Distinct repertoire development history during affinity maturation of antigen-responding VHH antibody

Tomonari Matsuda[1]†, Yoko Akazawa-Ogawa[2]†, Lilian-Kaede Komaba[2], Norihiko Kiyose[3], Nobuo Miyazaki[3], Yusaku Mizuguchi[4], Tetsuo Fukuta[4], Yuji Ito[5], & Yoshihisa Hagihara[2]*

[1]Research Center for Environmental Quality Management, Kyoto University, 1-2 Yumiyama, Otsu, Shiga 520-0811
[2]Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 1-8-31 Midorigaoka, Ikeda, Osaka 563-8577
[3]ARK Resource. Co., Ltd., 383-2 Nakahara-machi, Nishi-ku, Kumamoto 861-5271
[4]JSR Corporation, 25 Miyukigaoka, Tsukuba, Ibaraki 305-0841
[5]Graduate School of Science and Engineering, Kagoshima University, 1-21-35 Korimoto, Kagoshima 890-0065

*Corresponding author. Email: hagihara-kappael@aist.go.jp
†These authors contributed equally to this work.

Abstract

Time-resolved antibody repertoire analysis is essential to elucidate the antibody maturation process during immunization. However, most analyses are merely snapshot comparisons of unimmunized and fully immunized animals. In this study, an alpaca was immunized with multiple antigen injections and its repertoire development was traced by high-throughput next-generation sequencing every week for 3 months. The use of alpaca-derived, single-domain antibody circumvents the problems associated with conventional antibodies including heavy-chain and light-chain matching and labor-intensive production. The sequences were processed into clusters and the antibodies were generated to determine whether the clusters included antigen-binding antibodies. For most of the antigen-responding clusters, repertoire development exhibited distinct time-dependent patterns compared with clusters unresponsive to antigens. Moreover, the sequences at the tips and root of the phylogenetic tree had strong and weak antigen-binding affinity, respectively. The foregoing characteristics could be exploited in the prediction of antigen-responding clusters and the physical properties of antibodies.

Teaser

Clusters of antigen-responding, single-domain antibodies exhibited distinct time-dependent repertoire development history

Introduction

Antibodies accumulate somatic hypermutations and undergo affinity maturation upon exposure to antigens (*1*). Immunization exploits this mechanism to produce antibodies against the target antigens. Repetitive antigen injections introduce random mutations and increase the antigen affinity of the antibodies. The history of the mutational changes that occur in antibodies during immunization directly reflects the enhancement of the adaptive humoral immune response and helps elucidate the antibody maturation process in response to specific antigens.

High-throughput next-generation sequencing (NGS) of vast immune repertoires provides useful information for immunological system research and its practical applications (*2, 3*). Unlike conventional sequencing techniques, NGS enables us to draw a comprehensive picture of immune repertoires that respond to antigens. However, the utilization of NGS to analyze immunization-induced antibody repertoire development has been limited to snapshot comparisons of unimmunized and fully immunized animals at certain time intervals (*4-6*). Hence, precise time-resolved repertoire development during immunization is unclear.

It is difficult to link antibody repertoire development with the changes in protein level characteristic of antigen-responding antibodies. Despite the development of various empirical and bioinformatics technologies for nucleotide sequencing (*7-9*), correct light-chain and heavy-chain matching remains a challenging problem in the biophysical study of antibody obtained by high-throughput sequencing. Furthermore, preparation of full-length antibody from NGS sequence reads requires time-consuming recombinant strain construction and mammalian cell culture. Small antibody formats such as single-chain F_v fragment (scF_v) and F_{ab} can be produced by bacterial hosts. Nevertheless, this approach may result in aggregation, defective folding, and loss of activity. The V_H domain of camelid heavy-chain antibody (VHH) binds the antigen in single-domain format (*10, 11*) and can usually be produced rapidly, conveniently, and inexpensively in an *Escherichia coli* (*E. coli*) expression system (*12*). VHH is a suitable antibody format to examine numerous sequences and explore the physical effects of mutational changes induced by affinity maturation.

Here, we report VHH antibody repertoire development using pools of peripheral lymphocytes collected from alpaca blood every week for 3 months. The animal had undergone specific antigen immunization. The sequences were clustered according to length and similarity and were analyzed for time-dependent mutational changes. The cluster VHHs were produced and examined to determine whether they interacted with the immunized antigen. We also investigated temporal changes in the antigen-binding activity of various VHHs in the same cluster.

Results

Alpaca polyclonal antibody immune responses against injected antigens

We immunized an alpaca against ranibizumab, F_{ab} from trastuzumab, and a human κ C_L fragment produced by *E. coli* (Fig. 1). The animal had already been immunized with the same adjuvants before this experiment. Blood samples were collected immediately before immunization and every week thereafter for 14 weeks. We measured increases in the purified immunoglobulin G (IgG) IgG1, IgG2, and IgG3 subclass titers and completed immunization after six antigen injections (Fig. 1). The VHH cDNAs were synthesized by reverse transcription of mRNAs extracted from a lymphocyte pool. Short-hinge and long-hinge VHH-specific primers were used and the cDNAs were amplified for sequencing.

Clustering VHH sequences from immunized alpaca

After merging the overlaps, we obtained averages of 169,000 and 161,000 full-length VHH sequences from IgG2 (short-hinge between VHH and C_{H1}) and IgG3 (long-hinge between VHH and C_{H1}), respectively (Table S1). The same sequences were gathered into "unique sequences" and cleaned of any sequence errors. At each time point, we obtained averages of 3,000 and 25,000 unique sequences from IgG2 and IgG3, respectively. The

sequences were grouped according to their germ-line V and J combinations and their lengths (Fig. S1). Here, we refer to a set of DNA sequences as a "group" potentially consisting of various antibody families. D-region data were not used for grouping as they were too short and introduced ambiguity into the sequence matching. It was assumed that in most cases, the sequence lengths were the same for all members of each antibody family propagated from a single ancestral sequence to adapt a specific antigen.

