

Editor's Summary

Antibodies Act Their Age

One of the main advantages of the immune system in fighting infection is its ability to diversify—antibody and T cell receptor genes physically rearrange, creating a repertoire of potential responses that can be called upon and expanded if needed. However, the very diversity of this repertoire is what makes immune responses hard to study. Although we can know how any individual B cell or antibody responds to stimulation, getting the big picture is much more difficult. Adding in another variable, such as time, further complicates things. Now, Jiang *et al.* use high-throughput long read sequencing to characterize the human antibody response after influenza vaccination.

People of age are thought to have altered immune systems compared to younger individuals. However, why exactly how the antibody repertoire changes with age remains unclear. By analyzing more than 5 million antibody heavy chain sequences, the authors were able to compare isotype diversity, lineage structure, and mutational activity in differently aged populations. They found that the elderly have fewer lineages, with reduced diversity, compared with younger subjects; however, the antibodies present before vaccination had higher levels of mutation. These same techniques can be used to study individual-specific immune responses and may aid in optimizing vaccination in the future.

A complete electronic version of this article and other services, including high-resolution figures, can be found at:

<http://stm.sciencemag.org/content/5/171/171ra19.full.html>

Supplementary Material can be found in the online version of this article at:

<http://stm.sciencemag.org/content/suppl/2013/02/04/5.171.171ra19.DC1.html>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

INFLUENZA

Lineage Structure of the Human Antibody Repertoire in Response to Influenza Vaccination

Ning Jiang,^{1*†} Jiankui He,^{1*‡} Joshua A. Weinstein,^{2*§} Lolita Penland,¹ Sanae Sasaki,^{3,4} Xiao-Song He,^{4,5} Cornelia L. Dekker,⁶ Nai-Ying Zheng,⁷ Min Huang,⁷ Meghan Sullivan,⁷ Patrick C. Wilson,⁷ Harry B. Greenberg,^{3,4,5} Mark M. Davis,^{3,8} Daniel S. Fisher,^{9,10} Stephen R. Quake^{1,2,8,9¶}

The human antibody repertoire is one of the most important defenses against infectious disease, and the development of vaccines has enabled the conferral of targeted protection to specific pathogens. However, there are many challenges to measuring and analyzing the immunoglobulin sequence repertoire, including that each B cell's genome encodes a distinct antibody sequence, that the antibody repertoire changes over time, and the high similarity between antibody sequences. We have addressed these challenges by using high-throughput long read sequencing to perform immunogenomic characterization of expressed human antibody repertoires in the context of influenza vaccination. Informatic analysis of 5 million antibody heavy chain sequences from healthy individuals allowed us to perform global characterizations of isotype distributions, determine the lineage structure of the repertoire, and measure age- and antigen-related mutational activity. Our analysis of the clonal structure and mutational distribution of individuals' repertoires shows that elderly subjects have a decreased number of lineages but an increased prevaccination mutation load in their repertoire and that some of these subjects have an oligoclonal character to their repertoire in which the diversity of the lineages is greatly reduced relative to younger subjects. We have thus shown that global analysis of the immune system's clonal structure provides direct insight into the effects of vaccination and provides a detailed molecular portrait of age-related effects.

INTRODUCTION

The adaptive immune system produces a large and diverse set of antibodies, each with an individual evolutionary and clonal history. This so-called antibody repertoire protects each individual against insults such as infection and cancer and responds to vaccination with B cell proliferation in response to the antigenic stimulation. Hybridomas and antigen-specific fluorescence-activated cell sorting-based analysis have provided us with much insight on how the immune system generates the complex and diverse immune response required to protect the body from the wide variety of potential pathogens (1–3). However, these methods have not been sufficient to make global and unbiased characterizations of the clonal structure of the immune system of a particular individual, which could provide insights into how the diversity and clonal structures vary between individuals, with age or gender, and in response to specific antigen stimulation (4). With respect to antigen stimulation, although there is a great deal of data that have been

obtained by sorting antigen-specific B cells, there is less information on the effect of the antigen on the global response of the antibody repertoire (5–10).

We and others have begun applying high-throughput sequencing techniques to the immunogenetic characterization of antibody repertoire (11–17). Our previous work focused on zebrafish as a model organism, which enabled us to perform deep sequencing to exhaust the repertoire diversity in a manner that was independent of the physiology of the organism, that is, independent of where the B cells were residing (11). This work revealed that the repertoire of individuals has a surprisingly high fraction of shared sequences and a universal structure and that the balance between determinism and stochasticity in the repertoire is tilted more toward determinism both in early development and in the primary repertoire of mature organisms than had previously been suggested (15). Others have used similar approaches to measure the amount of residual disease from lymphoma B cell clones (12), to study gene segment frequency after lymphocytic ablation (16), and to bypass cloning and directly synthesize antibody from mining the high-throughput sequencing data acquired by using bone marrow plasma cells from immunized mice (13). Attempts to use this approach to study vaccination have not been able to resolve lineage relationships and have not demonstrated a functional link between repertoire and immune response (18). Here, we address the question of how the human immune repertoire responds to specific antigen stimulation, in particular by influenza vaccination. We determine the lineage structure of the repertoire before and after vaccination and demonstrate that some sequences in the repertoire correspond to vaccine-specific immunoglobulins. We further observe age-related changes in antibody isotype composition, lineage diversity and structure, as well as mutational load, thereby offering a molecular characterization of defects in humoral immune response resulting from aging.

¹Department of Bioengineering, Stanford University School of Medicine, Stanford, CA 94305, USA. ²Biophysics Graduate Program, Stanford University, Stanford, CA 94305, USA.

³Department of Immunology and Microbiology, Stanford University School of Medicine, Stanford, CA 94305, USA. ⁴VA Palo Alto Health Care System, Palo Alto, CA 94304, USA.

⁵Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA.

⁶Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305, USA. ⁷Department of Medicine, University of Chicago Pritzker School of Medicine, Chicago, IL 60637, USA. ⁸Howard Hughes Medical Institute. ⁹Department of Applied Physics, Stanford University, Stanford, CA 94305, USA. ¹⁰Department of Biology, Stanford University, Stanford, CA 94305, USA.

*These authors contributed equally to this work. [†]Present address: Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX 78712, USA.

[‡]Present address: Department of Biology, South University of Science and Technology of China, Shenzhen 518055, China.

[§]Present address: Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.

[¶]To whom correspondence should be addressed. E-mail: quake@stanford.edu

RESULTS

We analyzed antibody repertoires from peripheral blood drawn from 17 human volunteers who were immunized with the 2009 or 2010 seasonal influenza vaccines (table S1). These volunteers were recruited from three age groups—children (8 to 17 years of age), young adults (18 to 30 years of age), and elderly individuals (70 to 100 years of age)—and were randomly vaccinated with either trivalent inactivated influenza vaccine (TIV) or live attenuated influenza vaccine (LAIV), except for subjects in the 70 to 100 years group who could only receive TIV (table S1) (9). TIV and LAIV contain antigenically equivalent virus strains; however, LAIV is made of live attenuated viruses that are capable of limited proliferation after intranasal administration and are expected to induce a stronger mucosal immune response than TIV (19). The study included three pairs of identical twins to have repertoire control experiments with identical genetic background; these twins were in the age group of 8 to 17 and were randomly selected to receive either TIV or LAIV within a twin pair. Blood samples were collected from each participant at three time points: day 0 before vaccination (visit 1), day 7 or 8 (visit 2), and day 28 (± 4 , visit 3) after vaccination. Peripheral blood mononuclear cells (PBMCs) were collected at both visit 1 and visit 3. Naïve B (NB) cells and plasmablasts (PBs) were sorted from visit 2 blood samples by flow cytometry.

