hellmann, S. Reddy, S. Boyd, I.
426 Retter, J. Meier, S. Lewis, D. Gadala-Maria, G. Yaari, M. Uduman, S. Kleinstein, B. Haines, C. Angeles, A.
427 Parmelee, P. McLean, P. Brodeur, J. Ye, A. Feeney, R. Riblet, K. Jackson, B. Gaeta, A. Collins, B. Blomberg,
428 W. Geckeler, M. Weigert, E. Jouvin-Marche, M. Morgado, P. Brodeur, R. Riblet, A. Collins, M. Ikutani, D.
429 Puiu, G. Buck, A. Nadkarni, B. Gaeta, K. Larimore, M. McCormick, H. Robins, P. Greenberg, K. Jackson, B.
430 Gaeta, W. Sewell, A. Collins, J. Glanville, H. Yang, C. Scott, M. Potter, L. Reidl, C. Kinoshita, and L. Steiner.
431 2015. The mouse antibody heavy chain repertoire is germline-focused and highly variable between

432 inbred strains. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370: S41–S52.
433 51. Corcoran, M. M., G. E. Phad, N. V. Bernat, C. Stahl-Hennig, N. Sumida, M. A. A. Persson, M. Martin,
434 and G. B. K. Hedestam. 2016. Production of individualized v gene databases reveals high levels of
435 immunoglobulin genetic diversity. *Nat. Commun.* 7.
436 52. Cui, A., R. Di Niro, J. A. Vander Heiden, A. W. Briggs, K. Adams, T. Gilbert, K. C. O'Connor, F.
437 Vigneault, M. J. Shlomchik, and S. H. Kleinstein. 2016. A Model of Somatic Hypermutation Targeting in
438 Mice Based on High-Throughput Ig Sequencing Data. *J. Immunol.* 197: 3566–3574.
439 53. Fisher, C. R., H. J. Sutton, J. A. Kaczmarski, H. A. McNamara, B. Clifton, J. Mitchell, Y. Cai, J. N. Dups,
440 N. J. D'Arcy, M. Singh, A. Chuah, T. S. Peat, C. J. Jackson, and I. A. Cockburn. 2017. T-dependent B cell
441 responses to Plasmodium induce antibodies that form a high-avidity multivalent complex with the
442 circumsporozoite protein. *PLoS Pathog.* 13.
443 54. Gupta, N. T., K. D. Adams, A. W. Briggs, S. C. Timberlake, F. Vigneault, and S. H. Kleinstein. 2017.
444 Hierarchical Clustering Can Identify B Cell Clones with High Confidence in Ig Repertoire Sequencing Data.
445 *J. Immunol.* 198: 2489–2499.
446 55. Halliley, J. L., C. M. Tipton, J. Liesveld, A. F. Rosenberg, J. Darce, I. V Gregoretti, L. Popova, D.
447 Kaminiski, C. F. Fucile, I. Albizua, S. Kyu, K.-Y. Chiang, K. T. Bradley, R. Burack, M. Slifka, E. Hammarlund,
448 H. Wu, L. Zhao, E. E. Walsh, A. R. Falsey, T. D. Randall, W. C. Cheung, I. Sanz, and F. E.-H. Lee. 2015. Long-
449 Lived Plasma Cells Are Contained within the CD19(-)CD38(hi)CD138(+) Subset in Human Bone Marrow.
450 *Immunity* 43: 132–45.
451 56. Huang, J., B. H. Kang, E. Ishida, T. Zhou, T. Griesman, Z. Sheng, F. Wu, N. A. Doria-Rose, B. Zhang, K.
452 McKee, S. O'Dell, G. Y. Chuang, A. Druz, I. S. Georgiev, C. A. Schramm, A. Zheng, M. G. Joyce, M. Asokan,
453 A. Ransier, S. Darko, S. A. Migueles, R. T. Bailer, M. K. Louder, S. M. Alam, R. Parks, G. Kelsoe, T. Von
454 Holle, B. F. Haynes, D. C. Douek, V. Hirsch, M. S. Seaman, L. Shapiro, J. R. Mascola, P. D. Kwong, and M.
455 Connors. 2016. Identification of a CD4-Binding-Site Antibody to HIV that Evolved Near-Pan
456 Neutralization Breadth. *Immunity* 45: 1108–1121.
457 57. Jiang, N., J. He, J. a Weinstein, L. Penland, S. Sasaki, X.-S. He, C. L. Dekker, N.-Y. Zheng, M. Huang, M.
