Integrating B Cell Lineage Information into Statistical Tests for Detecting Selection in Ig Sequences

Mohamed Uduman, Mark J. Shlomchik, Francois Vigneault, George M. Church and Steven H. Kleinstein

Supplementary Material	http://www.jimmunol.org/content/suppl/2013/12/26/jimmunol.1301551.DCSupplemental.html
Subscriptions	Information about subscribing to *The Journal of Immunology* is online at: http://jimmunol.org/subscriptions
Permissions	Submit copyright permission requests at: http://www.aai.org/ji/copyright.html
Email Alerts	Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/cgi/alerts/etoc

Integrating B Cell Lineage Information into Statistical Tests for Detecting Selection in Ig Sequences

Mohamed Uduman,* Mark J. Shlomchik,[†,‡] Francois Vigneault,[§,¶] George M. Church,[§] and Steven H. Kleinstein*,[‖]

Detecting selection in B cell Ig sequences is critical to understanding affinity maturation and can provide insights into Ag-driven selection in normal and pathologic immune responses. The most common sequence-based methods for detecting selection analyze the ratio of replacement and silent mutations using a binomial statistical analysis. However, these approaches have been criticized for low sensitivity. An alternative method is based on the analysis of lineage trees constructed from sets of clonally related Ig sequences. Several tree shape measures have been proposed as indicators of selection that can be statistically compared across cohorts. However, we show that tree shape analysis is confounded by underlying experimental factors that are difficult to control for in practice, including the sequencing depth and number of generations in each clone. Thus, although lineage tree shapes may reflect selection, their analysis alone is an unreliable measure of in vivo selection. To usefully capture the information provided by lineage trees, we propose a new method that applies the binomial statistical framework to mutations identified based on lineage tree structure. This hybrid method is able to detect selection with increased sensitivity in both simulated and experimental data sets. We anticipate that this approach will be especially useful in the analysis of large-scale Ig sequencing data sets generated by high-throughput sequencing technologies. *The Journal of Immunology*, 2014, 192: 000–000.

A diverse repertoire of IgRs on B cells allows the adaptive immune response to recognize a universe of Ags. The initial diversity of naive B cells results from a complex genetic rearrangement process in the bone marrow, where the IgR H and L chains are formed through the recombination of V(D)J segments, with additional diversity generated at the junction boundaries (1–3). During T cell–dependent immune responses, the affinity of B cell IgRs is fine-tuned by the dynamic process of affinity maturation. This critical component of adaptive immunity, which normally occurs within the specialized microenvironment of germinal centers (GCs), protects the host from recurring infections and ever-evolving pathogens (4). In the GC, Ag-activated B cells clonally expand and undergo somatic hypermutation (SHM), which introduces an average of approximately one point mutation per division in the Ig gene (5, 6). These mutations can alter the affinity of the IgR, and the resulting diversification provides the substrate for affinity maturation. B cells

acquiring affinity-increasing mutations are preferentially expanded (positive selection), most likely through a survival advantage (7). B cells that acquire mutations that decrease affinity or adversely impact structural integrity of the IgR are removed from the population (negative selection) (8, 9). Multiple rounds of division, mutation, and selection result in populations of high-affinity memory and long-lived plasma B cells, which help to clear the current infection, as well as protect the host against future infections (10–12).

The ability to detect selection, especially positive selection, in experimentally derived Ig sequences is critical to understanding the role of affinity maturation in physiological immune responses, as well as in pathological ones. Autoreactive B cells derived from lupus-prone mice are clonally expanded and carry numerous somatic mutations with a distribution that suggests a role for selection (8, 13). In rheumatoid arthritis, B cells form ectopic GC-like microstructures in the synovium and are the source of autoreactive rheumatoid factors (14, 15). Similarly, expanded B cell clones contribute to CNS immunopathology in multiple sclerosis, and analysis of Ag-driven selection could help to identify mechanisms that drive the disease (16–18). Along with providing insights into disease pathology, the ability to detect selection may have prognostic value, as in the case of B cell lymphomas (19–21).

SHM introduces point mutations into the variable (V) region of Ig gene at the rate of $\sim 10^{-3}$ per base pair per division (5, 6). In the absence of selection, the fraction of mutations that results in an amino acid change (i.e., replacement [R]) can be estimated based on a model of SHM hot/cold spots and substitution bias (22–25). Because only R mutations can modify IgR affinity, we expect selection to skew the observed ratio of R and silent (S) mutations. If the frequency of R mutations is higher than expected, this is assumed to indicate positive selection, whereas a reduced frequency is associated with negative selection. Selection is usually analyzed separately for the CDRs, which are most likely to interact with Ag, and the framework regions (FWRs), which are mainly responsible for maintaining structural integrity of the Ig molecule (13).

*Interdepartmental Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520; †Department of Laboratory Medicine, Yale University School of Medicine, New Haven, CT 06520; ‡Department of Immunobiology, Yale University School of Medicine, New Haven, CT 06520; §Department of Genetics, Harvard Medical School, Boston, MA 02115; ¶AbVitro, Inc., Boston, MA 02116; and ‖Department of Pathology, Yale University School of Medicine, New Haven, CT 06520

Address correspondence and reprint requests to Dr. Steven H. Kleinstein, Department of Pathology, Yale University School of Medicine, 300 George Street, Suite 505, New Haven, CT 06511. E-mail address: steven.kleinstein@yale.edu

The online version of this article contains supplemental material.

Abbreviations used in this article: AUC, area under the curve; FWR, framework region; GC, germinal center; R, replacement; ROC, receiver operating characteristic; S, silent; SHM, somatic hypermutation.

The earliest methods to rigorously test for selection operated on single sequences and compared the observed frequency of R mutations with their expected frequency under the null hypothesis of no selection using a binomial statistical framework (13, 26, 27). However, these methods have been criticized for their high false-positive rate (28, 29). This low specificity was due, at least in part, to a failure to fully account for the intrinsic biases of SHM, which can give the appearance of selection. In previous work (25), we developed the Focused binomial test, which improved specificity through a better model of intrinsic biases and corrected for the effects of cross-talk between positive and negative selection using a carefully derived null model. A further improvement, the Focused Z-score method, allowed for the combined analysis of multiple sequences, potentially using different germline V(D)J segments, which increased the sensitivity for the detecting selection while maintaining specificity (30). All of these methods result in a *p* value that can be used to reject the null hypothesis that the observed mutation pattern occurred in the absence of selection. Most recently, we developed a statistical framework for Bayesian estimation of Ag-driven SELectIoN (BASELINe), which shifts the problem from one of simply detecting selection to one of quantifying selection (31). Despite these improvements, these R/S-based methods generally suffer from low sensitivity (32).