Reduction of relative immunoglobulin M abundance after vaccination decreases with age

All five isotypes were detected in all the samples processed but with different relative amounts. Immunoglobulin M (IgM), IgA, and IgG were more abundant than IgD and IgE, which together account for less than 4% of all sequencing reads (Fig. 1A). Most NB cells express IgM on their membrane and then, upon antigen stimulation, undergo an activation process that changes their constant region from IgM to

other isotypes and increases the antibody transcript copy numbers in each cell (20–22). We tracked the changes of isotype distribution between visit 1 and visit 3 and noticed a decrease of relative IgM abundance for all volunteers except one (Fig. 1B and table S4). On average, relative IgM usage decreased $11.1 \pm 3.2\%$ (SEM) (age 8 to 17), $6.5 \pm 2.4\%$ (SEM) (age 18 to 30), and $6.0 \pm 2.9\%$ (SEM) (age 70 to 100) at visit 3 (Fig. 1B, black lines). An independent measurement by digital polymerase chain reaction (PCR) (23, 24) was used to verify the relative isotype abundance in visit 1 and 3 samples and the reduction of relative IgM abundance from visit 1 to visit 3 (fig. S3). This decrease coincided with an increase of relative IgA and IgG abundance (fig. S3), suggesting that a portion of the NB cells may have undergone isotype switching and the antibody transcript copy numbers of these isotypes may also have increased as a result of the vaccine stimulation. This interpretation is supported by flow cytometry data, which did not show changes in the relative abundances of IgM-expressing cells between visit 1 and visit 3. Therefore, the large difference in relative IgM usage is due to antibody transcript copy number changes as a result of antigen stimulation. We also directly observed isotype-switched lineages in a small fraction of the sequence data. These sequences contained common complementarity determining region 3 (CDR3) sequences and had extensive mutations throughout the variable regions (fig. S5), which suggests that they are not template-switched PCR artifacts. Moreover, the number of lineages containing isotype switches decreased with age (fig. S4), which is consistent with our observation that reduction of relative IgM transcript abundance from visit 1 to visit 3 decreases with age regardless of vaccine types (Fig. 1C). Although LAIV receivers have less change in relative IgM usage than individuals who received TIV, there is a strong age dependence of isotype relative abundance change in TIV receivers—children who received TIV were more likely to have an increased relative IgA usage compared to young adults ($P = 0.03$, Mann-Whitney U test) or the elderly ($P = 0.05$, Mann-Whitney U test) (fig. S3).

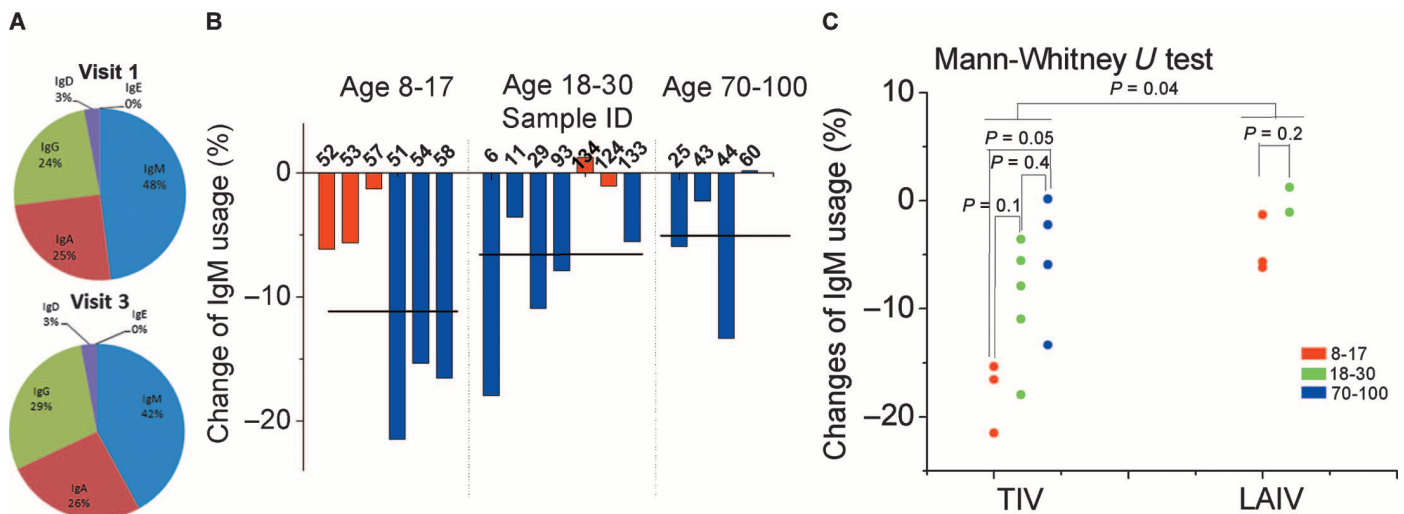


Fig. 1. Antibody isotype distribution changes after vaccination. **(A)** Antibody isotype composition in PBMCs at visit 1 (before vaccination, top) and visit 3 (28 \pm 4 days after vaccination, bottom) averaged for all subjects. **(B)** Percent change in individual's relative IgM usage in PBMCs from visit 1 to visit 3. The subject IDs were labeled on horizontal axis. **(C)** Comparison of relative change in IgM in different age group and vaccine types. P was

calculated by Mann-Whitney U test (three samples for TIV receivers of age 8 to 17, five samples for TIV receivers of age 18 to 30, four samples for TIV receivers of age 70 to 100, three samples for LAIV receivers of age 8 to 17, and two samples for LAIV receivers of age 18 to 30). Red, LAIV receivers; blue, TIV receivers. Percent change = (IgM reads in visit 1/total reads in visit 1) – (IgM reads in visit 3/total reads in visit 3).

Single linkage clustering enables informative definition of lineages

To dissect differences between somatic hypermutations and lineages and analyze the detailed mutations within a lineage, we developed a clustering scheme that focused on the CDR3 of the antibody sequence, which covers the region between the end of the V and the beginning of the J gene segments. We first converted the nucleotide sequences into amino acid sequences for each read. Translation rescue was performed for out-of-frame sequences that were mostly due to sequencing errors in the V, D, or J segments (fig. S1). To set the clustering threshold, we analyzed the amino acid distance between reads in the CDR3 region. The resulting distribution showed two distinct peaks: the first is at 1 amino acid and the other covers 4 to 10 amino acids (fig. S6). This suggests that the first peak contained sequencing reads that were mutations within a lineage and the second peak contained sequencing reads that had distinct CDR3 sequences that were generated during the VDJ recombination process. The amplitude of these two peaks changed between different samples that were collected at different time points and had varying NB and PB cell fractions. Sorted NB cells had the lowest first peak and highest second peak, whereas sorted PBs has the highest first peak and the lowest second peak. PBMC samples from visits 1 and 3 fell in between NB cells and PBs, whereas visit 2 PBMCs (available only for selected subjects) were similar to visit 2 PBs. These trends are consistent with dynamics of antibody-mediated immune response and distribution of NB cells and PBs in peripheral blood. These trends further support the interpretation that the first peak is due to mutation and the second peak is due to junctional diversity.

