Effects of Aging, Cytomegalovirus Infection, and EBV Infection on Human B Cell Repertoires

Chen Wang, Yi Liu, Lan T. Xu, Katherine J. L. Jackson, Krisha M. Roskin, Tho D. Pham, Jonathan Laserson, Eleanor L. Marshall, Katie Seo, Ji-Yeun Lee, David Furman, Daphne Koller, Cornelia L. Dekker, Mark M. Davis, Andrew Z. Fire and Scott D. Boyd

*J Immunol* published online 11 December 2013
http://www.jimmunol.org/content/early/2013/12/11/jimmunol.1301384

Supplementary Material
http://www.jimmunol.org/content/suppl/2013/12/11/jimmunol.1301384.DC1

Why The JI?
- **Rapid Reviews! 30 days*** from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Speedy Publication!** 4 weeks from acceptance to publication

*average

Subscription
Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

Permissions
Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts
Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts

*The Journal of Immunology* is published twice each month by The American Association of Immunologists, Inc., 1451 Rockville Pike, Suite 650, Rockville, MD 20852
Copyright © 2013 by The American Association of Immunologists, Inc. All rights reserved.
Print ISSN: 0022-1767 Online ISSN: 1550-6606.
Effects of Aging, Cytomegalovirus Infection, and EBV Infection on Human B Cell Repertoires

Chen Wang,* 1 Yi Liu,* 1,7,1 Lan T. Xu,* Katherine J. L. Jackson,* Krishna M. Roskin,* Tho D. Pham,* Jonathan Laserson, ‡ Eleanor L. Marshall, § Katie Seo,*, Ji-Yeun Lee,*, David Furman,† Daphne Koller, ‡ Cornelia L. Dekker, ‡ Mark M. Davis,‡¹,** Andrew Z. Fire,‡§ and Scott D. Boyd*

Elderly humans show decreased humoral immunity to pathogens and vaccines, yet the effects of aging on B cells are not fully known. Chronic viral infection by CMV is implicated as a driver of clonal T cell proliferations in some aging humans, but whether CMV or EBV infection contributes to alterations in the B cell repertoire with age is unclear. We have used high-throughput DNA sequencing of IGH gene rearrangements to study the BCR repertoires over two successive years in 27 individuals ranging in age from 20 to 89 y. Some features of the B cell repertoire remain stable with age, but elderly subjects show increased numbers of B cells with long CDR3 regions, a trend toward accumulation of more highly mutated IgM and IgG Ig genes, and persistent clonal B cell populations in the blood. Seropositivity for CMV or EBV infection alters B cell repertoires, regardless of the individual's age: EBV infection correlates with the presence of persistent clonal B cell expansions, whereas CMV infection correlates with the proportion of highly mutated Ab genes. These findings isolate effects of aging from those of chronic viral infection on B cell repertoires and provide a baseline for understanding human B cell responses to vaccination or infectious stimuli. The Journal of Immunology, 2014, 192: 000–000.

Any elderly individuals have a compromised immune system, leading to increased susceptibility to infectious diseases and decreased responses to vaccination (1). Aging has been reported to impair innate immunity, T cells, and Ab-producing B cells (1–5). Humoral responses are critical for responding to pathogens such as Streptococcus pneumoniae and influenza viruses that cause increased morbidity and mortality in the elderly, but age-related changes in human B cells and Ig repertoires are only beginning to be understood (6–8).

Advanced age has been reported to lead to increased or decreased B cell counts in the peripheral blood, increased, decreased or unchanged proportions of naive B cells, and increased CDS* B cell populations (3, 5, 9–13). Changes in serum Ab production, including decreases in vaccine-specific Abs, and decreased isotype switching associated with lower expression of activation-induced cytidine deaminase in B cells also have been described previously (8, 10, 14). Understanding the effects of aging on B cell function is further complicated by the common chronic viral infections seen at higher rates in the aging population, such as CMV and EBV. CMV infection is correlated with increased counts of LFA-1*CD8* memory T cells and reduced naive CD8* T cells, whereas total B cell counts in the blood are reportedly increased in CMV-seropositive individuals (15–17).