After excluding lone sequences, the DNA sequences within a group were divided by phylogenetic tree analysis into "clusters". We hypothesized that the clusters had the same properties as the antibody families. It was assumed that the sequences bound the same antigens and shared the same ancestors. However, it was difficult to conclude that each cluster covers entire sequences that evolved from the initial sequence. Moreover, we could not rule out the possibility that each cluster contained sets of different antibodies recognizing various antigens. Before the experiment, the alpaca used here had already been immunized with the same adjuvants. Therefore, we discarded clusters including sequences expressed before immunization and assumed that any clusters interacting with the adjuvants were removed at this step. We obtained 321 clusters comprising 923 and 3,546 sequences derived from IgG2 and IgG3, respectively. Numerous identical sequences were observed in IgG2 and IgG3. Hence, the total number of unique sequences was, in fact, less than the sum of the unique sequences derived from IgG2 and IgG3.

VHH sequence propagation by immunization

We attempted to elucidate the characteristics of the 16 predominant clusters showing the highest maximum percentage appearance. The percentage appearance is the sum of the percentage occupancy of the IgG2 and IgG3 sequences in the cluster relative to all IgG2 and IgG3 sequences per week. The maximum percentage appearance is the largest percentage appearance of the cluster during immunization. The sequences for clusters 7 and 15 were identified by bio-panning the M13 phage library for the cDNA of the VHHs collected at week 9. We also examined clusters 69, 99, and 210 which were identified by bio-panning and cluster 33 which was identified by comparing the sequence propensities before and after bio-panning (*13*, *14*). We prepared VHH proteins for all 20 clusters (Table S2) and used enzyme-linked immunosorbent assay (ELISA) and surface resonance spectroscopy (SPR) to evaluate their antigen affinities (Fig. S2).

Only the VHH clones in clusters 2, 5, 7, 10, 15, and 16 exhibited antigen binding ("hit clusters"). No antigen binding was detected for the clones in clusters 1, 3, 6, 8, 9, 11, 12, 13, or 14 ("miss clusters"). SPR revealed that clone S38 in cluster 4 bound aberrantly to the C_L fragment. Thus, cluster 4 could not be designated an antigen-binding clone and was excluded from further analysis. All clones had the same nomenclature as the sequence ID. The initial S and L indicate sequences derived from short-hinge antibody (IgG2) and long-hinge antibody (IgG3), respectively. The numbers following the S and L in the sequence ID refer to the descending order of maximum sequence frequency. ELISA and SPR showed that clones S11 (cluster 2), L926 (cluster 15), L792 (cluster 16), and L252126 (cluster 99) exhibited affinity for ranibizumab and F_{ab} from trastuzumab. Thus, the epitopes of these clones constituted the F_v and/or C_{H1} domain framework regions. Clones S1139 (cluster 7) and L54 (cluster 33) had profiles resembling those of the other clones. At high VHH concentrations, however, the ELISA signal against the C_L domain interfered with the determination of the epitope location. The ELISA and SPR signals were positive against the C_L fragment and F_{ab}s. Hence, the clone L2477 (cluster 69) epitope occurred on

the C_L domain. Clone L19 (cluster 10) only produced an ELISA signal against F_{ab}s. Its epitope may have been buried in the antigen and then exposed by its interaction with the plastic surface. The latter can induce protein unfolding. There were ELISA signals in clones L38 (cluster 5) and L15235 (cluster 210) against all antigens. However, the SPR signal was against the C_L fragment alone. We hypothesized that these epitopes occurred on the C_L and C_{H1} binding surfaces and were exposed by protein unfolding induced by the plastic surface.

To visualize cluster propagation in each independent sequence, we evaluated sequence appearance/disappearance transition and timing in the hit-clusters and miss-clusters (Figs. 2A & 2B). The cluster sequence transition was evaluated using the bit score parameter in the Basic Local Alignment Search Tool (BLAST) (*15*). It plotted the distance between the V gene region of each sequence and the ancestral germinal V gene sequence. The bit score increases with similarity of the query sequences to the reference. All hit-clusters had negative slopes for bit score vs. time of sequence appearance. Thus, the sequences in the clusters interacting with the antigen continually changed and became more remote from the ancestral sequence during immunization. By contrast, five of the miss-clusters (1, 3, 8, 11, and 14) showed positive slopes for bit score vs. time of sequence appearance. Clusters 6, 9, 12, and 13 exhibited slightly negative slopes for bit score vs. time of sequence appearance (Fig. 2B). Negative bit score slopes indicated sequence evolution and were good indicators of clusters that include antigen-binding clones.

Most hit-clusters displayed continuous sequence turnover which is displayed by the change in symbol color indicating the timing of the first appearance of each sequence (Fig. 2). This effect was clearly seen in clusters 7, 10, 15, and 16. Cluster 2 disappeared at the late immunization stage even though it predominated at the early immunization stage. In cluster 5, an early sequence occurred at week 1, reappeared at weeks 7 and 9, and disappeared thereafter (yellowish triangles). Sequences that had predominated between weeks 7 and 10 had persisted after 12 weeks. However, new sequences also emerged. Sequence turnover was only evident for miss-clusters 9 and 11. New sequences replaced old ones especially after week 12. The same was true for hit-cluster 5. Distinct sequence turnover may be a hallmark of clusters affected by the immune response. However, it was difficult to detect in clusters 5, 9, or 11.

The Y-axis intercept at the onset of the experiment (initial bit score) may pertain to progression of the immune response against a new antigen. For the hit-clusters and miss-clusters, the initial bit scores were 446 ± 90 and 319 ± 58, respectively. The alpaca was not pre-exposed to the antigens used in this experiment. Therefore, the hit-cluster sequences should not have been optimized before immunization and should have resembled the ancestral germinal V gene sequences. By contrast, clusters with low initial bit scores were unresponsive to the immunized antigens as they might have consisted of sequences that had matured before immunization.

Four clusters including the empirically identified clones were analyzed by the same plot (Fig. 2C). Negative bit score slopes, sequence turnover, and high initial bit scores were observed in all cases. These findings underscore the relationships among the hit-clusters and the foregoing criteria.

We sought clusters matching these criteria among those with low maximum appearances. We manually selected clusters 93, 103, 126, 139, 143, 175, 245, and 275 (maximum appearance range: 0.2–0.03%). Their bit score slopes were negative and their initial bit scores were very high (> 400) (Figs. 3 and S3). Newly appearing sequences predominated on a weekly basis and there was sequence turnover. Hence, we had selected putative hit-clusters. The sequences at the tips of the phylogenetic trees constructed for these clusters were prepared as antibody proteins (Table S2). Antigen binding was tested by ELISA and SPR. The VHH clones of clusters 103, 126, 139, 143, 175, 245, and 275 bound the antigen. ELISA and SPR indicated no interaction between the antigen and the clones from cluster 93.

