

Improved methods for detecting selection by mutation analysis of Ig V region sequences

Uri Hershberg^{1,2}, Mohamed Uduman³, Mark J. Shlomchik^{1,2} and Steven H. Kleinstein^{3,4}

¹Department of Laboratory Medicine, ²Department of Immunobiology, ³Interdepartmental Program in Computational Biology and Bioinformatics and ⁴Department of Pathology, Yale University School of Medicine, Yale University, The Anlyan Center, New Haven, CT 06520, USA

Keywords: adaptive immunity, affinity maturation, B cells, cancer, germinal center

Abstract

Statistical methods based on the relative frequency of replacement mutations in B lymphocyte Ig V region sequences have been widely used to detect the forces of selection that shape the B cell repertoire. However, current methods produce an unexpectedly high frequency of false positives and are sensitive to intrinsic biases of somatic hypermutation that can give the appearance of selection. The new statistical test proposed here provides a better trade-off between sensitivity and specificity compared with previous approaches. The low specificity of existing methods was shown *in silico* to result from an interaction between the effects of positive and negative selection. False detection of positive selection was confirmed *in vivo* through a re-analysis of published sequence data from diffuse large B cell lymphomas, highlighting the need for re-analysis of some existing studies. The sensitivity of the proposed method to detect selection was validated using new Ig transgenic mouse models in which positive selection was expected to be a significant force, as well as with a simulation-based approach. Previous concerns that intrinsic biases of somatic hypermutation could give the appearance of selection were addressed by extending the current mutation models to more fully account for the impact of microsequence on relative mutability and to include transition bias. High specificity was confirmed using a large set of non-productively rearranged Ig sequences. These results show that selection can be detected *in vivo* with high specificity using the new method proposed here, allowing greater insight into the existence and direction of antigen-driven selection.

Introduction

The ability of the immune system to adapt in response to antigenic challenge protects from recurrent infections, helps guard against rapidly mutating pathogens and is the basis for vaccines. In humoral immunity, this adaptation is based on the positive selection of rare higher affinity B cell clones generated through somatic hypermutation, along with negative selection of B cells with lower affinity or non-functional receptors (1, 2). The ability to detect selection, especially positive selection, in experimentally derived Ig sequences is a critical part of many studies. Indeed, Google Scholar identifies nearly 500 citations of four important papers defining methods to detect selection (3–6). Such techniques are useful not only for understanding the immune response to pathogens but are also critical to determine the role of antigen-driven selection in autoimmunity (3, 7), B cell cancers (8, 9) and in the diversification of pre-immune repertoires in certain species (10).

Current methods for detecting selection are based on the analysis of mutation patterns in Ig sequences. The most

common tests compare the frequency of replacement (*R*) mutations (i.e. mutations leading to a change in amino acid) found in mutated B cell Ig sequences to their expected frequency under the null hypothesis of no selection. Elevated frequencies indicate positive selection, while decreased levels indicate negative selection with significance determined by a binomial test (3). Separate tests are performed for *R* mutations that occur in the complementary-determining regions (CDRs), where most contact residues for antigen binding are found, and in the framework (FW) regions, which provide the structural backbone of the receptor. Pooling mutations by functional region increases statistical power, but tests based on this division are limited to detecting one type of selection in CDR or FW (most commonly positive selection in the CDR and negative selection in the FW). The various tests investigated here accept this division into CDR and FW, and mainly differ in how they define the expected frequency of *R* mutations.

The first use of a binomial test to study selection focused on detecting positive selection by comparing the fraction of R mutations in the CDR to all other mutations (3). The expectation that 50% of FW R mutations would be purged through negative selection was built in to the test by doubling the observed number of R mutations in the FW (3). This approach was extended by Chang and Casali (4) to account for the observation that codons with a higher propensity for R mutations are preferentially found in the CDRs. At the same time, this test removed the assumption of a fixed level of negative selection by considering the total number of mutations in the FW without doubling the number of R mutations. Lossos *et al.* (5) claimed that these binomial-based tests failed to account for the fact that selection occurs on the whole Ig sequence, and that a test based on a multinomial distribution, which considers the frequency of mutations in four categories (CDR R , CDR silent (S), FW R and FW S), would be more accurate. We prove here that this claim is incorrect. The multinomial test described by Lossos *et al.* is equivalent to a binomial test and is virtually the same test proposed by Chang and Casali (4). In this paper, we refer to the multinomial test (5) as a 'global binomial test' to more accurately reflect its true nature. Through a simulation-based validation approach, we further demonstrate that an interaction between positive and negative selection is responsible for the lower than expected specificity exhibited by this test (6), leading to a Type I error rate that cannot be defined in practice. A re-analysis of sequences from diffuse large B cell lymphomas (DLBCL) (5, 8) confirms that this lack of specificity has led to erroneous conclusions when applied to experimental data.

A source of serious criticism for all methods based on the frequency of R mutations has been the uncertainty inherent in determining this expected frequency under the null hypothesis of no selection. Dunn-Walters and Spencer (11) showed that the intrinsic biases of somatic hypermutation can give the appearance of selection, while Bose and Sinha (6) demonstrated that false-positive results can occur even when microsequence specificity is incorporated into the null hypothesis. However, as we show here, both of these studies were based on a flawed statistical test. In addition, the model developed by Bose and Sinha (6) only partially accounted for the effect of microsequence specificity and did not integrate the transition bias in somatic mutation.

To address these issues, we propose a new 'focused binomial test' for selection. In addition to correcting the specificity problem of the global binomial test, this method more fully accounts for the effects of microsequence specificity and also introduces the well-characterized transition bias of somatic hypermutation for the first time into the null model of mutation. The performance of this new method is validated on both synthetic and experimental data. In particular, we show that the focused binomial test does not detect selection in non-productive Ig sequences, but can detect positive selection *in vivo* (and *in silico*) when expected.

Methods

Statistical tests for detecting selection

All the tests considered here determine whether the observed number of R mutations (x) in either CDR or FW is significantly

different than expected relative to a larger group of mutations (n). The expected frequency (x/n) under the null hypothesis of no selection (of the type being tested for) is given by p . Mutations are assumed to be independent, and significance is calculated using a binomial probability model. For $x/n \leq p$ (an indication of negative selection), the significance of the test is calculated as the probability of observing x or fewer R mutations by adding half the probability density function (PDF) at x to the cumulative distribution function (CDF) at $(x - 1)$:

$$P = \text{Binomial CDF}(x - 1|n, p) + 0.5 \times \text{Binomial PDF}(x|n, p),$$

where:

$$\text{Binomial PDF}(x|n, p) = \binom{n}{x} p^x (1 - p)^{n-x},$$

$$\text{Binomial CDF}(x|n, p) = \sum_{i=0}^x \text{Binomial PDF}(i|n, p).$$

The standard convention of using half the PDF at x allows us to calculate the P value for $x/n > p$ (an indication of positive selection), as one minus the above. Note that since we perform a two-tailed test (i.e. for both positive and negative selection), these values of P are multiplied by two before comparing with the critical value (α). All the tests described below use the same definition for x , but differ in how they define n and p .