An alternate approach to detect selection is based on the analysis of B cell lineage trees, which can be created from sets of clonally related Ig sequences using methods such as maximum parsimony (33–35). Unlike the R/S-based methods, lineage trees preserve the substantial information present in the pattern of shared and unique mutations among clonal sequences, which reflects the dynamics of the underlying diversification and selection processes (5, 8, 36, 37). Lineage tree analysis was instrumental in establishing GCs as the site of SHM, as well as the concept of step-wise affinity maturation (38, 39). B cell lineage tree shapes have been suggested to reflect the degree of underlying selection, with "pruned" trees reflecting greater selection compared with "bushy" trees (37, 40). Simulation studies have been carried out to identify linkages between specific lineage tree shape measures and underlying selection (34, 41). These tree shape measures have been applied in several studies (42–44) to compare selection in different lymphoid tissues, autoimmune diseases, and lymphomas. However, the specificity of these methods remains uncertain because current approaches do not account for the potentially confounding influences of experimental factors, such as the number of generations in the clone and the number of sequences sampled, which might give the appearance of spurious selection.

In this article, we quantify the relationship between B cell lineage tree shape measures and immune selection. Despite the correlation between several tree shape measures and selection, we show that existing measures are affected by experimental factors likely to lead to poor specificity in practice. This problem is overcome by using a hybrid method that combines lineage tree information within an R/S-based framework that significantly improves the sensitivity for detecting selection over existing methods.

Materials and Methods

Simulation model of B cell affinity maturation

The simulation is based on the "Clone" model and its more recent extensions (6, 9, 25). It includes B cell clonal expansion, SHM, and affinity-dependent death. The simulation is seeded with a single B cell and proceeds in discrete, synchronous generations. During each generation, all cells divide once and accumulate a Poisson distributed number of mutations (average $\mu = 0.3$). The impact of every mutation is determined stochastically, according to the distribution described in figure 1 in Ref. 25.

Briefly, the probability that a mutation falls in the FWR is given by $pFWR = 0.75$; otherwise, it falls in the CDR ($1 - pFWR = pCDR = 0.25$). Within each region, the probability of a mutation being an R is given by $pR = 0.75$; otherwise, it is an S ($1 - pR = pS = 0.25$). All S mutations are neutral. R mutations that fall in the FWR are considered lethal with probability $\lambda = \{0, 0.5\}$, and cells that accumulate such mutations are removed at the end of each generation. A value of $\lambda = 0$ indicates no negative selection, whereas $\lambda = 0.5$ is used for realistic levels of negative selection (9, 13).

Negative selection is modeled by death (or removal) of the B cells that acquire lethal mutations at the end of each generation. Independent of lethal mutations, a fraction of B cells die at the end of every generation. R mutations in the CDR are considered advantageous (i.e., affinity increasing) with probability $fCDR\_A = 0.75$, and these mutations provide a survival advantage to the cell by decreasing the death rate (7). Specifically, the death rate (d_i) for a given B cell (i) is given by:

$$d_i = \frac{d_{max}}{s^{a_i}}$$

where $d_{max} = 0.4$ is the maximum death rate, a_i is the number of advantageous mutations accumulated by the cell, and the selection factor ($s = \{1, \ldots, 7\}$) determines the cumulative effect of advantageous mutations. An s value of 1 indicates no positive selection in the simulation, and increasing s causes each mutation to have a greater proportional effect on the death rate. Moderate values of s produce the greatest average affinity increases, with $s = 7$ being optimal (data not shown).

Generating lineage trees

During the simulation, the entire B cell lineage was tracked. After D generations, q cells are randomly sampled from the population, and the remaining cells are pruned from the lineage. The resulting tree, which is a subtree of the exact simulated lineage, is comparable to a lineage tree obtained using experimentally derived sequences created using maximum parsimony (6). Unless otherwise specified $D = 22$ and $q = 20$. All of the simulated data used in this study are available at: http://clip.med.yale.edu/papers/Uduman2013JI.

Lineage trees–shape measures

Twenty-nine graph theoretical properties were computed to quantify the shape of each lineage tree. These shape measures involve terms based on counting various types of nodes (sequences) and the distances (number of mutations) between them. In many cases, the maximum, minimum, and average values for each shape measure were also calculated. For example, the numbers of mutations from the root to the leaves of a lineage tree have an average (across all leaves), a maximum, and a minimum. Although all 29 shape measures were thoroughly analyzed, detailed results are presented for the seven shape measures that were reported to be highly correlated with selection (45) and three additional measures that we found to be significantly correlated (Table I). Please refer to figure 1 in Ref. 45 for a visual representation of lineage tree shape measures.

Quantifying the influence of selection on lineage tree shape measures

To determine the effect of selection on a lineage tree shape measure, its empirical probability distribution was determined from a simulation that includes selection and compared with one in the absence of selection ($\lambda = 0$, $s = 1$). Several kinds of selection were considered: only positive selection ($\lambda = 0$, $s = 7$), only negative selection ($\lambda = 0.5$, $s = 1$), and a combination of negative and positive selections ($\lambda = 0.5$, $s = 7$). The shift in these two probability distributions was quantified using numeric integration to estimate the area under the curve (AUC), as described in equation 7 of Ref. 31.

Quantitative framework for detecting selection using lineage tree shape measures

To estimate the true- and false-positive rates for a shape measure's ability to detect selection, we first determined the distribution of the shape measure under the null hypothesis of no selection ($\lambda = 0$, $s = 1$). Based on this null distribution, a cutoff value for the measure is given by the 5% tail of the distribution. The direction of the tail was determined beforehand based on the expected influence of selection on the tree shape measure's distribution. For each lineage tree simulated in the presence of selection, the null hypothesis of no selection was rejected if the tree shape measure was more extreme than the cutoff. The fraction of trees rejected over many simulated lineage trees provides the sensitivity or 1-specificity of the shape measure's ability to detect selection.