Following V(D)J rearrangement to generate functional Ig genes in B cells, the Ig repertoire during a human's life span is further shaped by negative selection against self-Ags, clonal expansion of B cells stimulated by Ag, activation-induced mutation of Ig genes, and receptor editing, among other processes. Ineffective Ab responses in the elderly have been attributed to decreased diversity of Ab repertoires with accumulation of memory B cells and decrease of naive B cell populations (18). Influenza vaccination responses in the elderly are associated with decreased numbers of vaccine-stimulated B cells (8), and a recent study that included four elderly subjects show decreased diversity of influenza vaccine-stimulated B cells (19). However, there is also evidence of relatively preserved Ig repertoire diversity in tonsillar tissue of aged humans and increased proportions of naive B cells in some elderly individuals (20). Mutation of IGHV in B cell populations reportedly changes with aging, with one study reporting modestly increased mutation in IgG but not memory IgM B cell populations in the blood, whereas data from tonsillar B cells indicate increased mutation in memory IgM B cells but not other subsets (20, 21). Most prior studies of IGHV gene rearrangements in young versus elderly subjects have been limited to examination of tens to hundreds of sequences, from small numbers of individuals, or have not assessed the potentially confounding effects of chronic herpesvirus infections (20–23). Seropositivity for CMV, in particular, increases with age in human populations and should be controlled for in studies of the effects of aging on the immune system (24).
In this study, we characterize peripheral bloodIGH repertoires measured with >500,000 sequences from a cohort of healthy young (n = 10) and older (n = 17) people over two consecutive years and analyze features that change with age, CMV, or EBV infection. Some B cell repertoire features are stable with age, but we find that elderly individuals show increased numbers of B cells expressing long IGH CDR3 regions and that the proportion of highly mutated B cells, particularly in IgM and IgG populations, shows a trend toward increasing with age and is increased in subjects infected with CMV. Unusual large persistent clonal populations of B cells are common in the oldest individuals in our data set and are absent from younger individuals; notably, the contribution of both large and small persistent B cell clones over the yearlong time course is correlated with EBV seropositivity, regardless of age. Taken together, these findings isolate age-specific and CMV- or EBV-associated alterations in the B cell repertoire and provide a baseline for further study of impaired Ag-specific responses and increased autoreactivity in the elderly.

Materials and Methods
Specimen collection

Human peripheral blood was collected from a cohort of 27 healthy participants aged 20–89 y during each of two consecutive years: Y2 (2008) and Y3 (2009). Participants’ ID, age, and other demographic information is described in Supplemental Table I. The participants were grouped into three age categories: 20–31 y (n = 10), 61–69 y (n = 7), and 72–89 y (n = 10). Recruitment of patients, documentation of informed consent, collection of blood specimens, and experimental measurements were carried out with Institutional Review Board approval at Stanford University.

CMV and EBV serotyping

Serum was separated by centrifugation of clotted blood and stored at −80°C before use. After all samples were collected, specimens were thawed at room temperature and assayed for the presence of CMV and EBV Abs (both IgG and IgM) using the CMV or EBV ELISA Kit from Calbiotech (Spring Valley, CA), as recommended by the manufacturer. No specimens were positive for virus-specific IgM.

Genomic DNA and cDNA template preparation and PCR amplification

Isolation of PBMCs was performed as described previously, and isolation of genomic DNA (gDNA) and RNA was performed using Allprep kits (Qiagen) (22). PCR of IGH V region rearrangements from each sample was carried out with six independent 100 ng gDNA aliquots to generate six independent bar-coded libraries per sample. Multiplexed primer sets hybridizing to framework regions (FR1 and FR2), and a J-primer were based on the BIOMED-2 design (25). For isotype-specific Ig libraries, mRNA was reverse-transcribed to cDNA using random hexamer primers, and cDNA corresponding to 100 ng RNA was used for PCR using the same 5’- FR1 and FR2 multiplex primers and 3’-isotype-specific primers located in the first exon of the C region. Sample identity and replicate library identity were encoded by 10-nucleotide “bar code” sequences in the primers. PCR was carried out with AmpliTaq Gold (Roche) following the manufacturer’s instructions and used a program of 94°C 5 min, 35 cycles of (94°C 30 s, 60°C 45 s, and 72°C 90 s), and final extension at 72°C for 10 min. PCR products from each replicate library were quantitated, pooled in equimolar amounts, and then gel-purified and gel extracted (Qiagen). High-throughput sequencing was performed on the 454 (Roche) platform using Titanium chemistry.