As detailed in the Appendix, we have re-formulated the multinomial test of Lossos *et al.* (5) as an equivalent binomial test (referred to here as the global binomial test). In this case, the observed number of R mutations (x) is compared with all other mutations:

$$n = r_{\text{CDR}} + s_{\text{CDR}} + r_{\text{FW}} + s_{\text{FW}}$$

$$p = \frac{P(R_{\text{region}})}{P(R_{\text{CDR}}) + P(S_{\text{CDR}}) + P(R_{\text{FW}}) + P(S_{\text{FW}})},$$

where region $\in \{\text{CDR}, \text{FW}\}$, and r_{CDR} , s_{CDR} , r_{FW} and s_{FW} are the observed number of R and S mutations in CDR and FW, respectively. $P(R_{\text{region}})$ and $P(S_{\text{region}})$ are the probability of a random mutation being a R or S mutation in the specified region under the null model (Fig. 1). We describe below how to calculate each of these probabilities including the possibility of microsequence specificity and transition bias.

The focused binomial test excludes R mutations from outside the region of interest and normalizes p by the likelihood of all mutations under consideration, thus 'focusing' this test on the forces of selection in only one region:

$$n = r_{\text{region}} + s_{\text{CDR}} + s_{\text{FW}}$$

$$p = \frac{P(R_{\text{region}})}{P(R_{\text{region}}) + P(S_{\text{CDR}}) + P(S_{\text{FW}})},$$

Finally, the local binomial test considers mutations in a single region (CDR or FW):

$$n = r_{\text{region}} + s_{\text{region}} \quad p = \frac{P(R_{\text{region}})}{P(R_{\text{region}}) + P(S_{\text{region}})}.$$

The probability of a random mutation being a R or S mutation in the specified region under the null model [e.g. $P(R_{\text{region}})$

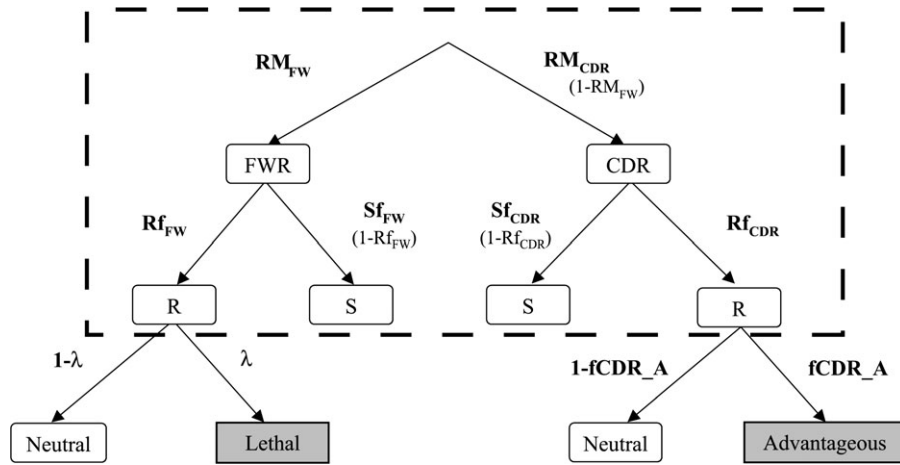


Fig. 1. Mutation decision tree used to determine the effect of each mutation in the simulation model of B cell clonal expansion and the distribution of mutations under the null hypothesis of no selection (area in dashed line). The parameters associated with each branch indicate the probability that an individual mutation will fall into the specified category. For example, in the simulation there is a $RM_{FW} = 75\%$ chance that a mutation will fall into the FW and a $RM_{CDR} = 25\%$ chance that it will fall into the CDR. Similar splits are shown for replacement (R) and silent (S) mutations. Parameters underlying the null hypothesis of the statistical tests investigated here are calculated directly from the germline sequence as described in Methods. In the simulation, mutations are accumulated at a Poisson rate of μ per division. CDR R mutations are affinity increasing with probability $fCDR_A$, while FW R mutations are lethal with probability λ . Default parameter values are provided in Table 1.

and $P(S_{region})$] is given by following the appropriate branches of the boxed area in Fig. 1. RM_{region} is the relative mutational propensity, and Rf_{region} is the expected frequency of R mutations in the specified region. Following the approach of (12) to account for microsequence-specific mutability and transition bias, we have calculated these values as:

$$RM_{FW} = \frac{\sum_{(a \rightarrow b) \in FW} f(a) \times P(b|a)}{\sum_{(a \rightarrow b) \in CDR+FW} f(a) \times P(b|a)} \quad \text{and}$$

$$Rf_{region} = \frac{\sum_{(a \rightarrow b) \in R_{region}} f(a) \times P(b|a)}{\sum_{(a \rightarrow b) \in R_{region} + S_{region}} f(a) \times P(b|a)},$$

where $(a \rightarrow b)$ are individual point mutations that fall in the specified categories. The functions $f(a)$ and $P[b|a]$ account for microsequence specificity leading to the difference in rate of mutation and that leading to transition/transversion bias, respectively. In particular, $f(a)$ is the factor increase/decrease in the mutation rate for the nucleotide a given its immediate surrounding bases (13). $P[b|a]$ is the transition/transversion bias as estimated separately for each nucleotide in (14, 15). Specifically, given that nucleotide a mutates, $P[b|a]$ is the probability that the mutation is to base b as opposed to either of the other two possibilities. Microsequence-specific effects on mutability can be excluded by setting $f(a) = 1$, while transition bias can be excluded by setting $P[b|a] = 1/3$ for all nucleotides. Finally, the required probabilities are obtained from multiplying the probabilities on the branches in the mutation decision tree (Fig. 1):

$$P(R_{FR}) = RM_{FR} \times Rf_{FR},$$

$$P(S_{FR}) = RM_{FR} \times (1 - Rf_{FR}),$$

$$P(R_{CDR}) = (1 - RM_{FR}) \times Rf_{CDR},$$

$$P(S_{CDR}) = (1 - RM_{FR}) \times (1 - Rf_{CDR}).$$

Simulation model of B cell clonal expansion

In order to test the different binomial tests on synthetic data sets with known selection pressures, we developed a stochastic simulation of B cell clonal expansion including proliferation, mutation and death (default parameter values are given in Table 1). The simulation is based directly on the 'Clone' model described (16) as implemented in the study by Kleinstein *et al.* (17). It is initiated with a single seeding cell. During each discrete time step, a number of processes take place (in a random order to prevent bias):

- All cells divide once and accumulate a Poisson distributed number of mutations with average μ . The impact of each individual mutation is stochastic and follows a distribution described by the parameters in Table 1, which are mapped to the decision tree extended from (16) (Fig. 1). Negative selection in our model is defined as the occurrence of lethal mutations, i.e. R mutations that lead to a non-functional or non-specific receptor and ultimately cell death. A fraction (λ) of all FW R mutations will fall into this category. Cells with lethal mutations are removed from the simulation at the end of every time step (or generation).
- Mutation-independent cell death occurs with probability d_i for cell i . Positive selection is implemented as a decrease in this death rate (in other words, as a survival advantage). This is consistent with our recent experimental findings that affinity-based selection is controlled by death and not proliferation (S. Anderson, A. Khalil, Y. Louzoun, S. Kleinstein, U. Herhshberg, A. Haberman and M. Shlomchik). We have adapted the basic framework provided by the B cell dynamics simulation Clone (16) so that the death rate (d_i) for B cell i is given by:

$$d_i = \frac{d_{max}}{S^{a_i}},$$

Table 1. Parameter descriptions and values for the simulation of B cell clonal expansion