IgG sequence data set from a healthy male

PBMCs were isolated from a healthy Personal Genome Project subject (46), and RNA was extracted from B lymphocytes. IgH genes were amplified using gene-specific reverse transcription and sequenced by emulsion PCR and 454 GS FLX. The sequences were processed through IMGT/HighV-QUEST (47) to identify V(D)J germline gene segments and aligned using IMGT numbering (48). Sequences were further filtered for functionality, quality redundancy, and isotype, resulting in 7738 unique IgG sequences. Clonally related sequences were identified by first grouping based on common V and J gene family assignments and on junction region length. Within these larger groups, sequences differing from one another, in the junction region, by three or fewer point mutations were defined as clones. This resulted in 1369 clones containing multiple sequences (ranging from 2 to 36 sequences), with an average size of 2.9 sequences/clone. These data are available at: http://clip.med.yale.edu/papers/Uduman2013JI. BASELINe (31), using the Focused test statistic (25, 30), was applied to estimate selection strength.

Nonterminal branch mutations

Nonterminal branch mutations are defined as the set of mutations that occurs on branches of the lineage tree that are not connected to a leaf (Supplemental Fig. 3B). For computational efficiency, the set of nonterminal branch mutations was approximated as those that are shared between more than one sequence in a clone.

Results

Influences of selection on lineage tree shape measures

It is not possible to link tree shape measures with selection using experiments alone, because the parameters that influence selection are not easily modified or controlled. Instead, a computer simulation (described in *Materials and Methods*) was used to produce synthetic data sets with known levels of positive and/or negative selection pressures. Importantly, the simulation also allowed us to control experimental factors, such as number of sampled sequences (q) and cell generations (D). Using the simulation, we evaluated the influence of selection on 29 lineage tree shape measures, many of which were proposed previously to correlate with selection (34, 45). Some of these shape measures, such as the average out-degree (number of children per node; OD_Avg), increase with both positive and

negative selection (Fig. 1A). Other shape measures, such as the distance between adjacent split nodes (DASN_Avg) have a negative correlation with selection (Fig. 1D). Of the 29 lineage tree shape measures, a subset of 11 measures (Fig. 1, Supplemental Fig. 1, Table I) was chosen for detailed analysis based on either their ability to detect positive selection in simulated data with high sensitivity or having previously been reported as highly correlated with selection (45).

Although many tree shape measures were clearly modulated by selection, most exhibited properties that were not ideal for detecting positive and negative selection in practice. For example, the average distance between adjacent split nodes (DASN_Avg) decreases with both positive and negative selection, so this measure cannot differentiate between these different influences. Other shape measures were affected by positive and negative selection in opposing manners, allowing these forces to cancel each other out. This group included average outdegree (OD_Avg), average and minimum distance from the leaf to the split node closest to the root (DLFSN_Avg and DLFSN_Min), and the average and minimum path length (distance from the root to the leaves of the tree; PL_Avg and PL_Min). Interestingly, several other tree shape measures previously proposed to correlate with selection (45) did not show such a relationship in our model, including the minimum distance between adjacent split nodes (forks) in the trees (DASN_Min). Thus, although selection influences many lineage tree shape measures, it is not easy to interpret most of these in practice.

The most promising measures included the minimum distance from the root to any split node (DRSN_Min) and the distance from the root to the first node in the tree (T). These shape measures were increased by positive selection but were not influenced by negative selection (Table I). However, the influence of positive selection on these measures was minor (AUC ~ 0.5). The most powerful tree shape measure was the ratio of branching early versus late in the lineage (OD-Ratio). This measure was specifically influenced by positive selection and, unlike DRSN_Min and T, the AUC increased significantly

FIGURE 1. The influence of positive and negative selection on lineage tree shape measures. The probability distribution of OD_Avg (**A**), PL_Avg (**B**), OD-Ratio (**C**), and DASN_Avg (**D**) was computed for simulated data under different selection pressures: neutral selection (gray shaded area; $s = 1$, $\lambda = 0$), only positive selection (green line; $s = 7$, $\lambda = 0$), only negative selection (red line; $s = 1$, $\lambda = 0.5$), and the combination of positive and negative selections (blue line; $s = 7$, $\lambda = 0.5$). $q = 20$ and $D = 22$ for all simulations.
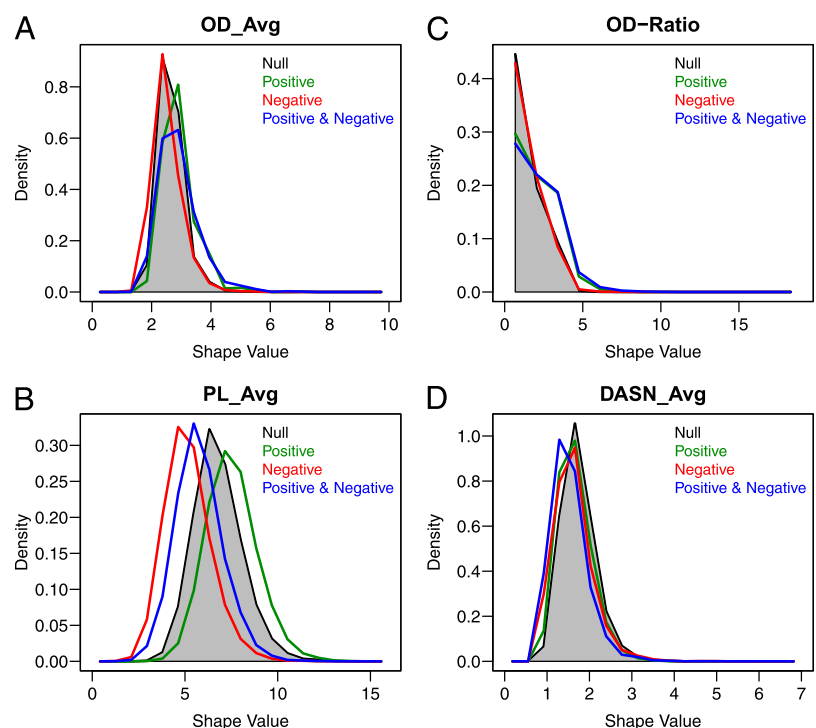
Table I. Lineage tree shape measures