458 Sullivan, P. C. Wilson, H. B. Greenberg, M. M. Davis, D. S. Fisher, and S. R. Quake. 2013. Lineage structure
459 of the human antibody repertoire in response to influenza vaccination. *Sci. Transl. Med.* 5: 171ra19.
460 58. Joyce, M. G., A. K. Wheatley, P. V. Thomas, G. Y. Chuang, C. Soto, R. T. Bailer, A. Druz, I. S. Georgiev,
461 R. A. Gillespie, M. Kanekiyo, W. P. Kong, K. Leung, S. N. Narpala, M. S. Prabhakaran, E. S. Yang, B. Zhang,
462 Y. Zhang, M. Asokan, J. C. Boyington, T. Bylund, S. Darko, C. R. Lees, A. Ransier, C. H. Shen, L. Wang, J. R.
463 Whittle, X. Wu, H. M. Yassine, C. Santos, Y. Matsuoka, Y. Tsybovsky, U. Baxa, J. C. Mullikin, K. Subbarao,

464 D. C. Douek, B. S. Graham, R. A. Koup, J. E. Ledgerwood, M. Roederer, L. Shapiro, P. D. Kwong, J. R.
465 Mascola, and A. B. McDermott. 2016. Vaccine-Induced Antibodies that Neutralize Group 1 and Group 2
466 Influenza A Viruses. *Cell* 166: 609–623.
467 59. Khan, T. A., S. Friedensohn, A. R. G. de Vries, J. Straszewski, H.-J. Ruscheweyh, and S. T. Reddy. 2016.
468 Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting. *Sci. Adv.*
469 2: e1501371–e1501371.
470 60. Levin, M., J. J. King, J. Glanville, K. J. L. Jackson, T. J. Looney, R. A. Hoh, A. Mari, M. Andersson, L.
471 Greiff, A. Z. Fire, S. D. Boyd, and M. Ohlin. 2016. Persistence and evolution of allergen-specific IgE
472 repertoires during subcutaneous specific immunotherapy. *J. Allergy Clin. Immunol.* 137: 1535–1544.
473 61. Levin, M., F. Levander, R. Palmason, L. Greiff, and M. Ohlin. 2017. Antibody-encoding repertoires of
474 bone marrow and peripheral blood—a focus on IgE. *J. Allergy Clin. Immunol.* 139: 1026–1030.
475 62. Li, X., X. Duan, K. Yang, W. Zhang, C. Zhang, L. Fu, Z. Ren, C. Wang, J. Wu, R. Lu, Y. Ye, M. He, C. Nie,
476 N. Yang, J. Wang, H. Yang, X. Liu, and W. Tan. 2016. Comparative analysis of immune repertoires
477 between bactrian Camel's conventional and heavy-chain antibodies. *PLoS One* 11.
478 63. Liao, H. X., R. Lynch, T. Zhou, F. Gao, S. Munir Alam, S. D. Boyd, A. Z. Fire, K. M. Roskin, C. A.
479 Schramm, Z. Zhang, J. Zhu, L. Shapiro, J. C. Mullikin, S. Gnanakaran, P. Hraber, K. Wiehe, G. Kelsoe, G.
480 Yang, S. M. Xia, D. C. Montefiori, R. Parks, K. E. Lloyd, R. M. Scearce, K. A. Soderberg, M. Cohen, G.
481 Kamanga, M. K. Louder, L. M. Tran, Y. Chen, F. Cai, S. Chen, S. Moquin, X. Du, M. Gordon Joyce, S.
482 Srivatsan, B. Zhang, A. Zheng, G. M. Shaw, B. H. Hahn, T. B. Kepler, B. T. M. Korber, P. D. Kwong, J. R.
483 Mascola, B. F. Haynes, J. Becker, B. Benjamin, R. Blakesley, G. Bouffard, S. Brooks, H. Coleman, M.
484 Dekhtyar, M. Gregory, X. Guan, J. Gupta, J. Han, A. Hargrove, S. L. Ho, T. Johnson, R. Legaspi, S. Lovett, Q.
485 Maduro, C. Masiello, B. Maskeri, J. McDowell, C. Montemayor, J. Mulliki, M. Park, N. Riebow, K.
486 Schandler, B. Schmidt, C. Sison, M. Stantripop, J. Thomas, P. Thomas, M. Vemulapalli, and A. Young.
487 2013. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* 496: 469–476.