Name	Default value ^a
Number of divisions (D)	22
Number of sequences sampled (q)	20
Mutation rate (μ)	0.3 per division
Maximum death rate (d_{\max})	0.4 per division
Selection factor (s)	1 or 7
Relative mutability of FW versus CDR (RM _{FW})	0.75
Probability of CDR R mutation (Rf _{CDR})	0.75
Probability of FW R mutation (Rf _{FW})	0.75
Probability of advantageous mutation (fCDR _A)	0.2
Probability of lethal mutation (λ)	0 or 0.5

^aWhere possible, values are based on published estimates of affinity maturation and experimental results: μ (12) and λ (3). The relative mutability of FW and probability of R mutations is roughly based on that found in murine λ light chains.

where d_{\max} is the maximum death rate (per division), a_i is the number of advantageous mutations carried by cell i and s is the selection factor.

During the simulation, the complete lineage history of all cells is tracked. There are two parameters that control the extent of positive selection, s and d_{\max} . Increasing the selection factors causes each mutation to have a greater proportional effect on the death rate. Note, however, that the first mutation will always have the greatest impact and, if s is too large, then subsequent mutations will have no meaningful effect. We have confirmed previous studies, which have shown that moderate values (e.g. $s \approx 7$) produce the greatest affinity maturation, as measured by the average number of affinity-increasing mutations per cell (16) (data not shown). The death rate of germline affinity cells (d_{\max}) determines the maximum possible survival advantage for higher affinity cells and thus controls the potential for positive selection. Higher values of d_{\max} lead to greater affinity maturation.

For each combination of negative selection ($\lambda = 0$ or 0.5, with $\lambda = 0$ being no negative selection) and positive selection ($s = 1$ or 7, with $s = 1$ being no positive selection), 25 000 clones were simulated for up to 22 generations. At various points in this expansion, 20 cells were randomly sampled and a genetic lineage tree was created as previously described (17). All results presented here are from clones at generation 22.

Sensitivity, specificity and the likelihood ratio

The performance of each test was assessed by estimating its sensitivity (or true positive rate), which is the fraction of true positives over the sum of true positives and false negatives, and its specificity (true negative rate), which equals the fraction of true negatives over the sum of true negatives and false positives. To express the trade-off between sensitivity and specificity in a single measure, we also calculated the likelihood ratio (LR) defined as sensitivity/(1 – specificity). This is a measure commonly used in the medical community. Unless otherwise stated, all measurements were made for binomial tests with a P value cutoff (α) of 0.05.

Counting mutations in clonal Ig sequences

When analyzing clonal Ig sequences, we counted all non-redundant mutations in a clone. This preserves the assumption of independence between mutation events that is implicit in the single-sequence methods under the null hypothesis of no selection.

Experimental data sets

We analyzed three different sets of experimental data. First we re-analyzed the DLBCL sequences from Lossos *et al.* (5, 8), verifying the previously reported results from the global binomial test and comparing these with the new local and focused binomial tests. To allow a proper comparison with the previous analyses of DLBCL (5, 8), we calculated all parameters of the decision tree (Fig. 2) in accordance with the prior publications. CDR and FW were divided according to Kabat nomenclature (18) and the potential for R mutation was calculated without microsequence specificity or a transition bias.

The second data set consists of 76 non-productively rearranged Ig V gene sequences from previously published studies (11, 19–22). Somatic Diversification Analysis was used to verify these sequences as non-productive (due to stop codons or mutations at invariant codons) and to identify the germline rearrangement (23). Each sequence was individually aligned with the appropriate germline gene (in ImMunoGeneTics (IMGT) format) using the Needleman–Wunsch pair-wise global alignment algorithm, implemented with the 'org.biojava.bio.alignment' package from BioJava 1.5 (24) and then manually corrected. Our analysis excluded 30 unmutated sequences and 7 sequences with insertions or deletions, resulting in a total of 39 mutated non-productive sequences.

The final data set is derived from the response to the nitrophenyl (NP) antigen in conventional Ig heavy-chain transgenic mice, which do not mutate or isotype switch their heavy-chain loci (25). B1-8 mice use a canonical V_H186.2 heavy chain which when paired with endogenous λ 1 light chains encode a B cell receptor with moderate affinity ($K_a \sim 9.64 \times 10^5 \text{ M}^{-1}$) (26, 27). In contrast, V23 mice use a non-canonical heavy chain that when paired with endogenous λ 1 encodes a B cell receptor with very low affinity ($K_a < 5.0 \times 10^4 \text{ M}^{-1}$) (26). These mice were backcrossed with Jh KO/Balb mice (27, 28) to ensure that all B cell receptors use the heavy-chain transgene exclusively. All mice were maintained under specific pathogen free conditions and used at 6–10 weeks of age. Mice were immunized intraperitoneally with 50 μg of NP₂₅-chicken gamma globulin precipitated in alum.

Freezing, sectioning and staining were performed essentially as described in (27). λ^+ GCs were microdissected from stained spleen sections at days 10 and 16 after immunization essentially as described in (27). V _{λ} 1 sequences were amplified by nested PCR using *Pfu* Turbo polymerase (Stratagene) using external primers 5'-GCACCTCAAGTCTTGAGAG-3' and 5'-ACTCTCTCTCTGGCTCTCA-3' and internal primers 5'-CTACACTGCAGTGGGTATGCAA CAATGCG-3' and 5'-GTTCTCTAGACCTAGGACAGTCAGT-TTGG-3'. Amplified DNA was cloned directly into pCR4

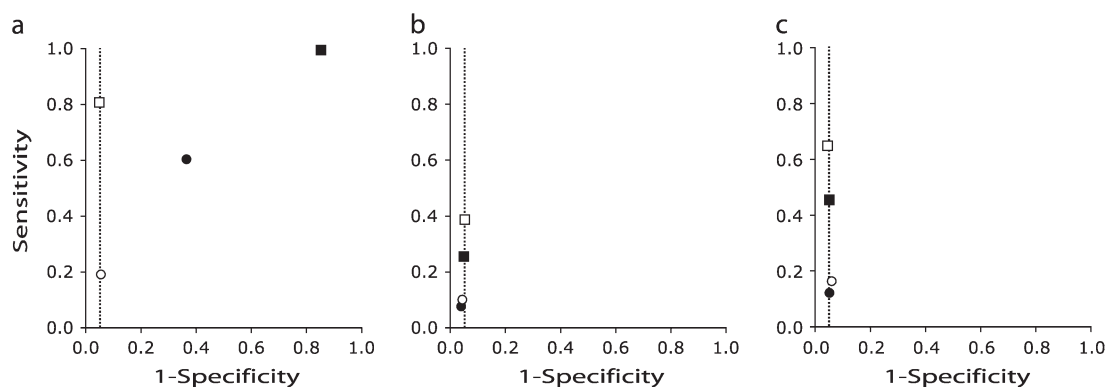


Fig. 2. Simulation-based validation showing the sensitivity and specificity for the (a) global, (b) local and (c) focused binomial tests for detecting selection. Each statistical test was applied to synthetic data subject to both positive and negative selection (solid markers) or positive selection only (open markers). Tests were performed using mutations from single clones (circles) or by combining mutations from groups of 10 clones (squares). The dashed vertical lines indicate the expected specificity of 0.95 (for $\alpha = 0.05$).