Shape Measure	Shape Description	Positive Selection Only (s=7, λ=0)	Negative Selection Only (s=1, λ=0.5)
OD-Root*	The outgoing degree (the number of nodes branching out) at the root. This represents the number of unique sequences branching from the germline.		
OD_Avg*	The average of outgoing degrees at each node of the tree. This represents the average number of unique sequences sharing an immediate common ancestor.		
OD-Ratio	The ratio of OD_Root and the average outgoing degree of the tree excluding the root. This represents the ratio of branching early versus late in the lineage.		
PL_Min*	The minimum length from root to leaf. This represents the minimum number of mutations from the germline to a leaf in the lineage.		
PL_Avg*	The average of length from root to each of the leaves. This represents the average number of mutations from the germline to all the leaves in the lineage.		
DLFSN_Min*	The minimum distance from the leaf to the first split node (from the root of the tree). This represents the minimum number of mutations between two sequences found as leaves in the lineage and their closest common ancestor.		
DLFSN_Avg*	The average distance from all the leaves to the first split node (from the root of the tree This represents the average number of mutations between any two sequences found as leaves in the lineage and their closest common ancestor.		
DASN_Min*	The minimum distance between adjacent split nodes of the tree. This represents the minimum number of mutations between two sequences that each has multiple children in the lineage.		
DASN_Avg*	The average of distance between adjacent split nodes of the tree. This represents the average number of mutations between two sequences that each has multiple children in the lineage.		
DRSN_Min*	The minimum distance from the root to any split node. This represents the minimum number of mutations from the germline to any sequence that has multiple children.		
T*	The number of mutations on the single branch leading out of the germline. If multiple branches lead from the germline, this value is zero. This represents the number of mutations on the trunk of the tree.		

| ≤ 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 0.95 | | 0.95 | 0.9 | 0.8 | 0.7 | 0.6 | ≤ 0.5 |

Lineage tree shape measures, their description, and heat map of the AUC quantifying the impact of positive and negative selection on the shape measures. Red indicates increased shape measure values, and green indicates decreased values.
*Shape measures from Ref. 45. Italicized shape measures were reported to be significantly correlated with selection.

(AUC = 0.85). Thus, OD-Ratio is a good candidate for detecting positive selection in practice.

The sensitivity of lineage tree shape measures for detecting positive selection

Having identified lineage tree shape measures with the desired properties, we next sought to quantify their sensitivity for detecting selection. The true positive rate was estimated by comparing the distribution of shape measures for simulated trees in the presence of selective pressures with a null distribution generated in the absence of any selective pressure (see *Materials and Methods*). Surprisingly, six of the shape measures proposed by Shahaf et al. (45), including all of the minimum value measures (DASN_Min, DLFSN_Min, DRSN_Min and PL_Min), along with OD-Root, showed no ability to detect selection, even for the strongest positive-selection pressure (Fig. 2). Of the remaining previously proposed measures, even the best, DLFSN_Avg, was able to detect only 13.9% of the lineage trees as being selected (for *s* = 7). Similar performance was estimated for OD_Avg (13.0%) and PL_Avg (12.7%). OD-Ratio, a new tree shape measure proposed in this study, had the highest sensitivity (17.7%). Both OD-Ratio and DLFSN_Avg detected selection better than the Focused Z-score (30), an R/S-based method that detected selection in
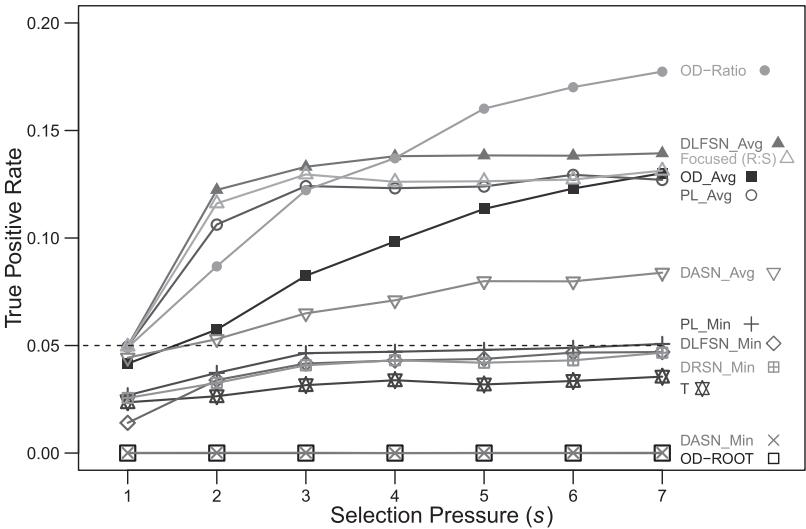
13.1% of the lineage trees (Fig. 2). Overall, these results demonstrate the potential of lineage tree shapes to detect selection with increased sensitivity.

The confounding influence of experimental factors on lineage tree shape measures

In practice, lineage tree analysis involves comparing the shape measure distributions between two or more groups of trees (e.g., case and control) (34, 41–43, 49). Unlike the analysis presented in the previous section, in which the null distribution only differed in parameters related to selection (λ and *s*), real experimental groups exhibit many differences. These experimental factors include the number of sequences used to create the lineage tree (sequencing depth) and the generation in the B cell clone at which the sequences were sampled. We hypothesized that differences in these factors alone could give the appearance of selection.

To evaluate the potential effect of sampling depth on lineage tree shape measures, several synthetic data sets were generated with varying numbers of sequences (*q* = {10,20,50,100,200,300}). In all cases, selection was absent, and the cells were sampled *D* = 22 generations. For each number of sequences (*q*), the number of statistically significant trees was calculated using *q* = 300 to generate the control. Because selection is absent in all cases, this

FIGURE 2. Sensitivity of lineage tree shape measures for detecting selection. Fifteen thousand trees were simulated for a range of positive selection pressures ($s = \{1,...,7\}$, $\lambda = 0$), and the power to detect selection was assessed for each shape measure (individual lines). True-positive rates were calculated by comparison with the null distribution of neutral selection ($s = 1$, $\lambda = 0$) using a p value cutoff of $\alpha = 0.05$, as described in *Materials and Methods*. $q = 20$ and $D = 22$ for all simulations.

provides an estimate for the false-positive rate as a function of the number of sequences. Of the five lineage tree shape measures that exhibited a reasonable sensitivity for detecting positive selection (Fig. 2), DASN_Avg is most affected by the sampling depth, with a false positive rate of 50.3% when trees contained 10 sequences ($q = 10$) (Fig. 3A). OD_Avg, DLFSN_Avg, and PL_Avg also exhibited high false-positive rates of 35.1, 22.6, and 17.1%, respectively (Fig. 3A). Even OD-Ratio, the shape measure with the highest sensitivity in detecting positive selection under ideal conditions, yielded a false-positive rate of 13.9% (Fig. 3A).