Comparison of VHH clones in the same cluster

To elucidate the effects of accumulating somatic hypermutation during immunization, we analyzed the antigen-binding affinity of the clones in cluster 7. The sequences of clusters 33 and 120 exhibited high similarity with one amino acid length difference. Thus, they were combined as cluster 33+120 and their clones were examined. The antibodies in these clusters bound F_{ab} from trastuzumab and ranibizumab (Fig. S2). We selected various sequences appearing at different times and positioned at different branches in the phylogenetic tree and produced them as antibody proteins (Fig. 4). Fourteen and six sequences were examined for clusters 7 and 33+120, respectively (Table S2). Trastuzumab F_{ab} was used in SPR evaluation of antibody-antigen interactions (Fig. S4).

Differences between the sequences and the root sequence increased as immunization progressed. The root sequence was closest to the ancestral germline V sequence and exhibited the highest bit score in the cluster. The phylogenetic tree appeared to expand randomly, especially remarkable in the cluster 7. In most cases, the clones near the root of the phylogenetic tree had lower affinities than those further away from the root. Therefore, the antigen affinity tended to increase with decreasing bit score or increasing numbers of mutations. However, clones S675 (L2848) and L2754 in cluster 7 had high binding affinity and relatively high bit scores as well (Fig. 4C). Both clones were located on branches relatively close to the root but present at the tips as well. The positional characteristics of the clones in the phylogenetic tree should also be established to avoid missing high-affinity antibodies.

There were wide variations in clone affinity in clusters 7 and 33+120 after the final antigen injection at week 10 (Figs. 4E and 4F). In cluster 33+120, the binding affinities of the clones that emerged after week 10 were higher than those of the clones that appeared before week 10. In cluster 7, the clones that emerged before and after week 10 had overlapping binding affinities. L15236 appeared once at week 7 and had higher binding affinity than S709 (L25474) or L2487 (S2363). The latter two first appeared at week 12. The K_D of the clones appearing after week 10 in clusters 7 and 33+120 were in the ranges of 10^{-8}–10^{-10} and 10^{-9}–10^{-10}, respectively.

Discussion

With certain exceptions, the repertoire development history of the sequences in the immune-responsive antibody clusters exhibited a distinct time-dependent pattern. The sequences continuously developed and accumulated diversity throughout immunizations greater than 10 weeks in duration. Many sequences and clusters seems to be recognized as potent by alpaca immune system, suggesting the robustness of humoral immunity at least

for the antigen we used. Furthermore, the sequences showed intensive turnover. The older sequences in the hit-clusters became extinct and were superseded by newly emerged sequences. It is unclear whether antibody affinity maturation during immunization causes antibody sequences to converge. Sequence divergence might have increased at the early stage and a few of the best-adapted sequences were selected at the late stage. However, the present study indicated that the sequences in the hit-clusters continuously changed during immunization and even 4 weeks after the final antigen injection. Prolonged sequence turnover suggests that antibody affinity maturation was ongoing and the hit-cluster sequences did not converge within the experimental period of 14 weeks.

The bit score plot is an excellent tool for identifying hit-clusters. It contains the frequency of appearance, the timing of the first appearance, and the bit score of each sequence within the same cluster. Typical patterns were observed for clusters 7, 15, and 16 (Fig. 2A) which had high maximum bit scores. Thus, they started from the sequence nearest that of the germ line. Over time, the sequence generation alternated and the bit score decreased. Hence, affinity maturation progressed. We selected eight clusters with relatively low abundance based on the bit score slopes, sequence turnover, and initial bit scores. Seven of these included clones that bound the immunized antigen and were designated hit-clusters. The overall ratio of hit-clusters to all clusters is unknown. The ratio for the top 16 clusters except cluster 4 was 6:15. The immune response increased antigen-responding antibody expression. Consequently, the hit-cluster ratio may increase with cluster appearance frequency. For these reasons, the proportion of 7/8 hit-clusters with relatively low frequency was significant in hit-cluster prediction.

When a hit-cluster was identified by bio-panning or bit score plot analysis, VHH clones with high affinity for the target antigen could be selected using the phylogenetic tree of the identified cluster. Our data suggested that the clones at the tip of the phylogenetic tree had relatively stronger antigen-binding ability than those at the root of the phylogenetic tree. Hsiao et al. found that the antigen-binding affinity increased with somatic mutation load in antibodies from immunized mice. Nevertheless, this trend did not apply to all mutants with high somatic mutation loads (*16*). Here, the correlation between affinity and bit score indicated the mutation load and was imperfect for clones S675 (L2848) and L2754 in cluster 7 (Fig. 4C). The impact of each mutation on binding affinity improvement should be unique. Therefore, mutation load may not always indicate an increase in affinity.

Clones with diverse affinity persisted even after the final antigen injection. In addition, clones with relatively high binding affinity were extinct by the middle of immunization. There may be a certain degree of antigen affinity tolerance and there may be no strict affinity cutoff even after substantial immunization progress. *In vivo* affinity maturation is a selection process driven by high affinity and other factors such antibody expression efficiency.

Clone S1139 (L27626) in cluster 7 and clone L54 in cluster 33 were identified either by bio-panning or a combination of bio-panning and NGS analysis (*13*, *14*). These empirically identified clones had low antigen affinity and were localized to the root of the phylogenetic tree. However, the clones at the tips of the phylogenetic tree had relatively higher antigen affinity than the empirically identified clones. Nevertheless, not all experimental clones have low antigen affinity. Turner *et al.* (*17*) showed that bio-panning-derived clones within the same clusters shared similar antigen affinity. The bit scores for

the other empirically identified clones in clusters 15, 69, 99, and 210 were 366, 470, 483, and 353, respectively, and their highest bit scores were 433, 486, 499, and 387, respectively (Figs. 2A and 2C). These empirically identified clones were remote from the root of the phylogenetic tree.