488 64. Lindner, C., I. Thomsen, B. Wahl, M. Ugur, M. K. Sethi, M. Friedrichsen, A. Smoczek, S. Ott, U.
489 Baumann, S. Suerbaum, S. Schreiber, A. Bleich, V. Gaboriau-Routhiau, N. Cerf-Bensussan, H. Hazanov, R.
490 Mehr, P. Boysen, P. Rosenstiel, and O. Pabst. 2015. Diversification of memory B cells drives the
491 continuous adaptation of secretory antibodies to gut microbiota. *Nat. Immunol.* 16: 880–888.
492 65. Meng, W., B. Zhang, G. W. Schwartz, A. M. Rosenfeld, D. Ren, J. J. C. Thome, D. J. Carpenter, N.
493 Matsuoka, H. Lerner, A. L. Friedman, T. Granot, D. L. Farber, M. J. Shlomchik, U. Hershberg, and E. T.
494 Luning Prak. 2017. An atlas of B-cell clonal distribution in the human body. *Nat. Biotechnol.* 35: 879–886.
495 66. Menzel, U., V. Greiff, T. A. Khan, U. Haessler, I. Hellmann, S. Friedensohn, S. C. Cook, M. Pogson, and

496 S. T. Reddy. 2014. Comprehensive evaluation and optimization of amplicon library preparation methods
497 for high-throughput antibody sequencing. *PLoS One* 9.
498 67. Mroczek, E. S., G. C. Ippolito, T. Rogosch, K. H. Hoi, T. A. Hwangpo, M. G. Brand, Y. Zhuang, C. R. Liu,
499 D. A. Schneider, M. Zemlin, E. E. Brown, G. Georgiou, and H. W. Schroeder. 2014. Differences in the
500 composition of the human antibody repertoire by B cell subsets in the blood. *Front Immunol* 5: 96.
501 68. Ota, M., B. H. Duong, A. Torkamani, C. M. Doyle, A. L. Gavin, T. Ota, and D. Nemazee. 2010.
502 Regulation of the B cell receptor repertoire and self-reactivity by BAFF. *J. Immunol.* 185: 4128–36.
503 69. Palanichamy, A., L. Apeltsin, T. C. Kuo, M. Sirota, S. Wang, S. J. Pitts, P. D. Sundar, D. Telman, L. Z.
504 Zhao, M. Derstine, A. Abounasr, S. L. Hauser, and H. C. Von B??dingen. 2014. Immunoglobulin class-
505 switched B cells form an active immune axis between CNS and periphery in multiple sclerosis. *Sci. Transl.*
506 *Med.* 6.
507 70. Prohaska, T. A., X. Que, C. J. Diehl, S. Hendrikx, M. W. Chang, K. Jepsen, C. K. Glass, C. Benner, and J.
508 L. Witztum. 2018. Massively Parallel Sequencing of Peritoneal and Splenic B Cell Repertoires Highlights
509 Unique Properties of B-1 Cell Antibodies. *J. Immunol.* 200: 1702–1717.
510 71. Rubelt, F., C. R. Bolen, H. M. McGuire, J. A. V. Heiden, D. Gadala-Maria, M. Levin, G. M. Euskirchen,
511 M. R. Mamedov, G. E. Swan, C. L. Dekker, L. G. Cowell, S. H. Kleinstein, and M. M. Davis. 2016. Individual
512 heritable differences result in unique cell lymphocyte receptor repertoires of naïve and antigen-
513 experienced cells. *Nat. Commun.* 7.
514 72. Zhu, J., G. Ofek, Y. Yang, B. Zhang, M. K. Louder, G. Lu, K. McKee, M. Pancera, J. Skinner, Z. Zhang, R.
515 Parks, J. Eudailey, K. E. Lloyd, J. Blinn, S. M. Alam, B. F. Haynes, M. Simek, D. R. Burton, W. C. Koff, N. C. S.
516 NISC Comparative Sequencing Program, J. C. Mullikin, J. R. Mascola, L. Shapiro, and P. D. Kwong. 2013.
517 Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and
518 phylogenetic pairing of heavy/light chains. *Proc. Natl. Acad. Sci. U. S. A.* 110: 6470–5.
519 73. Stern, J. N. H., G. Yaari, J. A. Vander Heiden, G. Church, W. F. Donahue, R. Q. Hintzen, A. J. Huttner, J.