This distribution provides a natural threshold when clustering and was used to group sequences according to their lineage identity. We also performed clustering directly on nucleotide sequences with varying thresholds (fig. S15). Here, “lineage” refers to antibody sequences that originated from the same VDJ recombination event and have the same junctional sequence, but may be further diversified because of antigen stimulation and somatic hypermutation. We clustered all sequencing reads from sorted PBs at visit 2 using 1 amino acid difference at CDR3 as a threshold. This means that two sequences will be grouped into the same lineage if they are in the same V and J family and their protein sequence in the CDR3 region differs by no more than 1 amino acid. Using these lineage data, one can construct a graphical representation of the clonal structure of the immune repertoire (Fig. 2 and fig. S7).

The central functional question regarding these informatically defined sequence lineages is to what extent they include influenza-specific antibodies. To examine this, we amplified influenza-specific antibody sequences from single-sorted PBs for two of the subjects in the 70- to 100-year-old group: 017-043 and 017-044. We then expressed monoclonal antibodies according to these sequences and verified their binding to each of the three virus strains in the vaccine. Eleven of 16 heavy chain sequences from single cell-cloned PBs were found within the lineages we measured, especially for the anti-vaccine high-affinity antibodies (Fig. 2 and table S6). For subject 017-044, the single cell-cloned sequences overlap with lineages containing smaller number of reads compared to sequences cloned from 017-043, where many single cell-cloned sequences are in the top lineage containing most of the reads (fig. S11). This may reflect structural differences in repertoire between these two subjects because one has a dominant lineage and the other one has a more even distribution (fig. S11). Together, these data confirm that the influenza-specific antibody responses are contained

within the globally measured immune repertoire sequences as well as the informatically defined lineages we derived from them.

Lineage structure analysis reveals distortion in some elderly subjects

Lineages belonging to PBs exhibit an apparent power-law distribution with a few lineages that dominate the repertoire, whereas those belonging to naïve cells do not (fig. S13). This is consistent with long-tailed distributions observed previously (11) and is the direct consequence of clonal expansion. The elderly have fewer lineages than other age groups both before (Fig. 3A) and after (fig. S12) vaccination, indicating an altered repertoire structure and potentially a smaller pool of diversity for the immune repertoire to draw upon in vaccine response.

Using the three parameters of diversity (unique protein sequences), average mutation, and number of reads in each lineage, one can visualize and compare the antibody repertoire in a quantitative manner (Fig. 4, A to F). In each individual, most of the lineages contain fewer than 10 reads and fewer than 2 unique amino acid sequences. The elderly vaccine recipients can be separated into two groups: One group had a distribution of lineages similar to the children (Fig. 4, A and B) and young adults (Fig. 4, C and D), and the other group had a very different distribution of lineages compared to the other age groups (Fig. 4, E and F). Elderly subjects in the second group had a few lineages that encompassed more than 80% of the reads. This is exemplified by subject 017-43 (Fig. 4F and figs. S8 and S11). Detailed sequence analysis revealed that 58% (subject 017-043) and 90% (subject 017-060) of the reads within the biggest lineage for these elderly were identical. This is consistent with the overall observation that influenza vaccination resulted in expansion of far fewer B cell lineages in the elderly compared to the other age groups (fig. S12A). This reduced clonal diversity when weighted with abundance may be related to a reduced antibody response to influenza vaccine in the elderly. Lineage analysis on IgG from visit 1 and 3 PBMCs also suggested that, in general, the elderly have a reduced B cell clonal diversity compared to the younger age groups (Fig. 3A and fig. S12B), which might explain the reduced clonal diversity in vaccine-activated B cells in the elderly.

Age affects somatic hypermutation and lineage diversity

One interesting question about reduced immune response to influenza vaccination in the elderly is whether the B cells that respond to the current vaccine had been primed by previous infections or vaccinations. If so, those B cells from the elderly will most likely be memory B cells that have a higher baseline mutation than responding B cells from younger volunteers, where most of these cells should be of naïve phenotype or relatively less antigen-experienced; therefore, they have fewer mutations. Another important question is whether those responding memory B cells in the elderly have the same ability to introduce new mutation upon antigen stimulation compared to responding B cells from young volunteers.

To answer these questions, we performed a detailed analysis of mutation statistics. Although 454 sequencing has a high error rate of about 1%, most of them are insertions and deletions (indels) (11, 15) and can be repaired (fig. S1). The substitution error rate (from sequencing and/or PCR) is estimated to be 0.065% per nucleotide (fig. S16 and Control library section in the Supplementary Materials). This is lower than the estimated somatic hypermutation rate, which is about 0.1% measured in nucleotides per cell division (25). Also, any B cells undergoing somatic hypermutation are likely to have several rounds of