Blunt-TOPO vector using the Zero Blunt TOPO PCR Cloning Kit for Sequencing (Invitrogen). $V_{\lambda}1$ DNA was further amplified by placing colonies directly into PCR reactions containing the following primers: M13 forward 5'-GTAAAACGACGGC-CAG-3' and M13 reverse 5'-CAGGAAACAGCTATGAC-3'. DNA was purified from the PCR reaction mixture with the QIAquick PCR Purification kit (Qiagen), mixed with sequencing primer, T3 5'-AATTAACCCTCACTAAAGGG-3', and sequenced by the Keck Biotechnology Resource Laboratory at Yale University School of Medicine using Applied Biosystems DNA sequencers. We typically recovered eight sequences per microdissection. Sequences were aligned to a rearranged germline $V_{\lambda}1/J_{\lambda}1$ sequence using Lasergene DNA analysis software.

To determine if sequences were clonally related, we developed a set of computer algorithms to handle the specific circumstances of Ig hypermutation analysis (U. Hershberg, T. Gianoulis, L. Tsaban, Y. Louzoun, S. Kleinstein, M. Shlomchik). By combinatorial matching of the end regions of $V_{\lambda}1$ and $J_{\lambda}1$, the algorithm determined all possible V-J junctions, including those potentially generated by P-nucleotides. This allowed us to differentiate between junctional diversity and somatic hypermutation in the region of the V-J junction, as those bases that could not be accounted for by any combination of the germline sequences were considered to be mutations. Junctional diversity was then used to separate independent clones that may have been found in the same microdissection. Since there is relatively little diversity at V-J junctions and virtually no N-region addition, there are multiple independent examples of the same junction among $V_{\lambda}1$ sequences. Therefore, sequences that shared one of a few very common junctions were considered independent unless they also shared at least one mutation, in which case they were considered clonally related. The computer algorithm was also used to identify cases of independent parallel mutations, which in most cases were attributed to hybridization in the PCR amplification process and were thus discarded. Since isolated independent parallel mutations are known to occur, particularly in hot spots, commonly observed single mutations seen in parallel were not discarded unless other evidence indicating PCR hybridization was found.

Sequences were divided into CDR and FW regions according to our published methodology (29) based on Kabat and IMGT (18, 29, 30). Details concerning the number of mutations in the CDR and FW of all sequences analyzed and the IMGT names of germline sequences used to calculate the expected frequency of each mutation type can be found in Supplementary Table 1 (available at *International Immunology* Online).

Results

The new statistical tests proposed here were created to correct flaws in current methods. In order to understand the rationale for the proposed changes, it is important to first understand the source of these flaws.

The specificity of the multinomial test is decreased by crosstalk

The P value cutoff used to determine statistical significance in the multinomial test of Lossos *et al.* (5) should predict the Type I error rate ($1 - \text{specificity}$). However, the specificity of the multinomial test on simulated data was lower than expected (6). We have found that this troubling finding, which should provide reason enough to reject the multinomial test, is due to an interaction between the effects of positive and negative selection. This could be seen most easily when the multinomial test was re-formulated as a mathematically equivalent (and computationally efficient) binomial test that compares R mutations in the region of interest (i.e. CDR or FW) to the sum total of all mutations (see Methods and proof of equivalence in Appendix). Consequently, in tests for positive selection, the multinomial test could misinterpret a decreased frequency of FW R mutations, the result of negative selection, as a relative overabundance of CDR R mutations. We refer to this potential for interactions of positive and negative selection to create a false signal in the multinomial test as 'crosstalk', as each type of selection interferes with the ability to correctly detect the other.

We used a simulation-based validation approach to verify that crosstalk completely explains the reduced specificity of the multinomial test. The use of a simulation allowed us to

create synthetic data sets with known levels of positive and negative selection so that specificity could be directly estimated. Known levels of positive and negative selection were included by adapting the basic framework of the B cell dynamics simulation Clone (16), so that a specified fraction of CDR R mutations are advantageous, while a fraction of FW R mutations are lethal (see Methods and Fig. 1, with default parameters given in Table 1). For reasons of computational efficiency, the binomial equivalent of the Lossos *et al.* multinomial test (referred to throughout this paper as the global binomial test) was used in all the analyses described in the rest of the paper. When positive and negative selection both occur, the specificity of the global binomial test for detecting positive selection was ~ 0.60 , much lower than the expected value of 0.95 (given by one minus the P value cutoff used to determine statistical significance). However, in the absence of crosstalk, which was accomplished in the simulation by including only positive selection (i.e. $\lambda = 0$), the global binomial tests had the expected specificity (Fig. 2a and Table 2). Thus, crosstalk completely explains the low specificity of the global binomial test.

Relationship of the multinomial test to the method of Chang and Casali

The re-formulation of the multinomial test described by Lossos *et al.* (5) as a binomial test (Appendix) highlights the close relationship between this method and the one proposed by Chang and Casali (4). In fact, there is only a slight mathematical difference in how the two methods calculate the P value. The P value of Lossos *et al.* correctly includes the possibility of producing either the observed number of R mutations or more extreme values, while Chang and Casali calculated only the probability of producing exactly the observed number of R mutations (i.e. Binomial PDF with the parameters of a global binomial test, as described in Methods). Two important insights can be gained from this perspective. First, the method of Lossos *et al.* will almost always produce P values that are more conservative than the Chang and Casali approach since it sums over a larger number of possible values for the number of R mutations when calculating the P value. The only exception to this occurs when there is a large probability of producing exactly the observed number of R mutations, since the method of Lossos *et al.* only includes half this probability in the P value. The second insight is that the two methods are actually testing different hypotheses. By focusing on the exact number of observed R mutations, the method of Chang and Casali implicitly tests both for an excess or scarcity of R mutations. In contrast, the direction of selection must be specified in the Lossos *et al.* approach since it sums over more extreme numbers of mutations in calculating the P value. This reasoning provides a simple, mathematical explanation for all six of the discrepancies noted by Lossos *et al.* when comparing their approach to the method of Chang and Casali on a set of 54 Ig sequences from DLBCL patients (5, 8).

Improved methods for detecting selection: the local and focused binomial tests

Crosstalk between positive and negative selection can be avoided by using a binomial test that considers the muta-

tions in CDR and FW separately. Such a 'local binomial test' has been used by us previously (29, 31), though never explicitly formulated or validated as we have done here. As shown in Fig. 2(b), this test had the expected specificity of ~ 0.95 (for $\alpha = 0.05$) and a sensitivity of ~ 0.10 for detecting positive selection on synthetic data. To compare this performance with the global binomial test, we quantified the trade-off between sensitivity and specificity using the $LR = \text{sensitivity} / (1 - \text{specificity})$, where higher values indicate better performance. Despite the higher sensitivity of the global binomial test (compare Fig. 2a and b), the local binomial test exhibited slightly better overall performance as measured by its LR (1.95 versus 1.65, see Fig. 3a and Table 2). The improved performance of the local binomial test was more dramatic when comparing the ability to detect negative selection (LR of 12.3 versus 7.23, Table 2), which is a much stronger force in the simulation. This advantage disappears when only negative selection is included in the synthetic data (Table 2), highlighting once again the detrimental influence of crosstalk on the performance of the global binomial test.