Sequence quality assessment and filtering

Sequences with matching bar codes were assigned to the appropriate replicates and samples and then were trimmed of bar codes and IGHV primer sequences. The V, D, and J regions and V-D (N1) and D-J (N2) junctions were identified by alignment with germline IGH sequences from ImMunoGeneTics and National Center for Biotechnology Information databases, using the alignment program iHMMune-align (26, 27). Amino acid sequences and location of CDR3 were defined by the conserved cysteine-104 and tryptophan-118 based on the ImMunoGeneTics nomenclature. Sequences were filtered to remove non-cc and c-sequences with V-gene insertion or deletions, and chimeric sequences. Samples with <100 sequences were excluded from further analysis (two samples from Y3). A total of 323,285 gDNA and 189,915 cDNA sequences were analyzed. The sequences can be accessed via the National Center for Biotechnology Information dbGAP online archive (http://www.ncbi.nlm.nih.gov/gap) with accession number phs000666.v1.p1.

V, D, and J usage, functional features, and V-mutation analysis

Sequences from all replicates of each sample were pooled, and unique clones were defined by the same V and J gene usage and identical CDR3 as sequence. To analyze V, D, and J usage frequencies, CDR3 features, and mutations in V, reads belonging to each unique clone were collapsed to a single read. Hydrophobicity of CDR3 peptides was calculated using the Kyte–Dolittle scale (29) and net charge was calculated using the Henderson–Hasselbalch equation (30). CDR3 segments from out-of-frame sequences were excluded from CDR3 feature analysis. Distribution of mutation along the V-region was obtained by alignment with gapped ImMunoGeneTics germ-line sequences and counting the mutation frequency per nucleotide position.

Clonal expansion analysis

To summarize the contribution of clonally expanded B cells to the repertoire while normalizing for sequencing depth obtained for each sample, we calculated a “clonality score,” adapted from a coincidence index used in cryptanalysis, sum(f[i]), (where f[i] is the frequency of component i) (31). To compensate for the effect of PCR amplification on sequence counts, we calculated the clonality score as follows: sum of Nc × Na (j=k) over i,j,k, divided by the sum of Tc × Ta (j=k) over j,k, where Nc and Na are the copy numbers of clone i observed in independent replicate PCR libraries j and k generated from independent aliquots of template DNA, and Tc and Ta are the corresponding replicate libraries. To quantify the contribution of persistent B cell clonal populations detected in the two time points 1 y apart in the study, we calculated a “persistent clonality score” exactly as above, except that only replicate libraries from different years were compared with each other. Large clones were defined as those observed in more than four of the six replicate libraries from a sample and comprising >1% of total reads from the sample.

Correlation and regression analysis for clonality

We tested for correlation between clonality score, persistent clonality score, IGHV segment mutation frequency, age, gender, and seropositivity for EBV and CMV using multiple linear regression, including all subjects in the analysis. Laplace smoothing was applied to avoid occurrence of zeroes prior to calculating the logarithm of the clonality score.

Phylogenetic visualization of large clones

ImmuniTree is a dedicated algorithm and implementation for building immune receptor sequence trees from high-throughput DNA sequencing data (32). The following briefly describes the conceptual basis for calculating in ImmuniTree, although additional details of the model-building and computation will be presented elsewhere. A variety of broadly applied phylogenetic algorithms deal with sequences derived from species that are presumed to have evolved from common ancestors but where the intermediate shared ancestors are no longer present to be observed and must therefore be inferred. In contrast, ImmuniTree takes into account the possibility that common ancestor B cells may still be present in the members of a clonal lineage that are observed.