In the present study, we showed that VHH sequences responding to injected antigen were continuously generated and persisted throughout 3 months of immunization. Clone development was associated with increasing target antigen affinity. These findings were consistent with the concept of immune affinity maturation. To the best of our knowledge, this work is the first to describe time-resolved antibody affinity maturation and experimentally evaluate various antibodies appearing at different times. Based on the foregoing results, it is feasible to predict VHH hit-clusters and the high-affinity VHH clones within them based on their bit score plots and their positions on the phylogenetic tree. A combination of bio-panning and NGS analysis successfully identified binding VHH clones not detected by conventional bio-panning alone (*13*, *17, 18*). Unlike prior studies, the prediction method proposed here requires no bio-panning. Here, we utilized an alpaca single-domain VHH to avoid the technical difficulties associated with producing several conventional antibodies and their fragments using correctly matched light-chain and heavy-chain combinations. Nevertheless, it remains to be determined whether the discoveries herein are applicable to conventional antibody and immune systems in other animals. Therefore, future research should use other camelid species, antigens, and animal species to validate the mechanism of affinity maturation in humoral immunity.

References

1. C. Berek, C. Milstein, Mutation drift and repertoire shift in the maturation of the immune response. *Immunol Rev* **96**, 23-41 (1987).
2. C. Parola, D. Neumeier, S. T. Reddy, Integrating high-throughput screening and sequencing for monoclonal antibody discovery and engineering. *Immunology* **153**, 31-41 (2018).
3. G. Georgiou *et al.*, The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol* **32**, 158-168 (2014).
4. S. T. Reddy *et al.*, Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat Biotechnol* **28**, 965-969 (2010).
5. V. Greiff *et al.*, Systems analysis reveals high genetic and antigen-driven predetermination of antibody repertoires throughout B cell development. *Cell Reports* **19**, 1467-1478 (2017).
6. M. M. VanDuijn, L. J. Dekker, I. W. F. J. van IJcken, P. A. E. Sillevis Smitt, T. M. Luider, Immune repertoire after immunization as seen by next-generation sequencing and proteomics. *Front Immunol* **8**, 1286 (2017).
7. T. M. Gierahn *et al.*, Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* **14**, 395-398 (2017).
8. L. D. Goldstein *et al.*, Massively parallel single-cell B-cell receptor sequencing enables rapid discovery of diverse antigen-reactive antibodies. *Commun Biol* **2**, 304 (2019).
9. B. Wang *et al.*, Functional interrogation and mining of natively paired human VH:VL antibody repertoires. *Nat Biotechnol* **36**, 152-155 (2018).
10. C. Hamers-Casterman *et al.*, Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446-448 (1993).
11. S. Muyldermans, A guide to: generation and design of nanobodies. *FEBS J*, (2020).

12. Y. Liu, H. Huang, Expression of single-domain antibody in different systems. *Appl Microbiol Biotechnol* **102**, 539-551 (2018).
13. N. Miyazaki *et al.*, Isolation and characterization of antigen-specific alpaca (Lama pacos) VHH antibodies by biopanning followed by high-throughput sequencing. *J Biochem* **158**, 205-215 (2015).
14. A. Rafique *et al.*, Efficient screening and design of variable domain of heavy chain antibody ligands through high throughput sequencing for affinity chromatography to purify fab fragments. *Monoclon Antib Immunodiagn Immunother* **38**, 190-200 (2019).
15. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J Mol Biol* **215**, 403-410 (1990).
16. Y. C. Hsiao *et al.*, Immune repertoire mining for rapid affinity optimization of mouse monoclonal antibodies. *MAbs* **11**, 735-746 (2019).
17. K. B. Turner *et al.*, Next-generation sequencing of a single domain antibody repertoire reveals quality of phage display selected candidates. *PLoS One* **11**, e0149393 (2016).
18. P. Deschaght *et al.*, Large diversity of functional nanobodies from a camelid immune library revealed by an alternative analysis of next-generation sequencing data. *Front Immunol* **8**, 420 (2017).
19. T. H. Jukes, C. R. Cantor, in *Mammalian Protein Metabolism,* H. N. Munro, Ed. (Academic Press, 1969), pp. 21-132.
20. D. S. Doering, P. Matsudaira, Cysteine scanning mutagenesis at 40 of 76 positions in villin headpiece maps the F-actin binding site and structural features of the domain. *Biochemistry* **35**, 12677-12685 (1996).
21. Y. Hagihara, T. Matsuda, N. Yumoto, Cellular quality control screening to identify amino acid pairs for substituting the disulfide bonds in immunoglobulin fold domains. *J Biol Chem* **280**, 24752-24758 (2005).
22. H. Edelhoch, Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**, 1948-1954. (1967).

Acknowledgments

Funding:
JSPS KAKENHI Grant No. 20K07009 (YAO)
JSPS KAKENHI Grant No. 20K20599 (TM, YAO, and YH)
JSR Corporation (TM, YAO, NM, YI, and YH)

Author contributions:
Conceptualization: TM, YAO, NM, YM, TF, YI, and YH
Formal analysis: TM
Methodology: TM, YAO, and YH
Investigation: TM, YAO, LKK, NK, NM, YM, TF, YI, and YH
Visualization: TM, YAO, NK, NM, and YH
Funding acquisition: TM, YAO, NM, YI, and YH
Project administration: TM, YAO, and YH
Writing-original draft: TM, YAO, NK, NM, and YH
Writing-review and editing: TM, YAO, LKK, NK, NM, YM, TF, YI, and YH

Competing interests:

YM and TF of JSR Corporation participated in the conceptualization and investigation of this work. No other funding sources participated in study design, data collection, analysis, or interpretation or report writing.
Kyoto University, AIST, ARK Resource Co. Ltd., and Kagoshima University applied for patents (Nos. JP/2019/080434). TM, YAO, NM, TF, YI, and YH were listed as the inventors. Kyoto University applied for patents (Nos. WO/2020/213730). TM, YAO, NM, TF, YI, and YH were listed as the inventors.

Data and materials availability:
Antibody nucleotide sequence data are available in the DDBJ Sequenced Read Archive under accession Nos. DRR305325–DRR305354. SAMD00387596–00387610 correspond to the data for blood collected over weeks 0–14. All other data are available in the main text or the supplementary materials. All data used in the analysis are available for experimental reproduction purposes and for building on the present analysis. The materials used in this work are available under a Materials Transfer Agreement (MTA).