The relatively high sensitivity of the global binomial test is due to the inclusion of more mutations in the analysis (specifically R mutations from both the CDR and FW), compared with the local binomial test, which only considers one region. However, the results of the previous section clearly demonstrate that including R mutations subject to selective forces that are not the focus of the test decreases specificity. Thus, we sought to determine whether extending the local binomial test by including all S mutations, while still focusing on R mutations from only one region (CDR or FW), could improve the sensitivity for detecting positive selection without sacrificing specificity. As shown in Table 2, this focused binomial test provided the best trade-off between sensitivity and specificity with a LR of 2.37 for detecting positive selection. Furthermore, the sensitivity of this test is comparable with the global binomial test, even for the idealized case of positive selection acting alone (Table 2 and Fig. 2c versus a). Most importantly, the focused binomial test does not suffer from crosstalk as does the global binomial test, so that the P value cutoff (α) relates to specificity in the normal way.

Although the results presented throughout this paper are for single levels of positive and negative selection, we have performed equivalent analyses using a wide range of parameter values (e.g. $s = 1, 2, \dots, 10$ and $\lambda = 0.0, 0.25, 0.50, 0.75, 1.00$) (data not shown). While the exact performance of each test is obviously dependent on the parameters governing selection (e.g. the sensitivity of all tests improves as the number of divisions in the clone increases and more mutations accumulate), the relative performance of the test and general conclusions do not depend on the particular parameter values used.

Re-analysis of DLBCL sequences finds no evidence of positive selection

To determine whether crosstalk in the global binomial test had produced misleading results in practice, we re-analyzed the sequences from DLBCL patients presented in the study by Lossos *et al.* (8). As shown in Fig. 4a and Supplementary Table 1 (available at *International Immunology Online*), the

global binomial test detected many instances of positive selection in these data (16/54 sequences). Most of these predictions were not corroborated by the focused binomial test, which detected positive selection in only two cases. Although these results could reflect differences in sensitivity, they could also reflect a high frequency of false-positive cases in the global binomial test due to crosstalk. The simulation-based validation suggested that positive selection in the CDR detected by the global binomial test could be the result of crosstalk if it occurred with negative selection in the FW. Indeed, significant negative selection was concurrently detected in 88% (14/16) of these DLBCL sequences. In contrast, all the statistical tests find a similar profile of negative selection, and there are many cases in which negative selection is detected in the absence of positive selection. (Fig. 4a and Supplementary Table 1 is available at *International Immunology* Online). We conclude that virtually all the positive selection detected by the global binomial test in these data was the result of crosstalk.

The two cases of positive selection found by the focused binomial test in the DLBCL sequences may simply reflect

Table 2. LR for detecting positive selection in the CDR and negative selection in the FW at generation 22 in simulated clones subject to positive and/or negative selection

Region	No. ^a	Sel. ^b	Global	Local	Focused
CDR	Single	+/-	1.65	1.95	2.37
	Single	+	3.55	2.31	2.73
	Comb.	+/-	1.17	5.19	8.78
	Comb.	+	16.17	7.37	14.32
FW	Single	+/-	7.23	12.3	13.41
	Single	-	13.18	11.21	11.88
	Comb.	+/-	2.63	22.92	22.62
	Comb.	-	20.39	19.69	20.65

^aNumber of clones being tested: mutations were combined from one clone (single) or a group of 10 clones (comb.). ^bThe type of selection (Sel.) pressure present in the simulation: positive and negative (+/-), only negative (-) and only positive (+).

the intrinsic biases of somatic hypermutation, which can produce the appearance of selection (11). We initially did not include these effects in order to compare the results using the focused test directly with those of Lossos *et al.* (5), whose version of the global binomial test did not include such considerations. Bose and Sinha (6) showed how microsequence effects could be incorporated into the estimate of relative mutability for the CDR and FW (i.e. RM_{CDR} and RM_{FW} in Fig. 1). However, their approach ignored the effect of microsequence on the relative probability of *R* and *S* mutations (i.e. Rf_{CDR} and Rf_{FW} in Fig. 1). Using methods we previously developed for modeling B cell clonal expansion (12), we extended the null model of somatic hypermutation to include the full impact of microsequence-specific mutability on the distribution of mutations (see Methods). We also included the impact of transition bias (32), which had not been included in any previous method for detecting selection. After accounting for these intrinsic biases, no significant positive selection was detected in any DLBCL sequences. Many sequences continued to exhibit negative selection in the FW, as well as the CDR (Fig. 4b). Thus, in contrast to what was previously reported (8), we conclude that the heterogeneity of mutations that are present in these DLBCL sequences should not be interpreted as heterogeneity of selective pressures, and that there is no evidence for positive selection during the clonal evolution of these lymphomas.

Positive selection can be detected *in vivo* with high specificity

To verify that the local and focused binomial tests were capable of detecting positive selection *in vivo*, we used two Ig heavy-chain transgenic mouse models, which do not mutate or isotype switch their heavy-chain loci (25). These mice provided an ideal system to study antigen-driven selection. While affinity maturation in the primary anti-NP is normally dominated by an amino acid exchange from tryptophan to leucine at codon 33 of the canonical heavy chain (33), this was not possible in these transgenic mice since mutations occur only in the λ light chain. The result is a highly skewed mutational distribution with all selection pressure on V_{λ} . In

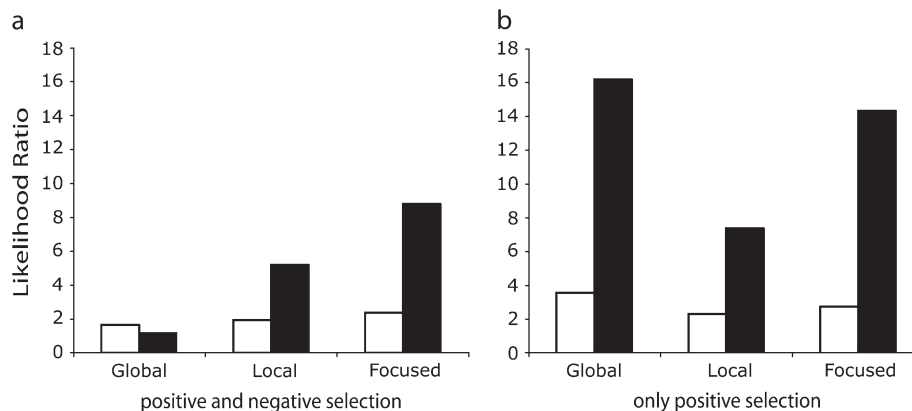


Fig. 3. LR summarizing the trade-off between sensitivity and specificity for the same synthetic data analyzed in Fig. 2. Open bars indicate results from the analysis of mutations from single clones, while solid bars indicate results from mutations combined from groups of 10 clones. (a) LR for detecting positive selection in the presence of both positive and negative selection. (b) LR for detecting positive selection in the presence of positive selection only (i.e. no negative selection).