In the ImmuniTree model of somatic hypermutation (SHM) of Ig gene rearrangements, potential nucleotides at each position on sequences representing parent and child subclones are modeled using a conditional multinomial distribution. Sequencing error is modeled similarly: the mutations from subclones to reads are also modeled using a conditional multinomial distribution. These distributions are subject to a given subclone, the SHM prior penalizes indels severely, whereas the sequencing error prior treats indels more permissively. This reflects the fact that true SHM is under selective pressure to produce in-frame Abs, whereas 454 sequencing is known to generate homopolymer errors. In the ImmuniTree model, internal subclones may or may not generate reads; also, internal subclones may or may not have multiple subclones as children. In contrast, existing phylogenetic algorithms generally associate reads only to the leaves of a binary tree. ImmuniTree thus builds on and extends existing phylogenetic algorithms, implementing the aforementioned fine-tuning designed to account for the complexity of Ig repertoires and sequencing, and uses a single sampling and optimization approach to compute a lineage that optimizes the overall likelihood. Any specific lineage has a likelihood consisting of the following parameters: SHM rates from parent subclones (nodes) to children subclones, sequencing error of reads corresponding to a given subclone, the best V and J reference segments, the parent subclone of each subclone, and subclone birth and death rates. The parameters are sampled using a Metropolis–Hastings Markov Chain Monte Carlo approach. After running for...
5000 iterations, each sampled lineage is examined and subjected to local optimization. The highest-likelihood lineage among the optimized sampled lineages is returned.

**Results**

**V, D, and J usage varies among individuals but is comparable between young and older people**

We first examined whether there were discernable age-related changes in the usage of V, D, or J gene in IGH rearrangements. Fig. 1 shows that the proportion of different genes was comparable across three age groups of samples in 2008 (Y2), with gene frequencies consistent with literature reports (33). V, D, and J gene usage from 2009 (Y3) follows a similar pattern and bears a high correlation with that from Y2 (data not shown). V gene segment usage in humans shows a different distribution in mutated versus nonmutated Ab gene rearrangements, so we separated these sequence subsets in our analysis (33). To be conservative in our mutation frequency estimates, given that 454 sequencing can demonstrate error rates of ∼0.3–1% per base (22), we consider sequences with <1% V-mutation frequency as unmutated and those with >1% as mutated. The V, D, and J gene usage did not show a significant difference between younger and older groups within either mutated and unmutated categories. Variation between individuals in gene usage patterns was seen, similar to that previously reported (34), consistent with individual differences in the intrinsic mechanisms that generate VDJ rearrangements, including copy-number variation in IGH germline sequences, or recombination signal sequence features, or individual variation in the B cell selection process.

**CDR3 length in both mutated and unmutated Abs increases with age**

We further analyzed biochemical features of the IgH CDR3 aa sequences, given the prominent role of this sequence region in determining Ab binding specificity, and the effects of B cell selection on the length of IgH CDR3 regions present in the B cell repertoire. It has been reported that B cell development in the bone marrow is accompanied by preferential removal of B cells expressing Abs with long CDR3s, potentially because of the increased levels of autoreactivity demonstrated by such Abs (35). Evidence of additional selection against B cells expressing Abs with long CDR3s, or in favor of sequences with shorter CDR3s during Ag-driven peripheral responses, is seen in the decreased IgH CDR3s present in B cells expressing mutated Ig genes compared with unmutated ones (36). To look for age-associated alterations in these selection processes, we analyzed the CDR3 features of mutated and unmutated sequences separately in our data set. Confirming prior observations, the unmutated sequences in our data set Fig. 2, middle panel, have longer CDR3s than mutated sequences Fig. 2, lower panel. Interestingly, the CDR3 regions of both unmutated and mutated sequences are significantly longer in older people compared with young individuals, especially within mutated sequences (Fig. 2, middle and lower panel), whereas there is no obvious change in hydrophobicity or net charge with age. The presence of increased numbers of B cells with longer CDR3 segments in both unmutated and mutated Ab H chain rearrangements in elderly individuals suggests an overall decreased level of selection against Abs with long CDR3s in the aging immune system, both in the generation of naive B cells, as well as during Ag-driven responses.

**IgM and IgG show an age- and CMV- associated increase in mutation**

We initially evaluated the levels of IGHV mutation regardless of Ab isotype in participants of various ages. We grouped sequences into three groups: <1% mutation frequency (unmutated), 1–10% mutation frequency (moderate), and >10% mutation frequency (high). As...
IGHV and EBV correlate with the extent of upper panel individuals aged (20–31), (61–69), and (72–89), respectively (Fig. 3B, comprises, on average, 68, 68, and 75% of the repertoire in individual's total B cell repertoire. Fig. 3B shows that highly mutated Ig showed that CMV has a stronger effect than that seen with regard to gender or EBV infection status. Multiple regression analysis of factors contributing to the proportion of reads falling in these groups in each individual is highly correlated between two consecutive years, as shown in Fig. 3A, the proportion of reads falling in these groups in each individual is highly correlated between two consecutive years, showing both the variation between individuals, and the stability of this feature of the B cell repertoire over time. Unmutated IGHV comprises, on average, 68, 68, and 75% of the repertoire in individuals aged (20–31), (61–69), and (72–89), respectively (Fig. 3B, upper panel).