520 D. Laman, R. M. Nagra, A. Nylander, D. Pitt, S. Ramanan, B. A. Siddiqui, F. Vigneault, S. H. Kleinstein, D. A.
521 Hafler, and K. C. O'Connor. 2014. B cells populating the multiple sclerosis brain mature in the draining
522 cervical lymph nodes. *Sci. Transl. Med.* 6.
523 74. Sundling, C., Z. Zhang, G. E. Phad, Z. Sheng, Y. Wang, J. R. Mascola, Y. Li, R. T. Wyatt, L. Shapiro, and
524 G. B. Karlsson Hedestam. 2014. Single-Cell and Deep Sequencing of IgG-Switched Macaque B Cells
525 Reveal a Diverse Ig Repertoire following Immunization. *J. Immunol.* 192: 3637–3644.
526 75. Tipton, C. M., C. F. Fucile, J. Darce, A. Chida, T. Ichikawa, I. Gregoretti, S. Schieferl, J. Hom, S. Jenks, R.
527 J. Feldman, R. Mehr, C. Wei, F. E. H. Lee, W. C. Cheung, A. F. Rosenberg, and I. Sanz. 2015. Diversity,

528 cellular origin and autoreactivity of antibody-secreting cell population expansions in acute systemic
529 lupus erythematosus. *Nat. Immunol.* 16: 755–765.
530 76. Turchaninova, M. A., A. Davydov, O. V Britanova, M. Shugay, V. Bikos, E. S. Egorov, V. I. Kirgizova, E.
531 M. Merzlyak, D. B. Staroverov, D. A. Bolotin, I. Z. Mamedov, M. Izraelson, M. D. Logacheva, O. Kladova,
532 K. Plevova, S. Pospisilova, and D. M. Chudakov. 2016. High-quality full-length immunoglobulin profiling
533 with unique molecular barcoding. *Nat. Protoc.* 11: 1599–1616.
534 77. Vander Heiden, J. A., P. Stathopoulos, J. Q. Zhou, L. Chen, T. J. Gilbert, C. R. Bolen, R. J. Barohn, M. M.
535 Dimachkie, E. Ciafaloni, T. J. Broering, F. Vigneault, R. J. Nowak, S. H. Kleinstein, and K. C. O'Connor.
536 2017. Dysregulation of B Cell Repertoire Formation in Myasthenia Gravis Patients Revealed through
537 Deep Sequencing. *J. Immunol.* 198: 1460–1473.
538 78. VanDuijn, M. M., L. J. Dekker, W. F. J. van IJcken, P. A. E. Sillevis Smitt, and T. M. Luider. 2017.
539 Immune repertoire after immunization as seen by next-generation sequencing and proteomics. *Front.*
540 *Immunol.* 8.
541 79. Vergani, S., I. Korsunsky, A. N. Mazzarello, G. Ferrer, N. Chiorazzi, and D. Bagnara. 2017. Novel
542 method for high-throughput full-length IGHV-D-J sequencing of the immune repertoire from bulk B-cells
543 with single-cell resolution. *Front. Immunol.* 8.
544 80. Wesemann, D. R., A. J. Portuguese, R. M. Meyers, M. P. Gallagher, K. Cluff-Jones, J. M. Magee, R. A.
545 Panchakshari, S. J. Rodig, T. B. Kepler, and F. W. Alt. 2013. Microbial colonization influences early B-
546 lineage development in the gut lamina propria. *Nature* 501: 112–115.
547 81. Wu, X., T. Zhou, J. Zhu, B. Zhang, I. Georgiev, C. Wang, X. Chen, N. S. Longo, M. Louder, K. McKee, S.
548 O'Dell, S. Perfetto, S. D. Schmidt, W. Shi, L. Wu, Y. Yang, Z.-Y. Yang, Z. Yang, Z. Zhang, M. Bonsignori, J. A.
549 Crump, S. H. Kapiga, N. E. Sam, B. F. Haynes, M. Simek, D. R. Burton, W. C. Koff, N. A. Doria-Rose, M.
550 Connors, N. C. S. NISC Comparative Sequencing Program, J. C. Mullikin, G. J. Nabel, M. Roederer, L.
551 Shapiro, P. D. Kwong, and J. R. Mascola. 2011. Focused evolution of HIV-1 neutralizing antibodies
552 revealed by structures and deep sequencing. *Science* 333: 1593–602.
553 82. Wu, X., Z. Zhang, C. A. Schramm, M. G. Joyce, Y. Do Kwon, T. Zhou, Z. Sheng, B. Zhang, S. O'Dell, K.