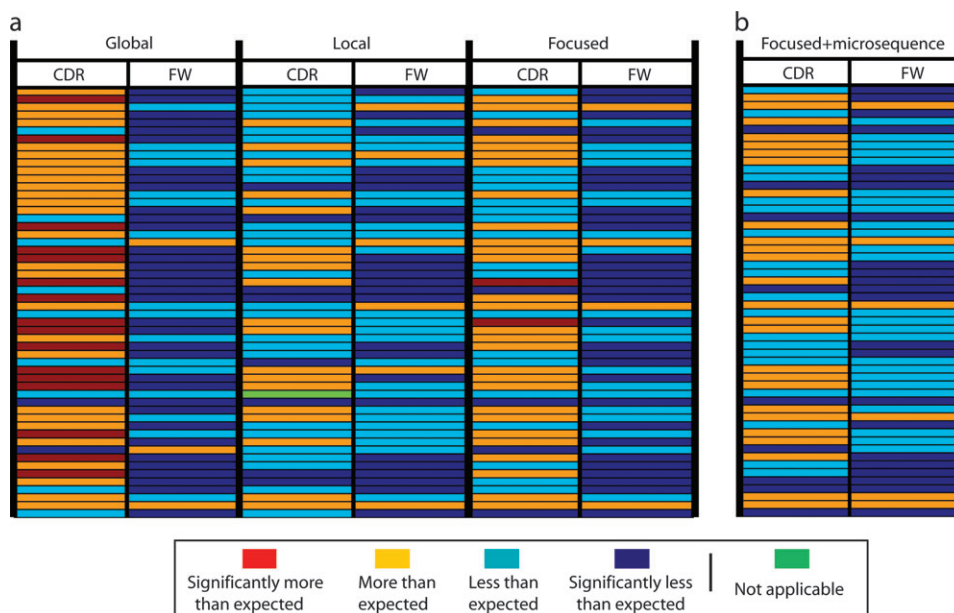


Fig. 4. Mutational analysis of DLBCL sequences from Lossos *et al.* (8). Each row represents a different sequence (order given in (8) and Supplementary Table 1 is available at *International Immunology Online*) and each column a statistical test (global, local and focused binomial) for detecting an excess/scarcity of *R* mutations in the CDR or FW region. *P* values are calculated (a) without microsequence specificity and transition bias or (b) including these intrinsic biases of somatic hypermutation. The results of each test are color coded as described in the legend. Note that each sequence and region was tested for both positive and negative selection (i.e. a two-tailed test with $\alpha = 0.10$), leading to the detection of two sequences exhibiting negative selection in the CDR that were not identified in the original one-tailed analysis of Lossos *et al.* (5). Results from the local and focused tests are indicated as not applicable when no mutations relevant to the test are found in the sequence. For instance if testing for an abundance of mutations in the CDR the local binomial test is not applicable if there are only mutations (*R* and/or *S*) in the FW and the focused mutation is not applicable if there are only *R* mutations in the FW. Exact *P* values are in Supplementary Table 1 (available at *International Immunology Online*).

addition, one of the V_H Ig we used fixes a very low initial affinity, which was improved by the selection of one or more recurrent CDR *R* mutations, as shown by site-directed mutation and re-expression experiments (A. Khalil, S. Anderson and M. Shlomchik, unpublished observations). We could thus, a priori, be confident that positive selection was present. This provided a unique opportunity to evaluate the relative sensitivity of each statistical test for detecting selection *in vivo*.

In order to maximize our chances of detecting positive selection, mutations were combined from sets of clonally related sequences taken from adjacent microdissected germinal center cells. B cells from a single microdissection were considered to be part of the same clone if they shared an equivalent V-J gene junction and if they shared at least one mutation. Despite the use of an idealized experimental system and summing of mutations over clones, no significant positive or negative selection was detected in any of the 79 clones using the local or focused binomial tests. This probably reflects the low sensitivity of these methods considering the small number of mutations present in these sequences (an average of 3.5 independent mutations per clone, compared with 29.7 per sequence in the DLBCL data of the previous section). As our simulation-based validation demonstrated, although the focused test offered an improved LR, the performance was still far from ideal. This is partly due to the weak selection pressure in the simulation model, but the importance of mutation frequency as a source of false negatives in the synthetic data is underscored by the significant negative correlation between the number of *R*

mutations in the CDR and the *P* value for detecting positive selection ($r = -0.58$, $P \ll 0.05$).

To increase the sensitivity for detecting selection, we combined mutations from several clones that arose under equivalent experimental conditions (i.e. four groups reflecting B1-8 or V23 mice at days 10 or 16). In our simulation-based validation, this kind of grouping raised the sensitivity of both the local and focused binomial tests without changing specificity, leading to a 3-fold improvement in the LR when mutation data from 10 clones was combined (Fig. 3 and Table 2). Indeed, when clones were grouped by experiment, significant positive selection was detected in the CDR for three of the four groups. [In contrast to the local and focused binomial tests, combining mutations from different clones actually degraded the performance of the global binomial test (Fig. 3a and Table 2). In this case, the inclusion of more mutations not only amplified the signal for both positive and negative selection but also increased the signal from the crosstalk between the two types of selection. This resulted in a lower LR, as the decrease in specificity surpassed the increase in sensitivity. It is clear that the lower LR of combined mutations is the result of crosstalk, as it is not found when only positive selection is present (Fig. 3a versus b and Table 2).]

Previous studies have questioned the specificity of statistical methods for detecting selection *in vivo* based on the findings of positive selection in non-productively rearranged Ig sequences (11) and sequences derived from T-independent responses (6), where selection was not

expected to influence the pattern of mutation. However, these studies used the method of Chang and Casali and the global binomial test and thus were subject to a high rate of false-positive results. To estimate the specificity of the focused binomial test *in vivo*, we collected a large set of non-productively rearranged Ig sequences [including those from (11)]. Positive selection was detected in <8% (3/39) of these sequences (with an average of 8.31 mutations per sequence), consistent with the expected Type I error rate. Indeed, after correcting for multiple hypothesis testing, no selection was detected in these sequences at a false discovery rate of 5% (34) (with $\lambda = 0$). In addition, positive selection was no longer detected in two sequences derived from a T-independent response (6). These results are consistent with the simulation-based validation, and suggest an acceptable specificity for the focused binomial test *in vivo*.

Discussion

The ability to detect antigen-driven selection in B cell Ig receptors is critical to understanding B cell development in health and disease. While direct measurements of receptor affinity can be made for specific antigens, they are impractical for large numbers of individual receptors reacting to complex antigens and impossible for cases where the antigen is unknown (such as B cell cancers). The most widely applied methods for detecting selection analyze the frequency of *R* mutations in somatically mutated Ig V sequences. In these approaches, an excess of *R* mutations relative to that expected by a 'random' mutation process indicates positive selection, while a scarcity of *R* mutations indicates negative selection. Previous studies have argued against the use of statistical tests based on this frequency because of a low specificity in practice due to the difficulty of accurately defining the features of a random mutation process (6, 11). However, we found that these criticisms were based on the use of incorrect statistical tests and incomplete models of mutation. When these problems are corrected using the methods proposed in this study, antigen-driven selection can be detected *in vivo* with high specificity.