We then asked whether gender or seropositivity for CMV or EBV correlate with the extent of IGHV mutation in each individual’s total B cell repertoire. Fig. 3B shows that highly mutated IGHV is increased in CMV-positive versus CMV-negative participants (p < 0.05). No significant differences in mutation rates were seen with regard to gender or EBV infection status. Multiple regression analysis of factors contributing to the proportion of highly mutated Ig showed that CMV has a stronger effect than that of age, gender, or EBV infection, with p = 0.11 in the model.

To evaluate IGHV mutational frequencies in greater detail, we examined sequences from Abs of particular isotypes to look for changes associated with aging or chronic viral infection. Fig. 4A shows that highly mutated IgM and IgG sequences show a trend toward being increased in the oldest individuals (72–89 y) compared with younger individuals, although one individual in the age 21–30 group (p65, described further below) had a proportion of >10% mutated IgM sequences that was unusually high for their age; this individual proved to be EBV and CMV seropositive. CMV infection is associated with an increased proportion of highly mutated sequences of IgM and IgG isotypes (Fig. 4B). No significant age- or CMV-related increase in mutation frequencies was seen for other isotypes including IgA or IgD (Supplemental Fig. 1). We considered IgM or IgD sequences with 1% or greater mutation to be derived from IgM memory B cells, whereas those with <1% mutation were from naive cells (38). We saw little difference in IGHV gene family usage between mutated IgM, IgD, IgG, or IgA. However, we saw significant differences in IGHV gene family usage between mutated IgM, IgD, IgG, or IgA (Supplemental Fig. 2). Similarly, age, CMV, or EBV infection had little impact on IGHV gene usage (Supplemental Fig. 2).

Previous studies have reported evidence that memory IgM and isotype switched memory B cell populations use IGHV families with differing frequencies, with IgM memory cells having a lower frequency of IGHV1 gene usage and higher frequency of IGHV3 gene usage than IgG and IgA isotype-switched memory B cells (37). We considered IgM or IgD sequences with ≥1% or greater mutation frequency to be derived from IgM memory B cells, whereas those with <1% mutation were from naive cells (38). Supplemental Fig. 2 shows that naive IgM and IgD sequences have significantly higher IGHV1 and lower IGHV3 usage than mutated IgM, IgD, IgG, or IgA. However, we saw little difference in IGHV gene family usage between mutated IgM, IgD, IgG, or IgA (Supplemental Fig. 2). Similarly, age, CMV, or EBV infection had little impact on IGHV gene usage (Supplemental Fig. 2).

Age and chronic EBV infections are correlated with persistent B cell clonal expansion

We quantified clonally expanded B cells in each individual, using a conservative clone definition of sharing the same V and J genes (191x455 to 473x709).
and having identical CDR3 amino acid sequences. Six independent *IGH* libraries amplified from gDNA template were examined from each sample, to distinguish between PCR amplification bias and true clonal expansion of B cells, by requiring that the same rearrangement be detected in at least two independent libraries from the sample to be considered clonally expanded. We calculated a “clonality score,” as described in Materials and Methods, which can be thought of as the probability that two B cells selected at random from the patient’s peripheral blood will belong to the same clonal lineage. Fig. 5A and 5B show that most samples from healthy individuals sequenced at this depth show low levels of B cell clonality, indicating that almost all of the sequences observed in each replicate library from a given individual are unique. Significantly increased clonality scores are seen in some elderly individuals, and correlate with CMV infection regardless of age. (A) Frequencies of gDNA sequences with different V-mutation levels are correlated strongly between samples from two consecutive years of the same participants. V-mutation frequencies are categorized to three levels: <1% (unmutated), 1–10% (moderate), and >10% (high) mutation. (B) Frequencies of sequences with different mutation levels are grouped by age, gender, CMV, or EBV status. Sequences are pooled from two years for each participant. The *p* values are calculated by two-sided Student *t* test, including all data points for each category. *p* < 0.05.