554 McKee, I. S. Georgiev, G. Y. Chuang, N. S. Longo, R. M. Lynch, K. O. Saunders, C. Soto, S. Srivatsan, Y.
555 Yang, R. T. Bailer, M. K. Louder, J. C. Mullikin, M. Connors, P. D. Kwong, J. R. Mascola, and L. Shapiro.
556 2015. Maturation and diversity of the VRC01-antibody lineage over 15 years of chronic HIV-1 infection.
557 *Cell* 161: 480–485.
558 83. Wu, Y. C. B., L. K. James, J. A. Vander Heiden, M. Uduman, S. R. Durham, S. H. Kleinstein, D. Kipling,
559 and H. J. Gould. 2014. Influence of seasonal exposure to grass pollen on local and peripheral blood IgE

560 repertoires in patients with allergic rhinitis. *J. Allergy Clin. Immunol.* 134: 604–612.
561 84. Zhou, T., R. M. Lynch, L. Chen, P. Acharya, X. Wu, N. A. Doria-Rose, M. G. Joyce, D. Lingwood, C. Soto,
562 R. T. Bailer, M. J. Ernandes, R. Kong, N. S. Longo, M. K. Louder, K. McKee, S. O'Dell, S. D. Schmidt, L. Tran,
563 Z. Yang, A. Druz, T. S. Luongo, S. Moquin, S. Srivatsan, Y. Yang, B. Zhang, A. Zheng, M. Pancera, T. Kirys, I.
564 S. Georgiev, T. Gindin, H. P. Peng, A. S. Yang, J. C. Mullikin, M. D. Gray, L. Stamatatos, D. R. Burton, W. C.
565 Koff, M. S. Cohen, B. F. Haynes, J. P. Casazza, M. Connors, D. Corti, A. Lanzavecchia, Q. J. Sattentau, R. A.
566 Weiss, A. P. West, P. J. Bjorkman, J. F. Scheid, M. C. Nussenzweig, L. Shapiro, J. R. Mascola, and P. D.
567 Kwong. 2015. Structural repertoire of HIV-1-neutralizing antibodies targeting the CD4 supersite in 14
568 donors. *Cell* 161: 1280–1292.
569 85. Zhu, J., S. O'Dell, G. Ofek, M. Pancera, X. Wu, B. Zhang, Z. Zhang, J. C. Mullikin, M. Simek, D. R.
570 Burton, W. C. Koff, L. Shapiro, J. R. Mascola, and P. D. Kwong. 2012. Somatic populations of PGT135-137
571 HIV-1-neutralizing antibodies identified by 454 pyrosequencing and bioinformatics. *Front. Microbiol.* 3.
572 86. Zhu, J., X. Wu, B. Zhang, K. McKee, S. O'Dell, C. Soto, T. Zhou, J. P. Casazza, J. C. Mullikin, P. D. Kwong,
573 J. R. Mascola, L. Shapiro, J. Becker, B. Benjamin, R. Blakesley, G. Bouffard, S. Brooks, H. Coleman, M.
574 Dekhtyar, M. Gregory, X. Guan, J. Gupta, J. Han, A. Hargrove, S. -l. Ho, T. Johnson, R. Legaspi, S. Lovett,
575 Q. Maduro, C. Masiello, B. Maskeri, J. McDowell, C. Montemayor, J. Mullikin, M. Park, N. Riebow, K.
576 Schandler, B. Schmidt, C. Sison, M. Stantripop, J. Thomas, P. Thomas, M. Vemulapalli, and A. Young.
577 2013. De novo identification of VRC01 class HIV-1-neutralizing antibodies by next-generation sequencing
578 of B-cell transcripts. *Proc. Natl. Acad. Sci.* 110: E4088–E4097.
579 87. Schroeder Jr, H. W. 2006. Similarity and divergence in the development and expression of the mouse
580 and human antibody repertoires. *Dev. Comp. Immunol.* 30: 119–135.
581 88. Zemlin, M., M. Klinger, J. Link, C. Zemlin, K. Bauer, J. A. Engler, H. W. Schroeder, and P. M. Kirkham.
582 2003. Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that
583 differ in their amino acid composition and predicted range of structures. *J. Mol. Biol.* 334: 733–749.
584 89. Olimpieri, P. P., P. Marcatili, and A. Tramontano. 2014. Tabhu: Tools for antibody humanization.
585 *Bioinformatics* 31: 434–435.
586