The current state-of-the-art method for detecting selection is a multinomial test proposed by Lossos *et al.* (5). This test has been widely used including, in the last year alone, several comparative studies of different lymphomas (35, 36), the germinal center reaction to simian HIV infection (37), the study of the role of mutation in the primordial immune systems of teleost fish (38) and the clinical study of a patient with chronic Lyme arthritis (39), among many others. We have shown here that the multinomial test is mathematically equivalent to the binomial test proposed earlier by Chang and Casali (4). The sole difference between the two methods is that the multinomial test corrects a serious statistical flaw in the Chang and Casali approach, whereby the *P* value was calculated as the probability of producing exactly the observed number of *R* mutations instead of including the possibility of producing more extreme values in the rejection region (i.e. 'more than' the observed number of *R* mutations when testing for positive selection and 'less than' the observed number when testing for negative selection). To more accurately reflect this derivation, we have referred to the

'multinomial test' as a global binomial test throughout this paper. Despite its widespread application, we recommend that this test be avoided because it has a Type I error rate ($1 - \text{specificity}$) that cannot be defined in practice due to crosstalk between positive and negative selection. For instance, the global binomial test cannot distinguish whether a relatively high frequency of *R* mutations in CDR is due to positive selection in the CDR or negative selection in the FW. The impact of crosstalk *in vivo* is apparent by comparing the CDR column patterns in Fig. 4. Whereas the frequency of CDR *R* mutations is more than expected for most sequences using the global binomial test (Fig. 4a, global), there is a trend toward less than the expected frequency as one removes the influence of crosstalk (Fig. 4a, local) and accounts for the intrinsic biases of somatic hypermutation (Fig. 4b). By not accounting for these effects, the global binomial test produces a high number of false-positive results *in vivo* leading to erroneous conclusions.

Two new methods that can eliminate crosstalk were proposed and evaluated here. The local binomial test looks separately at the CDR or the FW to detect selection, thus avoiding the potentially confounding signal of *R* mutations from outside the region of interest. However, the failure to make use of all available information (specifically, the mutations occurring in the other region) led to a decreased sensitivity compared with other methods. To maximize sensitivity, while maintaining an easily calculable and adjustable specificity, we proposed the focused binomial test, which combines *R* mutations in the region of interest with the total number of *S* mutations. To address previous concerns about low specificity *in vivo* (6, 11), our model also accounts for known biases of somatic hypermutation at two levels. First, the relative likelihood for each individual nucleotide to mutate depends on its local sequence context. This microsequence-specific mutability impacts both the distribution of mutations in CDR and FW and the relative frequency of *R* and *S* mutations. Previous studies accounted only for the former, producing results that were sometimes difficult to interpret (6). Second, nucleotides are not equally likely to be generated from the mutation of a particular position, with the consequence that transitions are about twice as likely as transversions (32). This transition bias, which has not been included in any previous method for detecting selection, can have a large impact on the relative frequency of *R* and *S* mutations (29). This improved model of mutation, combined with the focused binomial test, produces a high specificity in practice. In particular, positive selection was detected only at the level of chance in a large set of non-productively rearranged Ig sequences, and was no longer detected in two sequences derived from a T-independent response that were falsely detected by previous methods (6). Thus, our methods address previous concerns that intrinsic biases of somatic hypermutation would be misinterpreted leading to a low specificity in practice (6, 11).

A simulation-based validation using synthetic data sets with known levels of positive and negative selection found that the focused binomial test exhibited the best trade-off between sensitivity and specificity under biologically realistic conditions (Fig. 3a and Table 2). *In vivo* validation is problematic since the actual extent of selection may differ

significantly between clones, even for the same response, and there is no way to independently verify the results on a large scale. We overcame this problem using novel transgenic mouse models where positive selection was expected to dominate due to a fixed heavy chain, causing all mutation and selection to be focused on the λ light chain. Nevertheless, when clones from these mice were analyzed individually, few showed evidence for significant positive (or negative) selection. This low sensitivity results from the relatively small number of mutations, which simply reflects the biology of early time points in the germinal center, combined with the high frequency of R mutations expected even in the absence of selection. To address this natural limitation, we propose combining independent mutations from groups of sequences that share the same experimental parameters. When mutations were combined by mouse strain and time point, significant positive selection in the CDR was found in three of the four experimental groups demonstrating that positive selection could be detected *in vivo* using the focused binomial test. Analysis of synthetic data confirmed that this was not due to decreased specificity. Furthermore, grouping mutations based on the germline rearrangement increased detection of negative selection but did not lead to the detection of positive selection in a set of DLBCL sequences (results not shown).

As stated in the Introduction, the statistical tests investigated here all depend on a reasonable division of the Ig receptor into CDR and FW regions. This split helps to isolate the forces of positive and negative selection and allows for pooling of mutations to increase sensitivity. The location of each region was determined using standard conventions such as IMGT (30) and Kabat (18) (as specified) throughout this paper. The various conventions produce minor differences in the calculated parameters of the mutation model (Fig. 2), but even small changes can alter conclusions regarding significance for sequences/clones that are close to the P value cutoff. More work is required to determine the optimal division to detect selection, and we hope that future methods can be developed that do not depend on a pre-specified division.

Another area needing refinement concerns the null model for the distribution of random (i.e. non-selected) mutations (Fig. 1). In addition to estimating the relative probabilities using larger data sets of non-selected mutations, improvements in the mutation model may include expanding the limited sequence context for intrinsic bias (currently the two adjacent bases on either side for microsequence specificity and the nucleotide itself for transition bias), accounting for the distance to the promoter (40), as well as potential strand bias (15, 41). Updating the null model should be relatively straightforward in the framework we propose as the assumptions underlying the model of intrinsic biases for somatic hypermutation are isolated into two functions: $f(a)$ and $P[b|a]$ for microsequence specificity of mutation and transition bias, respectively (see Methods).

We conclude that statistical tests based on the frequency of R mutations can be used to accurately detect selection in Ig V region sequences. In particular, the focused binomial test provides high specificity and increased sensitivity, and can be used to detect positive and negative selection in single sequences, independent mutations from sets of clonally

related Ig sequences or mutations combined from groups of related sequences. In contrast, the global binomial test of Lossos *et al.* (and the closely related method of Chang and Casali) has a decreased specificity that cannot be defined in practice. We therefore strongly suggest that studies based heavily on global binomial tests should be reinterpreted using the focused binomial test proposed here, especially cases where evidence of positive selection was found in sequences also showing strong negative selection pressure (or the converse). To make the calculations underlying the focused binomial test easily accessible to researchers without programming expertise, we have created a web version that can be accessed at: <http://clip.med.yale.edu/selection>. Given a set of sequences (which may be clonally related) aligned with a germline sequence, this program will calculate the parameters underlying the null model of mutation (Fig. 1) and test for positive and negative selection in the CDR and FW.

Supplementary data

Supplementary Table 1 is available at *International Immunology* Online.

Funding

Informatics fellowship of the Pharmaceutical Research and Manufacturers of America foundation to U.H.; National Institutes of Health (A143603) to M.J.S.; National Science Foundation Integrative Graduate Education and Research Training (DGE-9972930) to S.H.K.