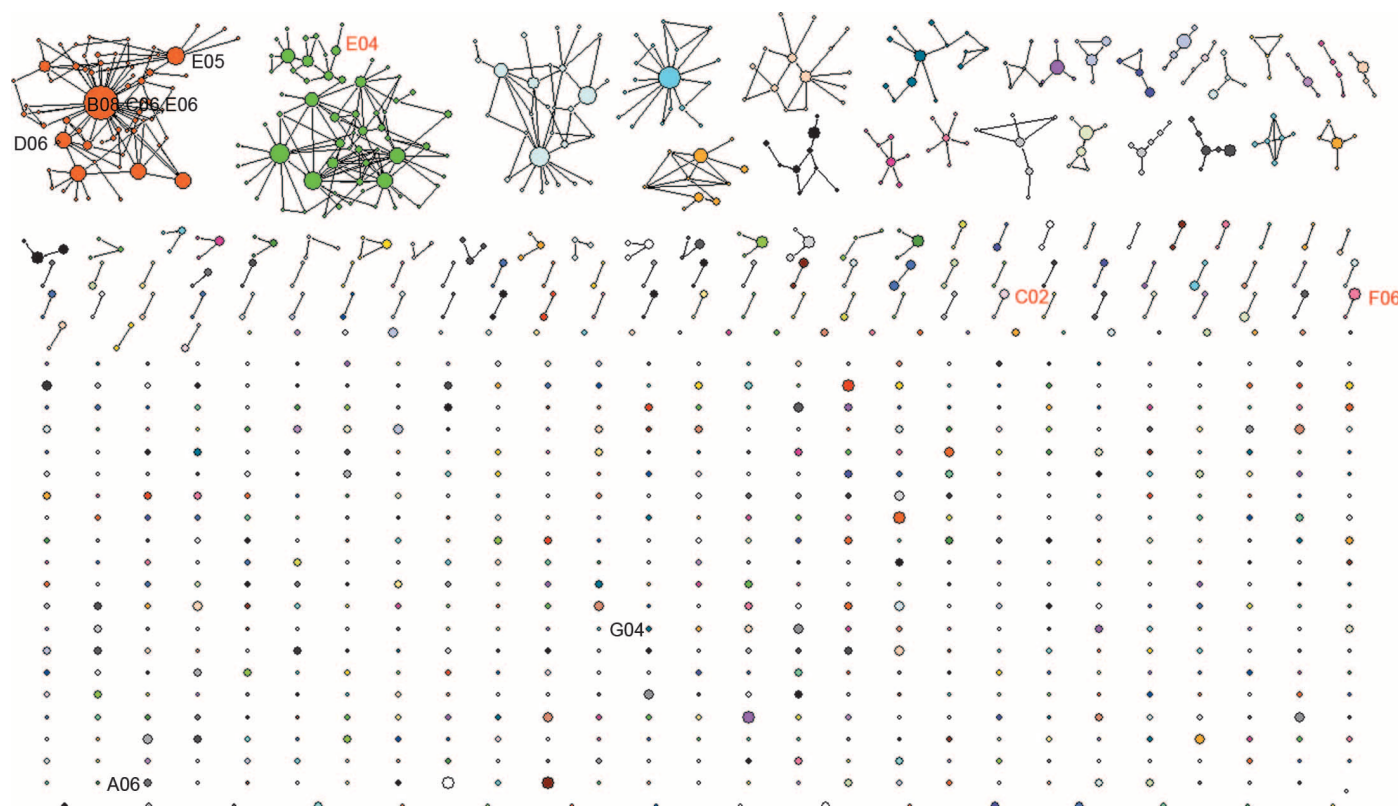


Fig. 2. Informatically defined lineages with influenza specificity. The intra- and interlineage structure of all IgG lineages visualized by sequencing the PBs sorted from blood sample collected at visit 2 (7 days after vaccination) from a volunteer in the 70- to 100-year-old group received TIV (subject 017-043). In this network representation, each cluster of dots connected by lines represents a lineage. Different colors were used to distinguish different lineages. Each dot represents a unique CDR3 protein sequence. Two dots are linked if they differ by 1 amino acid in the CDR3 region. This is the threshold used when

performing the single linkage clustering. The area of a dot is proportional to the number of reads with identical CDR3 protein sequences. Single cell-cloned antibodies are labeled with text. Red text indicates antibodies having a high affinity toward one of the virus strains used in the flu vaccine. Black text indicates antibodies with a low affinity toward one of the virus strains used in the flu vaccine or background level of binding toward all three virus strains. Eight of 10 single cell-cloned antibodies were found in the 454 sequences, except G04 and A06. All reads from 454 sequencing were used for this plot.

division, which increases their overall mutations per sequence. To analyze mutation statistics, we performed single linkage clustering by comparing the peptide sequences of CDR3 regions, using 1 amino acid as a threshold. We compared the average mutations per read from visit 1 PBMCs across different age groups. This number consistently increased with age in the IgG fraction (Fig. 3B) while remaining at the background level and with no difference between age groups in the IgM fraction (Fig. 3C), which is consistent with the fact that most of the IgM-expressing B cells are in a naïve state. We also applied the antibody lineage clustering performed previously using junctional nucleotide sequences, thresholded at 80% identity (15).

We found that mutations in general are far higher in IgG than in IgM, both when these mutations are measured relative to the germline reference sequences (Fig. 3, D and F) as well as to the most abundant sequence in each lineage. These observations point to mutational excursions of abundant class-switched sequences, as well as to diversification within the most abundant IgG lineages, respectively. In addition, there is a far greater parity between mutation loads measured at visits 1 and 3 among IgM antibodies ($R^2 = 0.92$) than among IgG ($R^2 = 0.54$). That is, even accounting for the variability among individuals, the IgM

repertoire is more similar between before and after vaccine samples than the IgG repertoire. This demonstrates that IgG antibodies undergo a far greater change in composition between the two time points compared to IgM. Furthermore, the elderly had the highest number of amino acid mutations in both visit 1 and visit 3 IgG fraction (Fig. 3D) while remaining low and similar to the IgM fraction in both visits of other age groups (Fig. 3F). This trend is consistent with our mutation analysis using clustering performed on amino acid sequences. IgG sequences had a higher number of mutations at visit 3 than at visit 1 when these were tallied in reference to the most abundant sequence in each lineage (off-diagonal line toward visit 3), suggesting that somatically hypermutated sequences persisted within the bloodstream 28 days after vaccination. At the same time, the elderly were not necessarily the group having the greatest number of mutations relative to these most abundant sequences (Fig. 3E). Therefore, because they lack any indications of greater intraclonal mutation compared to other age groups, these data suggest that the higher numbers of somatic mutations observed earlier in elderly individuals arise from clonal expansions that draw upon a pool of B cells having more somatic mutations to begin with.

DISCUSSION

Although the antibody repertoire is encoded by gene segments that are common to each individual human being, the various processes

of immunoglobulin diversity generation create a repertoire where the number of distinct immunoglobulin sequences in an individual exceeds the number of distinct genes in their consensus genome. The antibody repertoire is constantly evolving; it records the pathogenic