Figures and Tables

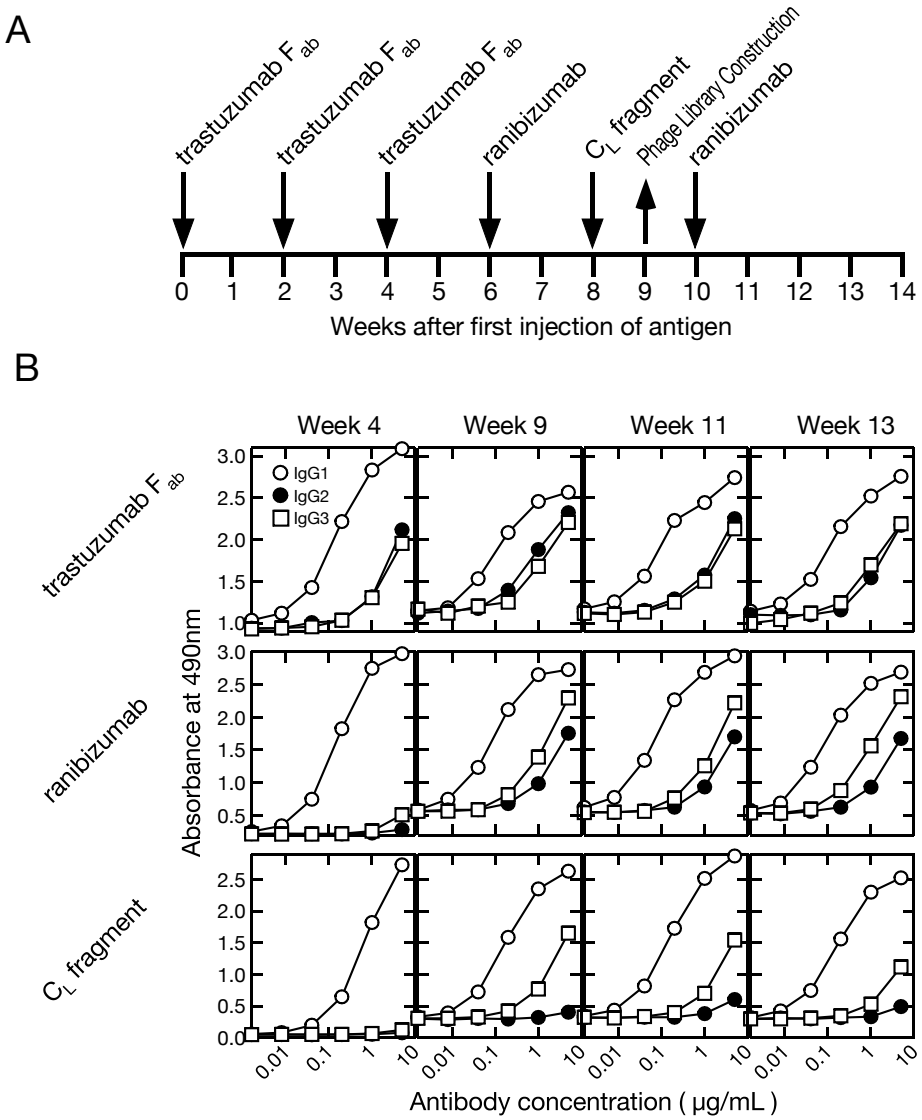


Fig. 1. Immunization and blood collection schedules (A) and time course of purified polyclonal alpaca antibody titer (B). (**A**) F_{ab} of trastuzumab, ranibizumab, and human *κ* C_L fragment antigens were injected into a single alpaca. Blood was collected once before immunization (week 0) and 14× after initiating immunization (weeks 1–14). A phage display screening library was prepared using blood collected at week 9. (**B**) Polyclonal IgG1 (conventional antibody), IgG2 (heavy-chain antibody with short hinge), and IgG3 (heavy-chain antibody with long hinge) were purified from blood collected at weeks 4, 9, 11, and 13. Titers were measured against F_{ab} of trastuzumab, ranibizumab, and human *κ* C_L fragment.

Fig. 1

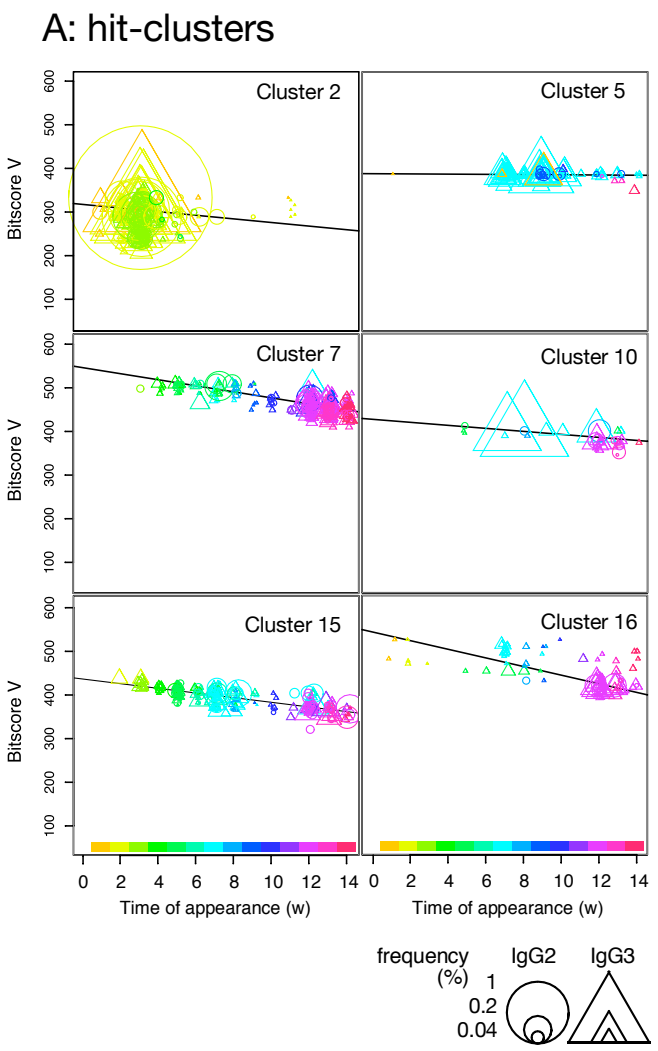

A: hit-clusters
Cluster 2
Cluster 5
Cluster 7
Cluster 10
Cluster 15
Cluster 16
Bitscore V
Time of appearance (w)
frequency (%)
IgG2
IgG3
1
0.2
0.04