Experimental data sets are also likely to vary in the number of generations (cell divisions), even when the sampling time can be fixed (e.g., equivalent number of days postimmunization). Even more problematic, comparisons are often carried out for different disease states (e.g., see Refs. 42, 43, 49), in which case it is not clear how long the B cell clones have been dividing. To evaluate the potential confounding effects of this variability, synthetic data sets were generated by sampling exactly $q = 20$ sequences at varying numbers of generations ($D = \{12,14,16,18,20,22\}$). Comparing these simulated tree shapes with a control based on $D = 22$ generations clearly shows that differences in the number of generations yield high false-positive rates for nearly all shape mea-

sures (Fig. 3B). PL_Avg and DLFSN_Avg were affected the most, with false-positive rates of 76.7 and 71.2%, respectively, when trees differed by 10 division cycles. High false-positive rates were also found for DASN_Avg (24.2%) and OD_Avg (20.7%), whereas the best performance was shown by OD-Ratio, which still had an unacceptably high false-positive rate of 9.8%. Overall, these results demonstrate that, even in the absence of any differences in selection pressure, variation in sampling depth and/or the number of generations can falsely give the appearance of selection. Thus, using lineage tree shape measures without properly accounting for such experimental variability can lead to incorrect interpretation of selection status.

Normalizations to correct for the effects of sampling depth and the number of generations were suggested by previous work (34, 41, 49). Specifically, it was proposed that the effect of sampling depth could be normalized by dividing the shape measure by the number of nodes (sequences) in a lineage tree (34), whereas the effects of generation could be scaled by the average path length (PL_Avg) (49). The goal of such normalization is to allow lineage tree data from different sources and conditions to be compared without bias. However, despite their use in several studies (34, 41–44, 49), these proposed normalizations have never been tested
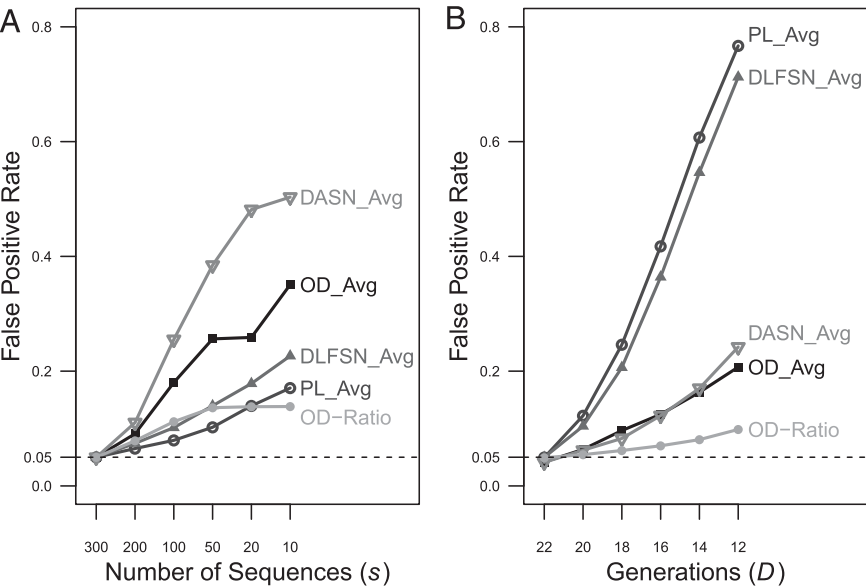

FIGURE 3. Low specificity of lineage tree–based approaches due to confounding experimental factors. Fifteen thousand trees were simulated assuming neutral selection ($s = 1$, $\lambda = 0$) for a range of the number of sequences sampled (**A**) and the generation at which the sequences were sampled (**B**). False-positive rates were calculated by comparison with the null distribution of neutral selection ($s = 1$, $\lambda = 0$) with $q = 300$ (A) or $D = 22$ (B) using a p value cutoff of $\alpha = 0.05$, as described in *Materials and Methods*.

directly. Results from simulated lineage trees suggest that the application of these normalizations actually decreases specificity (Supplemental Fig. 2). Note the trivial exceptions of DLFSN_Avg and PL_Avg, which achieved 100% specificity but lost sensitivity, because both of these shape measures approximate the average number of mutations, and the normalized measure is, thus, always around one (Supplemental Fig. 2B). In summary, currently proposed normalization methods do not correct for the confounding effects of sampling depth or generation and likely even decrease specificity.

Filtering for nonterminal branch mutations improves R/S-based methods

Given the specificity problems outlined above, along with the comparable sensitivity of lineage tree shape measures and R/S-based methods for detecting selection, we sought an alternate strategy that could combine the strengths of both approaches. We reasoned that the sensitivity of R/S-based methods might be improved by using information from lineage trees to identify a subset of mutations most likely to be selected. Because mutation and selection are ongoing, step-wise processes (5), it is expected that mutations appearing toward the root of lineage trees have been subject to the strongest selection. Indeed, we showed previously that recurrent mutations (most likely to increase affinity) tend to occur at "nonterminal" branches within a lineage tree (i.e., branches that are not connected directly to leaves) (7). This relationship is a direct consequence of the clonal expansion and selection process, because it also can be observed in simulated data (Supplemental Fig. 3). In this case, mutations appearing near lineage tree leaves do not show any increase in the R frequency, and their inclusion in R/S-based tests may actually decrease sensitivity. To test this hypothesis, we implemented an R/S-based test (the Focused Z-score method) using mutations exclusively from nonterminal branches in the lineage tree and applied this method on simulated trees to generate receiver operating characteristic (ROC) curves. These ROC curves show that using only nonterminal branch mutations improves the trade-off between sensitivity and specificity compared with using all unique mutations in the lineage tree (Fig. 4). For example, at the standard α cutoff of 0.05, using mutations from nonterminal branches yields a sensitivity of 23% compared with 18% when using all mutations. Importantly, this hybrid method maintains the expected specificity.

To verify that the proposed hybrid approach also can provide increased sensitivity for detecting selection in vivo, we analyzed isotype-switched IgG sequences derived from a blood sample from a healthy individual (see *Materials and Methods*). As expected, using mutations from nonterminal branches led to a significant increase in the estimated positive selection strength in the CDR ($p = 1.80E-05$) and negative selection strength in the FWR ($p = 2.74E-04$) (Fig. 5). Thus, when lineage tree information is available, focusing on the set of nonterminal branch mutations can provide a significant increase in sensitivity for detecting and quantifying selection.