**FIGURE 3.** V-mutation levels are increased in CMV-positive individuals. (A) Frequencies of gDNA sequences with different V-mutation levels are correlated strongly between samples from two consecutive years of the same participants. V-mutation frequencies are categorized to three levels: <1% (unmutated), 1–10% (moderate), and >10% (high) mutation. (B) Frequencies of sequences with different mutation levels are grouped by age, gender, CMV, or EBV status. Sequences are pooled from two years for each participant. The *p* values are calculated by two-sided Student *t* test, including all data points for each category. *p* < 0.05.

**FIGURE 4.** Frequencies of highly mutated IgM and IgG sequences are increased in some elderly individuals, and correlate with CMV infection regardless of age. (A) Frequencies of IgM (upper panel) and IgG (lower panel) cDNA sequences with different levels of V-mutation grouped by age of participants. (B) Frequencies of IgM (upper panel) and IgG (lower panel) cDNA sequences with different levels of V-mutation grouped by CMV status of participants. Sequences are pooled from two years. The *p* values are calculated by two-sided Student *t* test, including all data points for each category. *p* < 0.05.

We evaluated the correlation between B cell repertoire clonality and an individual’s age, gender, and serological status for EBV and CMV infection. To combine sequences from 2 y, we took the average of the clonality score from the 2 y. The heat map of Fig. 5C shows that age is correlated with B cell clonal expansions detected within each year of the study, with a correlation of 0.46 and two-sided *p* value of 0.016. We used multiple linear regression to address potential confounding variables, using gender, CMV status, and EBV status as covariates, correcting for multiple hypotheses. Age remained a significant factor, with a Benjamini–Hochberg corrected false-discovery rate of 0.061, and EBV infection status showed significance with a false-discovery rate of 0.083. Persistent clonality scores normalized for sequencing depth were calculated as described in Materials and Methods to quantifyate B cell clones persisting in an individual through the yearlong time course of the study. Correlation coefficients (Fig. 5D) showed that the persistent clonality score (i.e., the levels of B cell clones that were found to persist in a particular individual and were detected in both year Y2 and year Y3) was correlated with age and EBV status. Multiple linear regression analysis demonstrated that the effects of age and EBV infection on the persistent clonality score are significant after multiple hypothe-
represent abnormal neoplastic monoclonal B cell proliferations tent with prior or ongoing affinity maturation. All big clones in the IGHV changes (Supplemental Table II). The large clones showed mu- related CDR3 sequences differing only by apparent mutation.

An elderly individual (p23) showed three distinct larger persistent B cell clonal expansion. (A) and (B) Clonality scores of the Ig repertoire of each sample are grouped by age for time point Y2 (A, n = 27) and Y3 (B, n = 25). The pairwise comparison between different age groups is performed with two-sided Wilcoxon tests. **p < 0.01. (C) Correlation heat map of Y2-Y3 average clonality score (log) for each participant’s B cell repertoire, according to age, gender, CMV seropositivity, and EBV seropositivity in 27 individuals. (D) Correlation heat map of persistent clonality score (log) for clones detected in Y2 and Y3 of the study, according to age, gender, CMV, EBV, and frequency of highly mutated sequences (V-mut, freq.), in 25 individuals. The colored scale bar represents the strength of correlation for each variable pair. Details of the clonality score calculations, which yield a measure of the proportion of the B cell repertoire contributed by expanded clonal populations, normalized for the depth of sequencing carried out in each sample, are presented in Materials and Methods.

Discussion

We have analyzed >500,000 rearranged IgH sequences derived from the peripheral blood of young adults and elderly individuals in serial samples taken from each subject at two time points separated by ~1 y. Our data indicate that overall V, D, and J gene usage are similar in young and older individuals but that the aging B cell repertoire shows prominent changes in the lengths of IgH CDR3 regions that are present in B cells and in the levels of somatic mutation in Ab sequences. Importantly, the individuals studied in this paper were of known seropositivity status for CMV and EBV chronic infection, and there were sufficient numbers of seropositive and seronegative individuals to allow analysis of the impact of these viral infections on the B cell repertoire, independent of patient age. Our findings highlight an association between CMV seropositivity and levels of Ab gene somatic mutation, whereas EBV seropositivity is correlated with the contribution of persistent clonal populations to the B cell repertoire.