Fig. 2

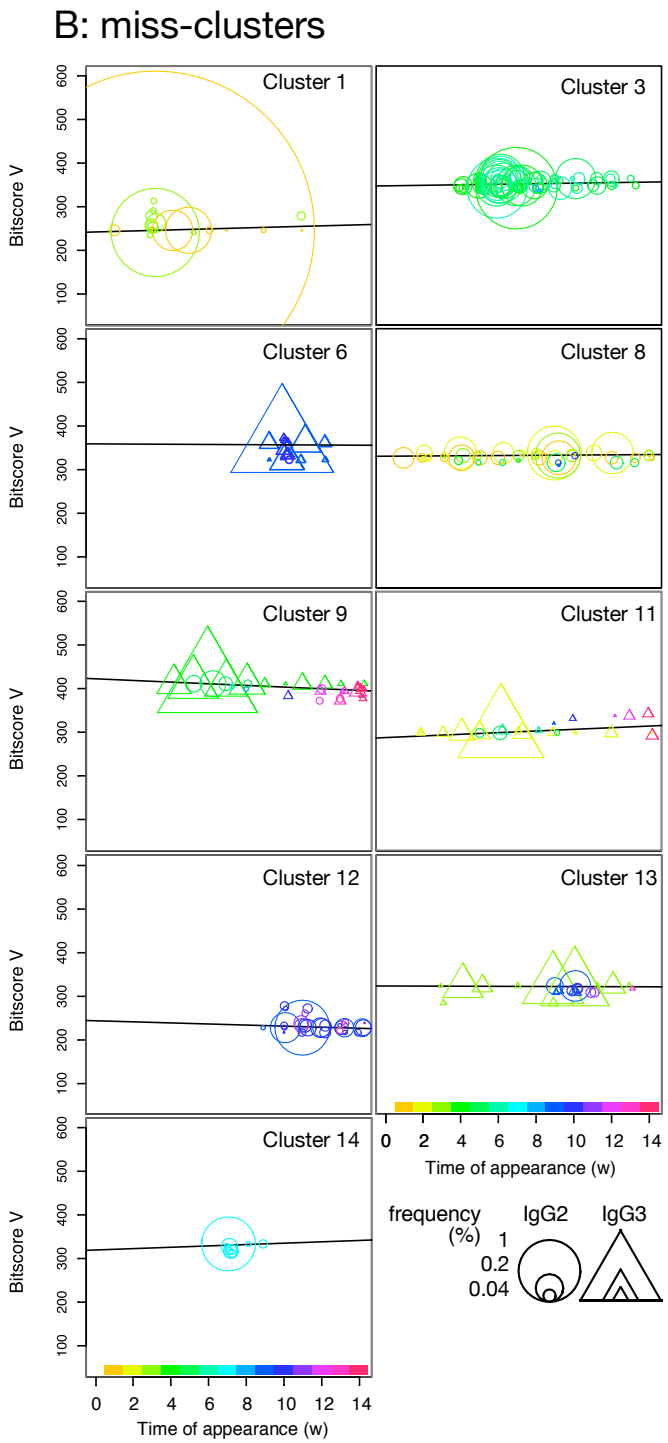

B: miss-clusters
Cluster 1
Cluster 3
Cluster 6
Cluster 8
Cluster 9
Cluster 11
Cluster 12
Cluster 13
Cluster 14
Bitscore V
Time of appearance (w)
frequency (%)
IgG2
IgG3
1
0.2
0.04

Fig. 2

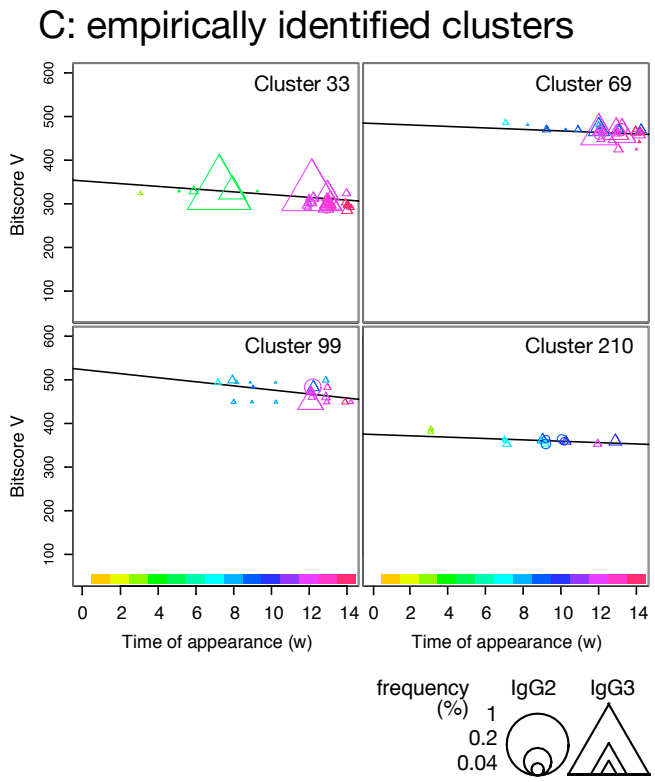


Fig. 2. Plot of sequence bit scores vs. time of appearance of sequences in hit-clusters (A), miss-clusters (B), and empirically identified clusters (C). Symbol colors indicate times of first appearance of each sequence. Colors corresponding to times of first appearance are indicated in color bar at panel bottom. Circles and triangles indicate sequences observed in short-hinge (IgG2) and long-hinge (IgG3) antibody, respectively. Symbol size corresponds to weekly clone frequency in IgG2 and IgG3 sequences.