Discussion

The most commonly used methods for detecting selection in Ig sequences are based on a binomial statistical framework that compares the observed frequency of R mutations with that expected under the null hypothesis of no (or neutral) selection. Despite recent advances, the overall sensitivity of these methods remains low (25, 30, 31). Methods based on the analysis of lineage trees provide an alternative approach to obtain insights into B cell clonal expansion and Ag-driven selection (34, 41). In this study, we applied a sim-
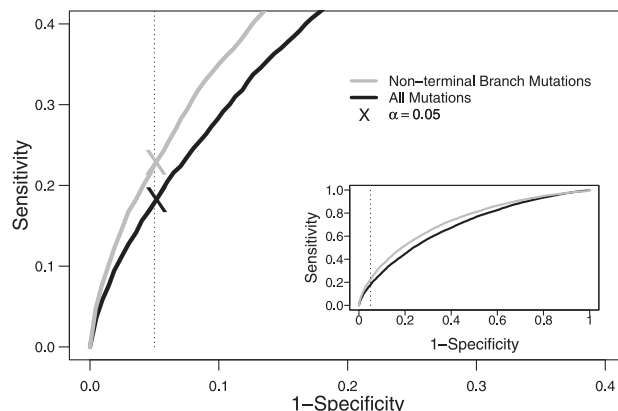

FIGURE 4. Performance of the hybrid method for detecting selection on simulated data. ROC curves comparing the performance of the Focused Z-Test (30) when using all mutations (black curve) or only mutations from nonterminal branches in the lineage tree (gray curve). Sensitivity is based on detecting positive selection in simulated data with both positive and negative selection ($s = 7$, $\lambda = 0.5$). Specificity is based on detecting positive selection in simulated data with only negative selection ($s = 1$, $\lambda = 0.5$). The dotted line indicates the expected specificity at $\alpha = 0.05$, and the "X" indicates the actual specificity at $\alpha = 0.05$ on the corresponding ROC curves. The *inset* shows the full ROC curves for the same data. $q = 20$ and $D = 22$ for all simulations.

ulation-based approach to identify several shortcomings in the use of lineage tree shape measures to detect selection. Rather than depending on these shape measures alone, we propose integrating this clonal information into an R/S-based statistical framework to improve the ability to detect selection in Ig sequences.

Qualitative analyses of B cell lineage tree shape measures were used in many studies to infer the dynamics of the underlying immune response. Within the past decade, a plethora of graph-theoretical measures have been proposed and applied to quantify lineage tree shapes and elucidate differences in selection dynamics (e.g., between samples from healthy individuals and those with various disease conditions) (42–44, 49). Shahaf et al. (45) dem-
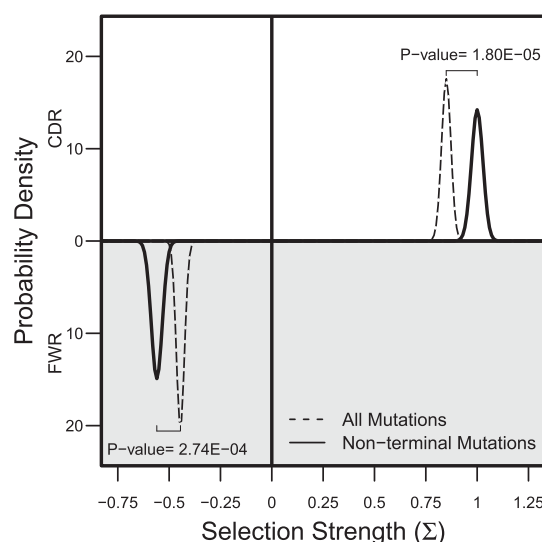

FIGURE 5. Hybrid method detects stronger selection in vivo. Estimated selection strength in isotype-switched IgG sequences from a healthy male. BASELINe was used to quantify selection in the CDR (top half of the plot) and the FWR (bottom half of the plot) using the Focused test statistic. The curves depict the posterior probability distributions of the selection strength using all unique mutations in the lineage tree (dashed) or only unique mutations on nonterminal branches (solid).

onstrated that several of these shape measures are statistically correlated with selection using a mathematical model of GC population dynamics. However, this previous study did not account for experimental factors, such as differences in sequencing depth or the time of sampling, which could lead to clones with widely varying numbers of cell divisions. Furthermore, the lineage trees analyzed by Shahaf et al. (45) were generated using all of the B cells in the simulation, so that some of the efficient shape measures simply reflect differences in the overall population size. Experimentally derived lineage trees are always a subsample of the full population. We hypothesized that differences in these experimental factors could confound the relationship between lineage tree shape measures and selection.

We developed a computer simulation of B cell clonal expansion and selection that included key experimental factors, such as the number of sequences sampled and the number of cell divisions/clone (associated with the time of sampling). Under conditions of strong positive selection, only two of the eight shape measures previously associated with selection by Shahaf et al. (45) remained significantly correlated (OD_Avg and PL_Avg, $p < 0.05$). Three additional measures (DLFSN_Avg, DASN_Avg, and OD-Ratio) also were significantly associated with selection. Although the absolute sensitivity estimated by our simulation approach can depend on parameters such as the number of sequences sampled, thus explaining some of the differences in efficiency between our study and that of Shahaf et al. (45), we always found that all of these tree shape measures had sensitivities that were comparable to the performance of the Focused Z-score method on the same data. Thus, detecting positive selection using lineage tree shape measures does not offer an obvious advantage over R/S-based methods in terms of sensitivity. In fact, a significant disadvantage of lineage tree–based methods concerns their reliance on a control group to make inferences. For example, the distribution of B cell lineage tree shape measures from an autoimmune setting could be compared with healthy samples. In such cases, it is difficult (if not impossible) to control for the number of sequences in each tree and the number of cell divisions/clone. This is a significant issue because we showed that differences in these experimental factors can falsely give the appearance of selection, and standard normalization approaches do not eliminate this effect. Thus, in practice, the performance of tree-based methods is likely to be significantly worse compared with R/S-based methods.

Despite these issues, the structure of B cell lineage trees can provide important information about the underlying biological process. We propose to leverage this information through a hybrid method that detects selection using an R/S-based test focused on the set of nonterminal branch mutations from a lineage tree. When tested on simulated data, this method exhibited increased sensitivity and the expected specificity. However, the actual performance of the method will depend on the affinity landscape of the response being analyzed. In the model, the landscape can be altered through the selection factor parameter (s). The strongest average selection occurs for $s \sim 7$, in which case the first advantageous mutation provides most of the selective advantage. To see that this is the case, consider that cells in the simulation start out with a 40% probability of dying at each generation and, when $s = 7$, this decreases to 6% after the first advantageous mutation (a decrease of 34%). The second advantageous mutation decreases the probability of death to 1% per generation (a decrease of only 5%), and further decreases provide little benefit given the relatively short time scale of B cell clonal expansion. Under these conditions, it is easy to see why there is little selective pressure on mutations near the leaves of a lineage tree. To assume a limit on affinity maturation and a scheme of diminishing returns seems reasonable, although whether real immune responses operate in a similar

fashion is unclear, because the shape of actual affinity landscapes has not been determined. For these reasons, it was critical to demonstrate the benefits of the hybrid method on experimental data. Indeed, using mutations from nonterminal branches in isotype-switched IgG sequences from a healthy individual led to a significant increase in the estimated positive- and negative-selection strength in the CDR and FWR, respectively. An additional benefit of the hybrid method is that many PCR errors, which are likely to appear on terminal branches, will be filtered.