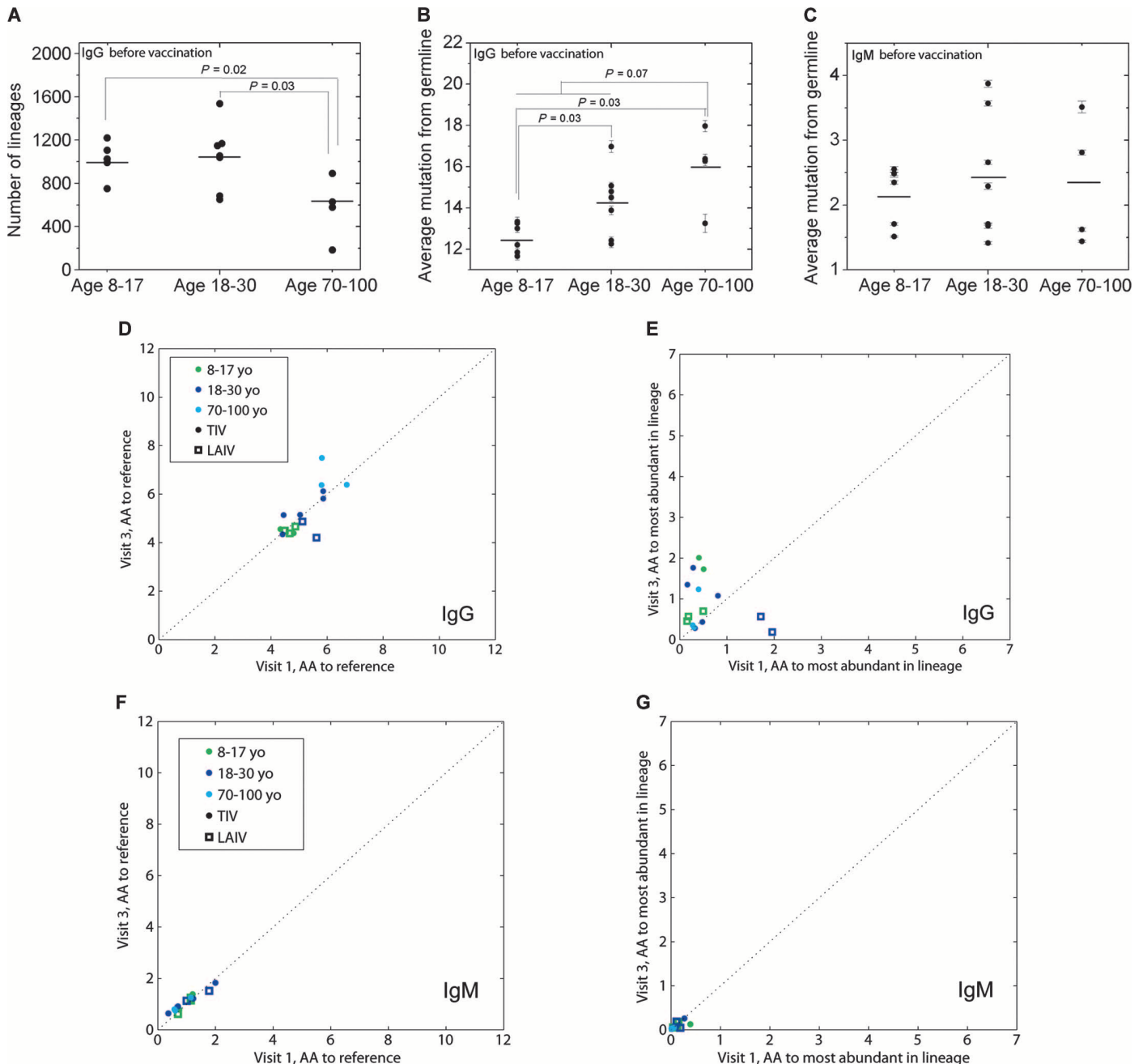


Fig. 3. Age-related repertoire diversity and mutation changes. **(A)** Repertoire diversity changes with age as measured by the number of lineages in IgG from visit 1 PBMCs. **(B and C)** Before vaccination mutation load as measured by averaging mutations at nucleotide level for IgG (B) and IgM (C) in visit 1 PBMCs, respectively. Mutations for each read were defined as the number of mismatches to germline reference in V, D, and J regions. **(D to G)** Lineage analysis, performed with 80% nucleotide sequence identity at the VDJ junctional region, gives measurements of amino acid mutations per read at V and J gene segments measured either to the germline

reference (D and F) or from the most abundant sequence of the lineage to which each belongs (E and G) for IgG (D and E) and IgM (F and G). *X* axes denote the measurement at visit 1, and the *y* axes denote the measurement at visit 3. Elderly patients show a higher number of IgG mutations from the germline (comparing 8- to 30-year-olds to 70- to 100-year-olds gives $P < 0.075$ before vaccination and $P < 0.0044$ after; restricting this analysis to TIV patients alone gives $P < 0.18$ and $P < 0.017$, respectively). Three thousand reads of subsampling were applied to all panels. All error bars are the SE. *P* values were calculated by Mann-Whitney *U* test.

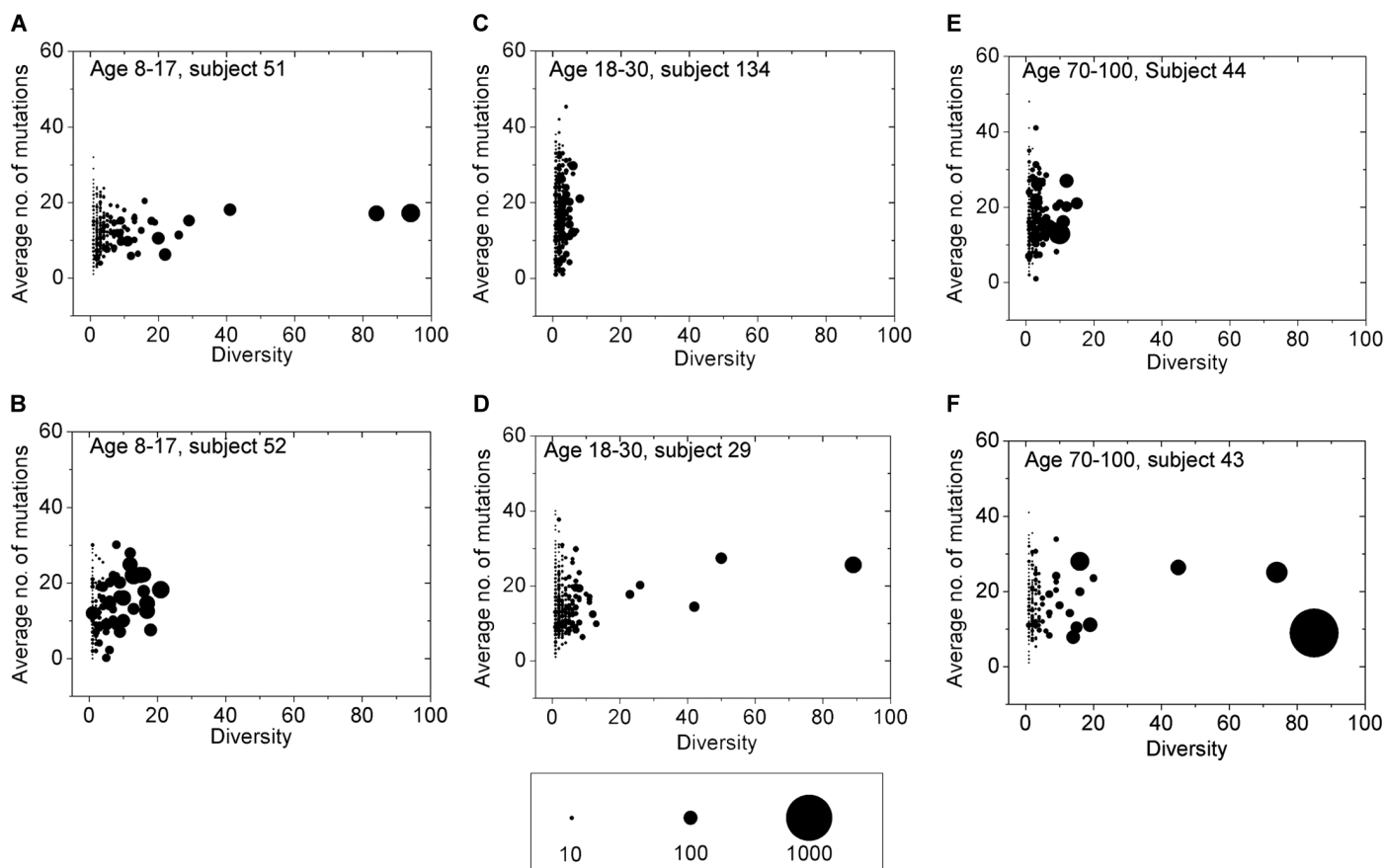


Fig. 4. Interlineage structure of IgGs in visit 2 PBMCs. (A to F) Interlineage structure of IgGs in visit 2 PBMCs is presented for six randomly selected subjects (A and B, age 8 to 17; C and D, age 18 to 30; E and F, age 70 to 100). Each dot represents a lineage of antibody sequences defined by single linkage clustering with 1 amino acid difference at CDR3 as the threshold. The area of the dot is proportional to the num-

ber of reads belonging to this lineage, as indicated in the scale bar. X axis is the diversity of the lineage that measures the number of unique protein sequences (full protein sequence, not just the CDR3 region) within the lineage. Y axis is the number of mutations at nucleotide level of the lineage averaged over reads. Three thousand reads of subsampling were applied.

exposure that one has experienced in the past and retains information on what it can protect us from. Therefore, it is of great interest to quantify and measure this dynamic system to understand how the repertoire responds to infection and vaccination and provide potential metrics for immune monitoring.