Fig. 2

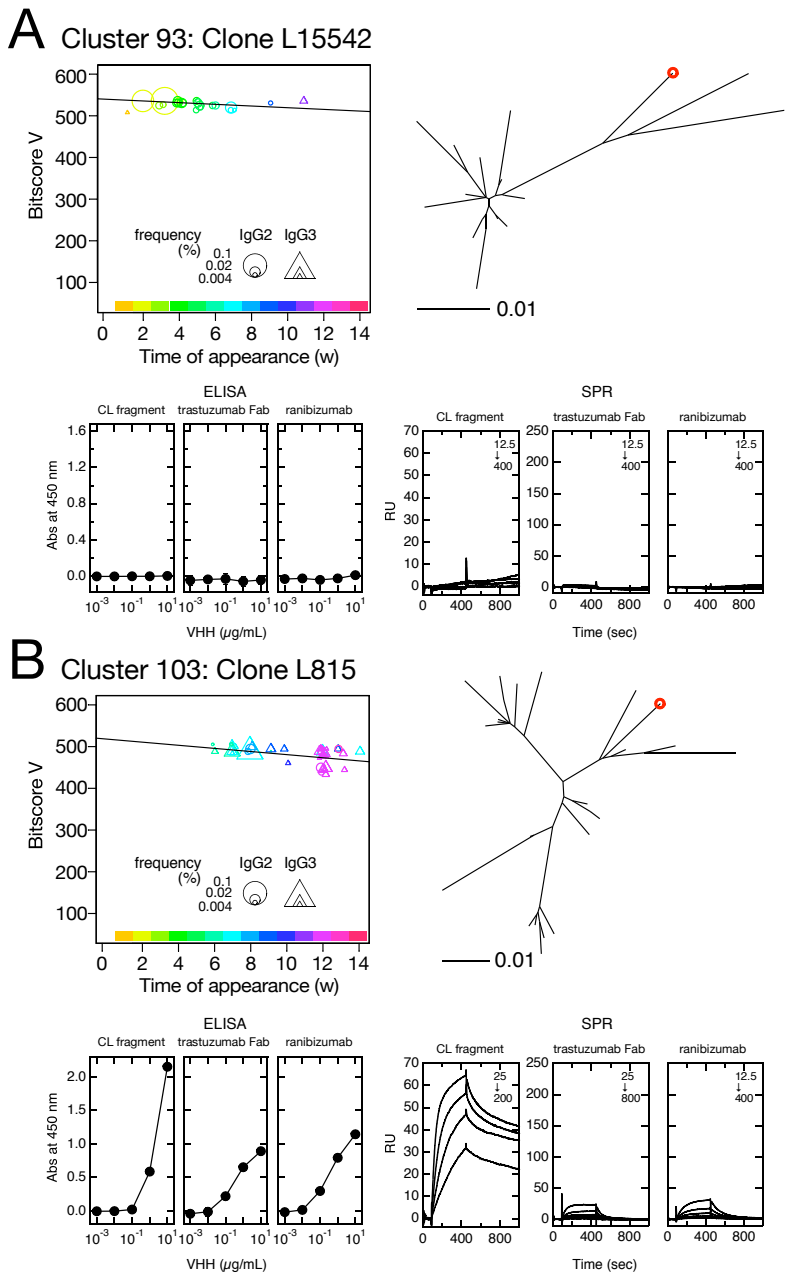
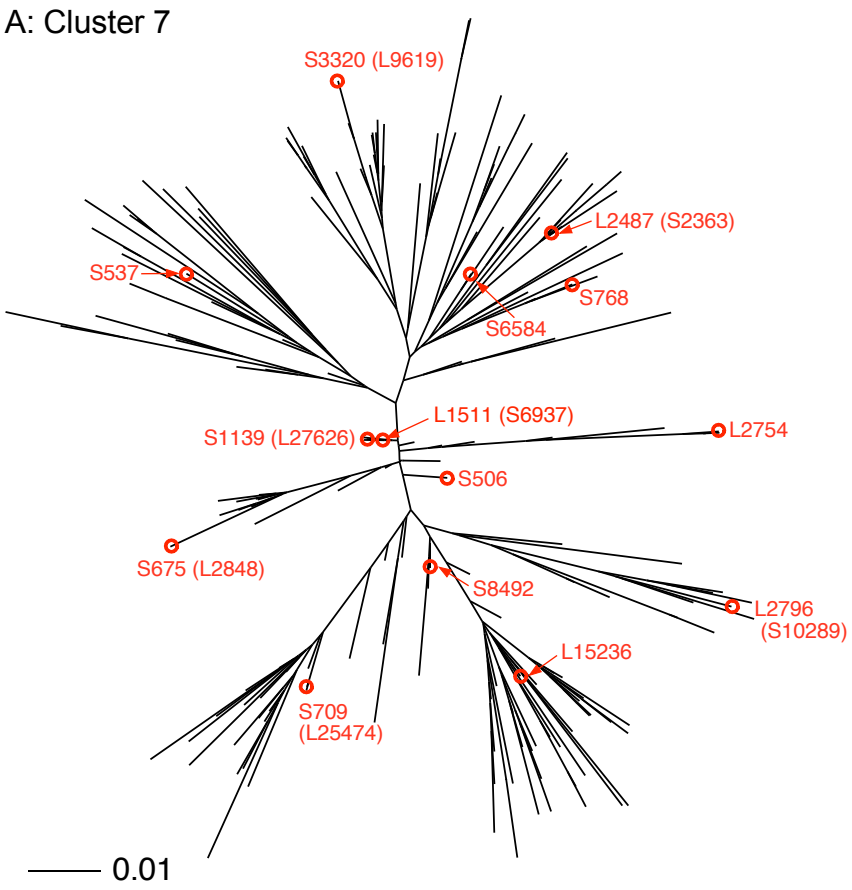


Fig. 3. Clusters with relatively low maximum percentages of appearance. Clusters 93 (**A**) and 103 (**B**) were selected based on their negative bit score slopes, distinct sequence turnover, and high initial bit scores depicted in bit score plot (upper left panels). Maximum percentages of appearance of clusters 93 and 103 were 0.23 and 0.20, respectively. Position of selected VHH clone in phylogenetic tree is indicated by red circle (upper right panel). Phylogenetic trees were drawn using DNA sequences including clusters. Bars below phylogenetic trees indicate distance = 0.01 calculated by JC69 (*19*) and corresponding to ~1% nucleotide sequence difference. Symbol size in bit score plot corresponds to weekly clone frequency in IgG2 and IgG3 sequences. Affinities of VHH clone for immobilized human κ C_L fragment (left), F_{ab} of trastuzumab (middle), and ranibizumab (right) are depicted by ELISA (lower left three panels) and SPR (lower right three panels). Values inside SPR panels indicate concentration ranges for VHH clones measured as analytes.