With greater prevalence and affordability of next-generation sequencing technologies, larger and deeper sequencing of Ig repertoires is becoming more common. The hybrid approach developed in this study can take advantage of the clonal information inherent in these data and provide analyses that will lead to more powerful and accurate insights into the evolution of B cell immune responses and malignant clones.

Disclosures

F.V. and G.M.C. have financial interests in AbVitro, Inc.

References

1. Weigert, M., L. Gatmaitan, E. Loh, J. Schilling, and L. Hood. 1978. Rearrangement of genetic information may produce immunoglobulin diversity. *Nature* 276: 785–790.
2. Weigert, M., R. Perry, D. Kelley, T. Hunkapiller, J. Schilling, and L. Hood. 1980. The joining of V and J gene segments creates antibody diversity. *Nature* 283: 497–499.
3. Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature* 302: 575–581.
4. Longo, N. S., and P. E. Lipsky. 2006. Why do B cells mutate their immunoglobulin receptors? *Trends Immunol.* 27: 374–380.
5. McKean, D., K. Huppi, M. Bell, L. Staudt, W. Gerhard, and M. Weigert. 1984. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc. Natl. Acad. Sci. USA* 81: 3180–3184.
6. Kleinstein, S. H., Y. Louzoun, and M. J. Shlomchik. 2003. Estimating hypermutation rates from clonal tree data. *J. Immunol.* 171: 4639–4649.
7. Anderson, S. M., A. Khalil, M. Uduman, U. Hershberg, Y. Louzoun, A. M. Haberman, S. H. Kleinstein, and M. J. Shlomchik. 2009. Taking advantage: high-affinity B cells in the germinal center have lower death rates, but similar rates of division, compared to low-affinity cells. *J. Immunol.* 183: 7314–7325.
8. Shlomchik, M. J., A. Marshak-Rothstein, C. B. Wolfowicz, T. L. Rothstein, and M. G. Weigert. 1987. The role of clonal selection and somatic mutation in autoimmunity. *Nature* 328: 805–811.
9. Shlomchik, M. J., P. Watts, M. G. Weigert, and S. Litwin. 1998. Clone: a Monte-Carlo computer simulation of B cell clonal expansion, somatic mutation, and antigen-driven selection. *Curr. Top. Microbiol. Immunol.* 229: 173–197.
10. MacLennan, I. C. 1994. Germinal centers. *Annu. Rev. Immunol.* 12: 117–139.
11. Rajewsky, K. 1996. Clonal selection and learning in the antibody system. *Nature* 381: 751–758.
12. Takahashi, Y., P. R. Dutta, D. M. Cerasoli, and G. Kelsoe. 1998. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. V. Affinity maturation develops in two stages of clonal selection. *J. Exp. Med.* 187: 885–895.
13. Shlomchik, M. J., A. H. Aucoin, D. S. Pisetsky, and M. G. Weigert. 1987. Structure and function of anti-DNA autoantibodies derived from a single autoimmune mouse. *Proc. Natl. Acad. Sci. USA* 84: 9150–9154.
14. Kim, H. J., and C. Berek. 2000. B cells in rheumatoid arthritis. *Arthritis Res.* 2: 126–131.
15. Silverman, G. J., and D. A. Carson. 2003. Roles of B cells in rheumatoid arthritis. *Arthritis Res. Ther.* 5(Suppl. 4): S1–S6.
16. Qin, Y., P. Duquette, Y. Zhang, P. Talbot, R. Poole, and J. Antel. 1998. Clonal expansion and somatic hypermutation of V(H) genes of B cells from cerebrospinal fluid in multiple sclerosis. *J. Clin. Invest.* 102: 1045–1050.
17. Zuckerman, N. S., H. Hazanov, M. Barak, H. Edelman, S. Hess, H. Shcolnik, D. Dunn-Walters, and R. Mehr. 2010. Somatic hypermutation and antigen-driven selection of B cells are altered in autoimmune diseases. *J. Autoimmun.* 35: 325–335.
18. von Büdingen, H.-C., T. C. Kuo, M. Sirota, C. J. van Belle, L. Apeltsin, J. Glanville, B. A. Cree, P.-A. Gourraud, A. Schwartzburg, G. Huerta, et al. 2012. B cell exchange across the blood-brain barrier in multiple sclerosis. *J. Clin. Invest.* 122: 4533–4543.
19. Lossos, I. S., C. Y. Okada, R. Tibshirani, R. Warnke, J. M. Vose, T. C. Greiner, and R. Levy. 2000. Molecular analysis of immunoglobulin genes in diffuse large B-cell lymphomas. *Blood* 95: 1797–1803.
20. Degan, M., R. Bomben, M. D. Bo, A. Zucchetto, P. Nanni, M. Rupolo, A. Steffan, V. Attadia, P. F. Ballerini, D. Damiani, et al. 2004. Analysis of IgV gene mutations in B cell chronic lymphocytic leukaemia according to antigen-

driven selection identifies subgroups with different prognosis and usage of the canonical somatic hypermutation machinery. *Br. J. Haematol.* 126: 29–42.

21. Bomben, R., M. Dal Bo, A. Zucchetto, E. Zaina, P. Nanni, P. Sonego, G. Del Poeta, M. Degan, and V. Gattei. 2005. Mutational status of IgV(H) genes in B-cell chronic lymphocytic leukemia and prognosis: percent mutations or antigen-driven selection? *Leukemia* 19: 1490–1492.

22. Smith, D. S., G. Creadon, P. K. Jena, J. P. Portanova, B. L. Kotzin, and L. J. Wysocki. 1996. Di- and trinucleotide target preferences of somatic mutagenesis in normal and autoreactive B cells. *J. Immunol.* 156: 2642–2652.