Here, we used seasonal influenza vaccine as a means of stimulation, and measured and quantified the changes in the antibody repertoire. First, we observed that the relative percentage of IgM sequences dropped after vaccination across all volunteers except for one. This reduction in IgM usage decreased with age, which is consistent with the hypothesis that the elderly are more likely to use memory B cells than NB cells to respond to influenza vaccination (26). We noted that children appear likely to increase relative IgA percentage in PBMCs compared to IgG.

A challenge of analyzing and quantifying the antibody repertoire is clonal expansion after antigen stimulation is not truly clonal as random mutations are introduced to the antibody genes at a rate of about 10^{-3} mutations per base pair per cell division (25). Using high-throughput sequencing in combination with informatics analysis, we were able to distinguish mutations, group sequences that differ by somatic hypermutation to the same clonal lineage, and follow the sequence evolution

within a lineage. This approach enabled us an unbiased measurement of the relative size among different lineages within one individual and the sequence diversities within each lineage.

A network representation of lineages allowed visual comparison of the intricate intra- and interlineage structure. Many of the top lineages were composed of extensively connected CDR3 sequences, each with varying number of sequencing reads. Sequence data from our single-cell cloning also confirmed that many of the top lineages are influenza-specific (Fig. 2 and fig. S11). Some single cell-generated sequences did not have a high affinity toward any one of the virus strains used in the vaccine; it is possible that they may not be representative or the recombinant antibodies may have been specific to internal viral proteins rather than the whole virus used in the enzyme-linked immunosorbent assay tests. The detailed topology of each lineage may contain information about how antigen selection and antibody affinity maturation work in concert in shaping the antibody repertoire. Studying the function of those informatically defined lineages may provide insight into this process.

Having several twin pairs among our subjects provides an interesting genetic control for the data. As one might expect, for the IgM

repertoire, the twins have closely related mutational loads, but these values diverge substantially for the IgG repertoire (fig. S19). We attribute this to the notion that the naïve repertoire is probably more strongly influenced by the background genetics of the individual, whereas the secondary repertoire incorporates a larger degree of stochasticity and randomness (15, 16). The mutational load versus diversity graphs for twins show little correlation (fig. S8, age 8 to 17 group); this is also to be expected as these data represent strong environmental and stochastic contributions to the immune system.

In conclusion, we have shown that it is possible to make personalized individual-specific measurements of immune repertoire with high-throughput DNA sequencing technology. These global repertoires contain a wealth of information and can be used to study individual-specific vaccine responses, and we have shown that analysis of the clonal structure provides direct insight into the effects of vaccination and provides a detailed molecular portrait of age-related effects. This approach to immune system characterization may be generally applicable to the development of new vaccines and may also help identify which individuals respond to a given vaccine.

MATERIALS AND METHODS

Human participants, vaccination protocol, blood sample collection, and cell sorting

Human participants, vaccination protocol, blood sample collection, and cell sorting were described by Sasaki *et al.* (9). Samples from a subgroup of volunteers were used in this study, and the demographic information of human participants was listed in table S1. The study protocols were approved by the institutional review boards at Stanford University. Informed consent was obtained from participants and the parents of pediatric participants. In addition, assent was obtained from the child participants. Participants were immunized with one dose of either the 2009 or the 2010 seasonal TIV (Fluzone, Sanofi Pasteur) or LAIV (FluMist, MedImmune). The 2009 vaccine contained an A/Brisbane/59/2007 (H1N1)-like virus, an A/Brisbane/10/2007 (H3N2)-like virus, and a B/Brisbane/60/2008-like virus. The 2010 vaccine contained an A/California/7/2009 (H1N1)-like virus, an A/Perth/16/2009 (H3N2)-like virus, and a B/Brisbane/60/2008-like virus. Blood samples were collected from each participant at three time points: day 0 before vaccination, day 7 or 8, and day 28 (± 4) after vaccination. PBMCs were isolated from the day 0 and day 28 blood samples with Ficoll-Paque Plus (GE Healthcare) following the manufacturer's instruction. Sorting of PBs was performed as previously described (9). In brief, B cells were isolated by negative selection with the RosetteSep Human B Cell Enrichment Cocktail (Stemcell Technologies) following the manufacturer's instructions from the day 7 to 8 whole blood samples. PBs were then sorted on the basis of the phenotype of CD3⁺CD19⁺CD20⁺CD27⁺CD38⁺, and NB cells were sorted on the basis of the phenotype of CD3⁺CD19⁺CD20⁺CD27⁺CD38⁺. Both populations reached a purity of 95%. Cells were lysed in RLT buffer (Qiagen) supplemented with 1% β -mercaptoethanol (Sigma) and stored at -80°C .

Primer design, RNA preparation, complementary DNA synthesis, and PCR

Two hundred forty-four human heavy chain variable gene segment sequences were downloaded from ImMunoGeneTics (27), excluding

pseudogenes. The leader regions of these sequences were used to design the 11 forward primers. The first 100 base pairs (bp) of the IgA, IgD, IgE, IgG, and IgM constant domain were used to design the reverse primers. Gene-specific primers were also designed for the reverse transcription step; these were located about 50 bp downstream from the PCR reverse primers. All primer sequences are listed in table S2.

Ten million PBMCs or sorted cells with varying numbers lysed in RLT buffer were used as input material for RNA purification. This was done with the AllPrep DNA/RNA Purification kit (Qiagen) following the manufacturer's instruction. The concentration of the RNA was determined with a NanoDrop spectrophotometer.

Complementary DNA (cDNA) was synthesized with SuperScript III reverse transcriptase (Invitrogen). One-fifth of the RNA purified from each sample was used for cDNA synthesis reactions with a total volume of 60 μL . All five constant region reverse transcription primers were added to the same reaction together with SUPERase-In (Ambion). RNase H (Invitrogen) was added to each reaction to remove RNA at the end of the cDNA synthesis step. All enzyme concentrations, reaction volumes, and the incubation temperature were based on the manufacturer's protocol for synthesis of cDNA with gene-specific primers.

For each sample, 11 PCRs were set up corresponding to 11 forward primers with a mixture of 5 reverse primers in each reaction. Two microliters of reverse transcription mixture was used in each PCR of 50 μL . Final concentration of 200 nM was used for each primer. The PCR program began with an initial denaturation at 94°C for 2 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing of primer to DNA at 60°C for 30 s, and extension by Platinum Taq DNA Polymerase High Fidelity (Invitrogen) at 68°C for 2 min. PCR products were first cleaned with QIAquick PCR Purification Kit (Qiagen) and then purified with the QIAquick Gel Extraction Kit (Qiagen). Concentration was measured with the NanoDrop spectrophotometer.