A: Cluster 7
S3320 (L9619)
L2487 (S2363)
S537
S768
S6584
L1511 (S6937)
S1139 (L27626)
L2754
S506
S675 (L2848)
S8492
L2796
(S10289)
L15236
S709
(L25474)
0.01

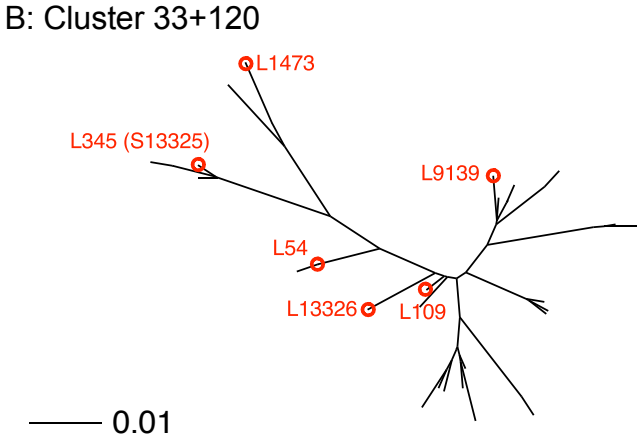

B: Cluster 33+120
L1473
L345 (S13325)
L9139
L54
L13326
L109
0.01

Fig. 4

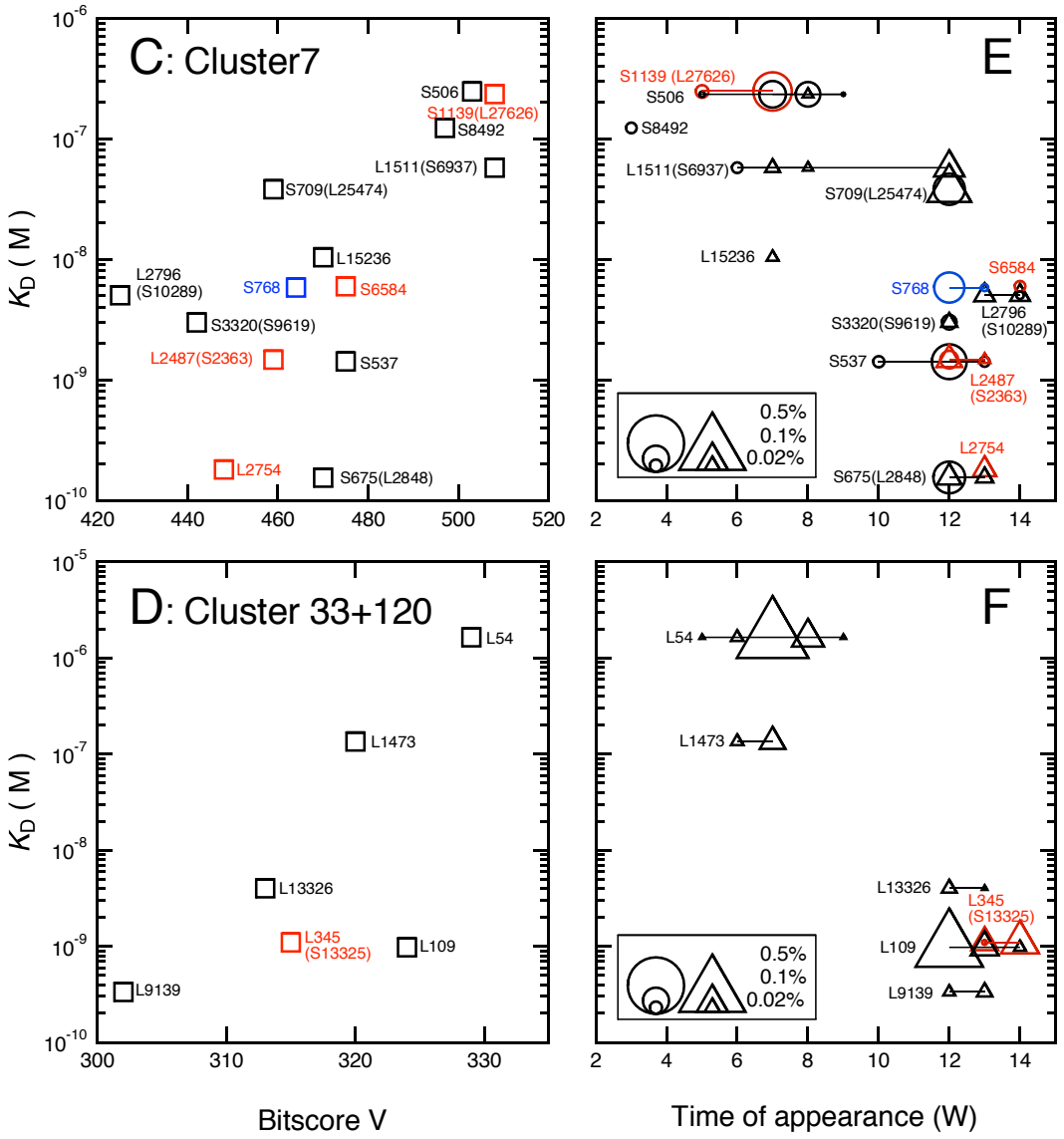


Fig. 4. Analysis of VHH clones in same cluster. **A & B:** Phylogenetic tree of clusters 7 (**A**) and 33+120 (**B**). S1139 (L27626) in cluster 7 and L54 in cluster 33 are empirically identified clones. **C & D:** Relationship between bit score (V) and dissociation constant (K_D) vs. F_{ab} of trastuzumab of clones in clusters 7 (**C**) and 33+120 (**D**). For clarity, symbols used for clones with similar K_D are indicated by different colors. Clone names are shown besides symbols. **E & F:** Time of appearance-dependent K_D change vs. F_{ab} of trastuzumab in clusters 7 (**E**) and 33+120 (**F**). Circle and triangle indicate clones derived from short-hinge (IgG2) and long-hinge (IgG3) antibody, respectively. Symbol size indicates weekly clone frequency in IgG2 and IgG3 sequences. Symbols of clones with similar K_D are indicated by different colors. Clone names are shown next to symbols.

Fig. 4

Supplementary Materials

Materials and Methods

Figs. S1–S4

Tables S1–S2