Increased levels of B cells expressing long IgH CDR3 regions in the elderly are seen in our data both in unmutated sequences, and in mutated sequences. Selection against B cells expressing Abs with long CDR3s during the process of generating the naive B cell repertoire has previously been correlated with the higher frequency of autoreactivity in Abs with long CDR3 regions (35). The effects of this selection process can be detected in the difference between the longer CDR3 regions present in unproductively rearranged IgH genes compared with productive expressed IgH sequences (43). The presence of longer CDR3 regions in mutated IgH sequences in the elderly in our data may simply reflect the fact that mutated sequences derive from the initially unmutated pool or could reflect additional age-related impairment in positive selection for B cells expressing Abs with shorter CDR3s in Ag-driven responses. Prior literature has documented increased levels of autoantibodies in the sera of elderly individuals compared with younger individuals (44–46). The results presented here may indicate that increased autoimmunity in the elderly stems from two different defects: impaired selection against Abs with long CDR3 regions in generation of the naive B cell repertoire, and decreased selection against long CDR3 Abs in B cell responses associated with Ab mutation and isotype switching. Other CDR3 features that have been associated with autoreactivity, such as hydrophobicity, did not show age-related changes in our data.

Both IgM and IgG mutated memory B cell pools in our data set show a trend toward age-associated accumulation of highly mutated sequences, although there is considerable variability between individuals in all age groups examined. These findings could
suggest that progressive antigenic exposure stimulating B cell proliferation and targeted Ig gene mutation can lead to the accumulation of increasingly highly mutated IGHV genes over the course of a human lifespan. Overall, there is considerable overlap in the levels of IGHV mutation seen in the B cell repertoires of young and elderly individuals. Implications of the latter result are that most long-lived memory B cells may be formed primarily from the relatively less-mutated members of a clonal proliferation or that memory B cells formed from more highly mutated clone members may have a shorter lifetime in the host. Indeed, very long-lasting memory B cells, such as those specific for the 1918 pandemic influenza strain in nonagenarian or centenarian survivors of the pandemic have shown IGHV mutation levels of ~10%, within the range of mutation frequencies that can be seen in memory B cells in younger individuals (47). The origin of IgM memory B cells is a subject of debate, with some reports proposing novel developmental pathways that are independent of germinal center stimulation, or independent of Ag stimulation entirely, whereas other interpretations describe these cells as the output of immune responses associated with decreased germinal center function or simply a subset of memory B cells typically produced in T cell–dependent immune responses (38, 48). If the generation of IgM memory cells is a consequence of impaired germinal center function, then age-associated increases in IgM mutation frequencies could be a consequence of decreased germinal center activity in the elderly, leading to greater reliance on B cell clonal lineages that show some Ig mutation, but have not been able to isotype switch.

We identified an independent correlation between CMV infection and the mutation levels in both IgM and IgG-expressing B cells in individuals regardless of age. This could be a direct effect of some B cells being specific for CMV, or alternately, could reflect the reported superantigen-like effects reported for CMV phosphoprotein pUL32 or other CMV proteins (49). Arguing against the latter possibility, the described superantigen-like activity of pUL32 was for B cells expressing IGHV1-69 or IGHV3-21 gene

**FIGURE 6.** Phylogenetic visualization of large clones. The large clones are visualized via immunitree. Each arrow points from an ancestral clone to a descendent clone. Sizes of the nodes correspond to the number of reads observed; the blue portion of the pie chart represents Y2, whereas the yellow portion represents Y3. The clones are labeled with the number of mutations from their immediate parent subclone. The root node’s numerical label counts the number of mutations from the best matching germline reference V segment.
and research assistant Kyrsten Spann for scheduling and conducting the study visits.

Disclosures
S.D.B. has consulted for and has stock interest in Immumetrix Co., a private company, but there is no overlap presenting a conflict of interest with the subject of this manuscript. Similarly, S.D.B., A.Z.F., E.L.M., and K.S. have a patent application pending regarding to immunological sequencing, but this does not represent a conflict of interest with the subject of this manuscript.

References

Acknowledgments
We thank Sally Mackey for project, regulatory, and data management, research nurses Sue Swope and Cynthia Walsh, phlebotomist Michele Ugur,