454 library preparation and sequencing

About 0.5 μg of QIAquick-cleaned PCR product for each sample was used to start the 454 library preparation process. 454 Titanium shotgun library construction protocol was followed for all samples. Briefly, double-stranded DNA was end-polished and ligated to sequencing adaptors, which contained a molecular identifier (MID, a nucleotide-based barcode system). The rest of the Roche 454 protocol was followed, which includes library immobilization, fill-in reaction, and single-stranded template DNA (sstDNA) library isolation. The sstDNA was quantified by a digital PCR method (28). Up to 16 libraries were pooled for one sequencing run, and Roche 454 emulsion PCR and sequencing protocols were followed for the rest of the sequencing procedure.

Data analysis

Detailed information on data analysis is included in the Supplementary Materials. In summary, reads from Roche 454 first entered into the primary analysis, which includes matching MID, filtering for minimum length of 250 bp, and then truncating to 220 bp. After V, D, and J assignment, protein translation rescue was formed on each reads. Sequencing reads that could not be rescued were discarded, which was about 13% of the sequencing reads. Single linkage cluster was performed at both nucleic acid and amino acid level with varying thresholds. Subsampling was applied in all analysis related to sequence and lineage diversity, mutation, and lineage structure except when looking for overlapping lineages between single cell-cloned sequences and high-throughput sequencing data.

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/5/171/171ra19/DC1

Materials and Methods

Fig. S1. Flow chart of bioinformatics pipeline.

Fig. S2. The composition of five antibody isotypes from PBMCs for each subject at visits 1 and 3.

Fig. S3. Isotype changes from visit 1 to visit 3.

Fig. S4. Young subjects have more lineages that are isotype-switched.

Fig. S5. Nucleotide sequence alignment for VDJ region exemplified for one isotype-switched lineage.

Fig. S6. Reads distribution based on relative sequence distance.

Fig. S7. The inter- and intralinear structure of all IgG lineages revealed by sequencing PBs sorted from the visit 2 blood samples for selected subjects.

Fig. S8. The interlineage structure of IgG from PBs sorted for all subjects at visit 2.

Fig. S9. The interlineage structure of IgG from PBMCs purified from subject 017-060 at visit 1.

Fig. S10. The interlineage structure of IgM and IgG for NB cells and PBs from one subject at visit 2.

Fig. S11. Overlapping of single cell-cloned antibody sequences with lineages.

Fig. S12. Repertoire diversity changes with age.

Fig. S13. Distribution of lineage size observes the power-law distribution.

Fig. S14. Mutation pattern of IgG for three age groups in visit 1 PBMCs.

Fig. S15. The mutation patterns for different age groups at threshold of 90% of nucleotide similarity in the CDR3 region.

Fig. S16. Zebrafish control data.

Fig. S17. Diversity and reads of IgG lineages of human PBs at visit 2.

Fig. S18. Synthetic sequence control data.

Fig. S19. Figure 3, B and C, respectively, from the main text with twin status indicated by arrows.

Table S1. Demographic information of human participants.

Table S2. Primer sequences.

Table S3. Summary of cell numbers and filtered reads for all samples.

Table S4. Raw reads for five isotypes in each sample.

Table S5. Summary of identifiable VJ reads for IgG in visit 2 PBs.

Table S6. Summary of single cell-cloned sequences.

Table S7. Control data information.