23. Shapiro, G. S., K. Aviszus, D. Ikle, and L. J. Wysocki. 1999. Predicting regional mutability in antibody V genes based solely on di- and trinucleotide sequence composition. *J. Immunol.* 163: 259–268.

24. Cowell, L. G., and T. B. Kepler. 2000. The nucleotide-replacement spectrum under somatic hypermutation exhibits microsequence dependence that is strand-symmetric and distinct from that under germline mutation. *J. Immunol.* 164: 1971–1976.

25. Hershberg, U., M. Uduman, M. J. Shlomchik, and S. H. Kleinstein. 2008. Improved methods for detecting selection by mutation analysis of Ig V region sequences. *Int. Immunol.* 20: 683–694.

26. Chang, B., and P. Casali. 1994. The CDR1 sequences of a major proportion of human germline Ig VH genes are inherently susceptible to amino acid replacement. *Immunol. Today* 15: 367–373.

27. Lossos, I. S., R. Tibshirani, B. Narasimhan, and R. Levy. 2000. The inference of antigen selection on Ig genes. *J. Immunol.* 165: 5122–5126.

28. Dunn-Walters, D. K., and J. Spencer. 1998. Strong intrinsic biases towards mutation and conservation of bases in human IgVH genes during somatic hypermutation prevent statistical analysis of antigen selection. *Immunology* 95: 339–345.

29. Bose, B., and S. Sinha. 2005. Problems in using statistical analysis of replacement and silent mutations in antibody genes for determining antigen-driven affinity selection. *Immunology* 116: 172–183.

30. Uduman, M., G. Yaari, U. Hershberg, J. A. Stern, M. J. Shlomchik, and S. H. Kleinstein. 2011. Detecting selection in immunoglobulin sequences. *Nucleic Acids Res.* 39(Web Server issue): W499–504.

31. Yaari, G., M. Uduman, and S. H. Kleinstein. 2012. Quantifying selection in high-throughput Immunoglobulin sequencing data sets. *Nucleic Acids Res.* 40: e134.

32. MacDonald, C. M., L. Boursier, D. P. D'Cruz, D. K. Dunn-Walters, and J. Spencer. 2010. Mathematical analysis of antigen selection in somatically mutated immunoglobulin genes associated with autoimmunity. *Lupus* 19: 1161–1170.

33. Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9: 1657–1659.

34. Dunn-Walters, D. K., A. Belelovsky, H. Edelman, M. Banerjee, and R. Mehr. 2002. The dynamics of germinal centre selection as measured by graph-theoretical analysis of mutational lineage trees. *Dev. Immunol.* 9: 233–243.

35. Barak, M., N. S. Zuckerman, H. Edelman, R. Unger, and R. Mehr. 2008. IgTree: creating Immunoglobulin variable region gene lineage trees. *J. Immunol. Methods* 338: 67–74.

36. Sablitzky, F., G. Wildner, and K. Rajewsky. 1985. Somatic mutation and clonal expansion of B cells in an antigen-driven immune response. *EMBO J.* 4: 345–350.

37. Jacob, J., and G. Kelsoe. 1992. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. II. A common clonal origin for periarteriolar lymphoid sheath-associated foci and germinal centers. *J. Exp. Med.* 176: 679–687.

38. Kocks, C., and K. Rajewsky. 1988. Stepwise intraclonal maturation of antibody affinity through somatic hypermutation. *Proc. Natl. Acad. Sci. USA* 85: 8206–8210.

39. Jacob, J., R. Kassir, and G. Kelsoe. 1991. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. I. The architecture and dynamics of responding cell populations. *J. Exp. Med.* 173: 1165–1175.

40. Jacob, J., J. Przylepa, C. Miller, and G. Kelsoe. 1993. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. III. The kinetics of V region mutation and selection in germinal center B cells. *J. Exp. Med.* 178: 1293–1307.

41. Dunn-Walters, D. K., H. Edelman, and R. Mehr. 2004. Immune system learning and memory quantified by graphical analysis of B-lymphocyte phylogenetic trees. *Biosystems* 76: 141–155.

42. Steiman-Shimony, A., H. Edelman, M. Barak, G. Shahaf, D. Dunn-Walters, D. I. Stott, R. S. Abraham, and R. Mehr. 2006. Immunoglobulin variable-region gene mutational lineage tree analysis: application to autoimmune diseases. *Autoimmun. Rev.* 5: 242–251.

43. Tabibian-Keissar, H., N. S. Zuckerman, M. Barak, D. K. Dunn-Walters, A. Steiman-Shimony, Y. Chowers, E. Ofek, K. Rosenblatt, G. Schiby, R. Mehr, and I. Barshack. 2008. B-cell clonal diversification and gut-lymph node trafficking in ulcerative colitis revealed using lineage tree analysis. *Eur. J. Immunol.* 38: 2600–2609.

44. Zuckerman, N. S., K. J. McCann, C. H. Ottensmeier, M. Barak, G. Shahaf, H. Edelman, D. Dunn-Walters, R. S. Abraham, F. K. Stevenson, and R. Mehr. 2010. Ig gene diversification and selection in follicular lymphoma, diffuse large B cell lymphoma and primary central nervous system lymphoma revealed by lineage tree and mutation analyses. *Int. Immunol.* 22: 875–887.

45. Shahaf, G., M. Barak, N. S. Zuckerman, N. Swerdlin, M. Gorfine, and R. Mehr. 2008. Antigen-driven selection in germinal centers as reflected by the shape characteristics of immunoglobulin gene lineage trees: a large-scale simulation study. *J. Theor. Biol.* 255: 210–222.

46. Church, G. M. 2005. The personal genome project. *Mol. Syst. Biol.* 1: 2005.0030.

47. Alamyar, E., V. Giudicelli, S. Li, P. Duroux, and M.-P. Lefranc. 2012. IMGT/HighV-QUEST: the IMGT web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res.* 8: 26.

48. Lefranc, M.-P., C. Pommié, M. Ruiz, V. Giudicelli, E. Foulquier, L. Truong, V. Thouvenin-Contet, and G. Lefranc. 2003. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.* 27: 55–77.

49. Steiman-Shimony, A., H. Edelman, A. Hutzler, M. Barak, N. S. Zuckerman, G. Shahaf, D. Dunn-Walters, D. I. Stott, R. S. Abraham, and R. Mehr. 2006. Lineage tree analysis of immunoglobulin variable-region gene mutations in autoimmune diseases: chronic activation, normal selection. *Cell. Immunol.* 244: 130–136.