REFERENCES AND NOTES

1. N. Maizels, Immunoglobulin gene diversification. *Annu. Rev. Genet.* **39**, 23–46 (2005).
2. C. D. Allen, T. Okada, J. G. Cyster, Germinal-center organization and cellular dynamics. *Immunity* **27**, 190–202 (2007).
3. J. Di Noia, M. S. Neuberger, Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**, 43–48 (2002).
4. R. Ahmed, M. B. Oldstone, P. Palese, Protective immunity and susceptibility to infectious diseases: Lessons from the 1918 influenza pandemic. *Nat. Immunol.* **8**, 1188–1193 (2007).
5. J. Wrammert, D. Koutsonanos, G. M. Li, S. Edupuganti, J. Sui, M. Morrissey, M. McCausland, I. Skountzou, M. Hornig, W. I. Lipkin, A. Mehta, B. Razavi, C. Del Rio, N. Y. Zheng, J. H. Lee, M. Huang, Z. Ali, K. Kaur, S. Andrews, R. R. Amara, Y. Wang, S. R. Das, C. D. O'Donnell, J. W. Yewdell, K. Subbarao, W. A. Marasco, M. J. Mulligan, R. Compans, R. Ahmed, P. C. Wilson, Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection. *J. Exp. Med.* **208**, 181–193 (2011).
6. J. Wrammert, K. Smith, J. Miller, W. A. Langley, K. Kokko, C. Larsen, N. Y. Zheng, I. Mays, L. Garman, C. Helms, J. James, G. M. Air, J. D. Capra, R. Ahmed, P. C. Wilson, Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* **453**, 667–671 (2008).
7. X. Yu, T. Tsibane, P. A. McGraw, F. S. House, C. J. Keefer, M. D. Hicar, T. M. Tumpey, C. Pappas, L. A. Perrone, O. Martinez, J. Stevens, I. A. Wilson, P. V. Aguilar, E. L. Altschuler, C. F. Basler, J. E. Crowe Jr., Neutralizing antibodies derived from the B cells of 1918 influenza pandemic survivors. *Nature* **455**, 532–536 (2008).
8. H. I. Nakaya, J. Wrammert, E. K. Lee, L. Racioppi, S. Marie-Kunze, W. N. Haining, A. R. Means, S. P. Kasturi, N. Khan, G. M. Li, M. McCausland, V. Kanchan, K. E. Kokko, S. Li, R. Elbein, A. K. Mehta, A. Aderem, K. Subbarao, R. Ahmed, B. Pulendran, Systems biology of vaccination for seasonal influenza in humans. *Nat. Immunol.* **12**, 786–795 (2011).
9. S. Sasaki, M. Sullivan, C. F. Narvaez, T. H. Holmes, D. Furman, N. Y. Zheng, M. Nishtala, J. Wrammert, K. Smith, J. A. James, C. L. Dekker, M. M. Davis, P. C. Wilson, H. B. Greenberg, X. S. He, Limited efficacy of inactivated influenza vaccine in elderly individuals is associated with decreased production of vaccine-specific antibodies. *J. Clin. Invest.* **121**, 3109–3119 (2011).
10. J. C. Krause, T. Tsibane, T. M. Tumpey, C. J. Huffman, B. S. Briney, S. A. Smith, C. F. Basler, J. E. Crowe Jr., Epitope-specific human influenza antibody repertoires diversify by B cell intraclonal sequence divergence and interclonal convergence. *J. Immunol.* **187**, 3704–3711 (2011).
11. J. A. Weinstein, N. Jiang, R. A. White III, D. S. Fisher, S. R. Quake, High-throughput sequencing of the zebrafish antibody repertoire. *Science* **324**, 807–810 (2009).
12. S. D. Boyd, E. L. Marshall, J. D. Merker, J. M. Maniar, L. N. Zhang, B. Sahaf, C. D. Jones, B. B. Simen, B. Hanczaruk, K. D. Nguyen, K. C. Nadeau, M. Egholm, D. B. Miklos, J. L. Zehnder, A. Z. Fire, Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci. Transl. Med.* **1**, 12ra23 (2009).
13. S. T. Reddy, X. Ge, A. E. Miklos, R. A. Hughes, S. H. Kang, K. H. Hoi, C. Chrysostomou, S. P. Hunnicke-Smith, B. L. Iverson, P. W. Tucker, A. D. Ellington, G. Georgiou, Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat. Biotechnol.* **28**, 965–969 (2010).
14. J. Glanville, W. Zhai, J. Berka, D. Telman, G. Huerta, G. R. Mehta, I. Ni, L. Mei, P. D. Sundar, G. M. Day, D. Cox, A. Rajpal, J. Pons, Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20216–20221 (2009).
15. N. Jiang, J. A. Weinstein, L. Penland, R. A. White III, D. S. Fisher, S. R. Quake, Determinism and stochasticity during maturation of the zebrafish antibody repertoire. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5348–5353 (2011).
16. J. Glanville, T. C. Kuo, H. C. von Büdingen, L. Guey, J. Berka, P. D. Sundar, G. Huerta, G. R. Mehta, J. R. Oksenberg, S. L. Hauser, D. R. Cox, A. Rajpal, J. Pons, Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20066–20071 (2011).
17. R. Arnaout, W. Lee, P. Cahill, T. Honan, T. Sparrow, M. Weiland, C. Nusbaum, K. Rajewsky, S. B. Koralov, High-resolution description of antibody heavy-chain repertoires in humans. *PLoS One* **6**, e22365 (2011).
18. Y. C. Wu, D. Kipling, D. K. Dunn-Walters, Age-related changes in human peripheral blood IGH repertoire following vaccination. *Front. Immunol.* **3**, 193 (2012).
19. R. J. Cox, K. A. Brokstad, P. Ogra, Influenza virus: Immunity and vaccination strategies. Comparison of the immune response to inactivated and live, attenuated influenza vaccines. *Scand. J. Immunol.* **59**, 1–15 (2004).
20. U. Schibler, K. B. Marcu, R. P. Perry, The synthesis and processing of the messenger RNAs specifying heavy and light chain immunoglobulins in MPC-11 cells. *Cell* **15**, 1495–1509 (1978).
21. H. M. Jäck, M. Wabl, Immunoglobulin mRNA stability varies during B lymphocyte differentiation. *EMBO J.* **7**, 1041–1046 (1988).
22. C. T. Yan, C. Boboila, E. K. Souza, S. Franco, T. R. Hickernell, M. Murphy, S. Gumaste, M. Geyer, A. A. Zarrin, J. P. Manis, K. Rajewsky, F. W. Alt, IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* **449**, 478–482 (2007).
23. L. Warren, D. Bryder, I. L. Weissman, S. R. Quake, Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17807–17812 (2006).
24. H. C. Fan, S. R. Quake, Detection of aneuploidy with digital polymerase chain reaction. *Anal. Chem.* **79**, 7576–7579 (2007).
25. V. H. Odegard, D. G. Schatz, Targeting of somatic hypermutation. *Nat. Rev. Immunol.* **6**, 573–583 (2006).
26. P. R. Dormitzer, G. Galli, F. Castellino, H. Golding, S. Khurana, G. Del Giudice, R. Rappuoli, Influenza vaccine immunology. *Immunol. Rev.* **239**, 167–177 (2011).
27. V. Giudicelli, D. Chaume, M. P. Lefranc, IMGT/GENE-DB: A comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res.* **33**, D256–D261 (2005).
28. R. A. White III, P. C. Blainey, H. C. Fan, S. R. Quake, Digital PCR provides sensitive and absolute calibration for high throughput sequencing. *BMC Genomics* **10**, 116 (2009).
29. E. A. Kabat, T. T. Wu, M. Reid-Miller, H. M. Perry, K. S. Gottesman, C. Foeller, *Sequences of Proteins of Immunological Interest* (Public Health Service, National Institutes of Health, Washington, DC, ed. 5, 1991), vol. 1.
30. S. H. Kleinstein, Y. Louzoun, M. J. Shlomchik, Estimating hypermutation rates from clonal tree data. *J. Immunol.* **171**, 4639–4649 (2003).

Acknowledgments: We thank N. Neff, B. Passarelli, J. Okamoto, N. Gobet, and G. Mantalas at the Stanford Stem Cell Genome Center for assistance with sequencing. We also thank S. Mackey for clinical project management and regulatory and data management; S. Swope and C. Walsh for research nurse support; G. Swam and members at the SRI International for help with twin volunteer recruitment; and H. Maecker, J. Bierre, and B. Varasteh at the Stanford Human Immune Monitoring Core for sample banking. **Funding:** This research was supported by NIH grant

U19 AI057229 (M.M.D., X.-S.H., H.B.G., and S.R.Q.), an NIH Pathway to Independence Award K99 AG040149 (N.J.), and an NSF graduate fellowship (J.A.W.). **Author contributions:** N.J., J.A.W., M.M.D., and S.R.Q. conceived the initial idea; N.J., J.A.W., X.-S.H., C.L.D., H.B.G., M.M.D., and S.R.Q. designed the research; C.L.D. was responsible for regulatory and clinical aspects of the study protocols; S.S. and X.-S.H. performed cell sorting; N.-Y.Z., M.H., M.S., and P.C.W. generated recombinant monoclonal antibodies and performed antibody functional assay; N.J., J.H., J.A.W., and L.P. performed high-throughput sequencing-related research; N.J., J.H., J.A.W., D.S.F., and S.R.Q. analyzed the data; and N.J., J.H., J.A.W., D.S.F., and S.R.Q. wrote the paper. **Competing interests:** A patent application entitled "Measurement and comparison of immune diversity by high-throughput sequencing" (application number: PCT/US2011/035507; inventors: S.R.Q., J.A.W., N.J., and D.S.F.) was filed by Stanford University based on this work. S.R.Q., M.M.D., D.S.F., and N.J. are advisors of ImmuMetrix, LLC, of which S.R.Q. is also a founder. The other authors declare that they have no competing interests. **Data and materials availability:** The

sequence data sets published in this paper can be found in the Sequence Read Archive, accession no. SRA058972.

Submitted 16 August 2012

Accepted 26 November 2012

Published 6 February 2013

10.1126/scitranslmed.3004794

Citation: N. Jiang, J. He, J. A. Weinstein, L. Penland, S. Sasaki, X.-S. He, C. L. Dekker, N.-Y. Zheng, M. Huang, M. Sullivan, P. C. Wilson, H. B. Greenberg, M. M. Davis, D. S. Fisher, S. R. Quake, Lineage structure of the human antibody repertoire in response to influenza vaccination. *Sci. Transl. Med.* **5**, 171ra19 (2013).