Appendix

Here, we show that the multinomial test proposed by Lossos *et al.* (5) is mathematically equivalent to a much simpler binomial test. The P value for the multinomial test for positive selection in the CDR is given by:

$$\sum_{\substack{r_{\text{CDR}} \leq k \leq n \\ r_{\text{FW}} + s_{\text{FW}} + k + s_{\text{CDR}} = n}} \binom{n}{r_{\text{FW}}, s_{\text{FW}}, k, s_{\text{CDR}}} \times P(R_{\text{FW}})^{r_{\text{FW}}} P(S_{\text{FW}})^{s_{\text{FW}}} P(R_{\text{CDR}})^k P(S_{\text{CDR}})^{s_{\text{CDR}}}.$$

Separating the summations, this can be written as:

$$\sum_{r_{\text{CDR}} \leq k \leq n} \left(\frac{n!}{k!} P(R_{\text{CDR}})^k \right) \times \sum_{r_{\text{FW}} + s_{\text{FW}} + s_{\text{CDR}} = (n-k)} \left(\frac{1}{r_{\text{FW}}! s_{\text{FW}}! s_{\text{CDR}}!} \right) \times P(R_{\text{FW}})^{r_{\text{FW}}} P(S_{\text{FW}})^{s_{\text{FW}}} P(S_{\text{CDR}})^{s_{\text{CDR}}}.$$

Multiplying the second summation by $(n-k)!/(n-k)!$ gives:

$$\sum_{r_{\text{CDR}} \leq k \leq n} \left(\frac{n!}{k!(n-k)!} P(R_{\text{CDR}})^k \right) \times \sum_{r_{\text{FW}} + s_{\text{FW}} + s_{\text{CDR}} = (n-k)} \left(\frac{(n-k)!}{r_{\text{FW}}! s_{\text{FW}}! s_{\text{CDR}}!} \right) \times P(R_{\text{FW}})^{r_{\text{FW}}} P(S_{\text{FW}})^{s_{\text{FW}}} P(S_{\text{CDR}})^{s_{\text{CDR}}}.$$

By the multinomial theorem, this is equivalent to:

$$\sum_{r_{\text{CDR}} \leq k \leq n} \left(\frac{n!}{k!(n-k)!} P(R_{\text{CDR}})^k \right) \times (P(R_{\text{FW}}) + P(S_{\text{FW}}) + P(S_{\text{CDR}}))^{(n-k)}.$$

Since $P(R_{\text{FR}}) + P(S_{\text{FR}}) + P(S_{\text{CDR}}) = 1 - P(R_{\text{CDR}})$, this can be rewritten as:

$$\sum_{r_{\text{CDR}} \leq k \leq n} \binom{n}{k} P(R_{\text{CDR}})^k (1 - P(R_{\text{CDR}}))^{(n-k)},$$

which is a binomial test with $p = P(R_{\text{CDR}})$. In practice the P value is conventionally calculated as $\Pr(R_{\text{CDR}} > r_{\text{CDR}}) + 0.5 \times \Pr(R_{\text{CDR}} = r_{\text{CDR}})$, but this does not change the logic of the above proof.

Acknowledgements

We would like to thank Joseph Chang and Abraham Tzou for helpful statistical advice, Ashraf Khalil and Shannon Anderson for helpful discussions and access to their data and David Schatz for his support and insight.

Abbreviations

CDF	cumulative distribution function
CDR	complementary-determining region
DLBCL	diffuse large B cell lymphomas
FW	framework (region)
LR	likelihood ratio
NP	nitrophenyl
PDF	probability density function
R	replacement (mutation)
S	silent (mutation)

References

- McKean, D., Huppi, K., Bell, M., Staudt, L., Gerhard, W. and Weigert, M. 1984. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc. Natl Acad. Sci. USA* 81:3180.
- Siskind, G. W. and Benacerraf, B. 1969. Cell selection by antigen in the immune response. *Adv. Immunol.* 10:1.
- Shlomchik, M. J., Aucoin, A. H., Pietsky, D. S. and Weigert, M. G. 1987. Structure and function of anti-DNA autoantibodies derived from a single autoimmune mouse. *Proc. Natl Acad. Sci. USA* 84:9150.
- Chang, B. and Casali, P. 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.
- Lossos, I. S., Tibshirani, R., Narasimhan, B. and Levy, R. 2000. The inference of antigen selection on Ig genes. *J. Immunol.* 165:5122.
- Bose, B. and Sinha, S. 2005. Problems in using statistical analysis of replacement and silent mutations in antibody genes for determining antigen-driven affinity selection. *Immunology* 116:172.
- William, J., Euler, C., Christensen, S. and Shlomchik, M. J. 2002. Evolution of autoantibody responses via somatic hypermutation outside of germinal centers. *Science* 297:2066.
- Lossos, I. S., Okada, C. Y., Tibshirani, R. et al. 2000. Molecular analysis of immunoglobulin genes in diffuse large B-cell lymphomas. *Blood* 95:1797.
- Bagnara, D., Callea, V., Stelitano, C. et al. 2006. IgV gene intraclonal diversification and clonal evolution in B-cell chronic lymphocytic leukaemia. *Br. J. Haematol.* 133:50.
- Flajnik, M. F. 2002. Comparative analyses of Ig genes: surprises and portents. *Nat. Rev. Immunol.* 2:688.
- Dunn-Walters, D. K. and Spencer, J. 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.
- Kleinstei, S. H. and Singh, J. P. 2003. Why are there so few key mutant clones? The influence of stochastic selection and blocking on affinity maturation in the germinal center. *Int. Immunol.* 15:871.
- Shapiro, G. S., Aviszus, K., Murphy, J. and Wysocki, L. J. 2002. Evolution of Ig DNA sequence to target specific base positions within codons for somatic hypermutation. *J. Immunol.* 168:2302.
- Smith, D. S., Creadon, G., Jena, P. K., Portanova, J. P., Kotzin, B. L. and Wysocki, L. J. 1996. Di- and trinucleotide target preferences of somatic mutagenesis in normal and autoreactive B cells. *J. Immunol.* 156:2642.
- Cowell, L. G. and Kepler, T. B. 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.
- Shlomchik, M. J., Watts, P., Weigert, M. G. and Litwin, S. 1998. "Clone": a Monte-Carlo computer simulation of B cell clonal expansion, somatic mutation and antigen-driven selection. *Curr. Top. Microbiol. Immunol.* 229:173.
- Kleinstei, S. H., Louzoun, Y. and Shlomchik, M. J. 2003. Estimating hypermutation rates from clonal tree data. *J. Immunol.* 171:4639.
- Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. and Gottesman, K. 1987. Sequences of Proteins of Immunological Interest, 4th edn. US Government Printing Office, Washington DC. 165.
- Souto-Carneiro, M. M., Sims, G. P., Girschick, H., Lee, J. and Lipsky, P. E. 2005. Developmental changes in the human heavy chain CDR3. *J. Immunol.* 175:7425.
- Parr, T. B., Johnson, T. A., Silberstein, L. E. and Kipps, T. J. 1994. Anti-B cell autoantibodies encoded by VH 4-21 genes in human fetal spleen do not require *in vivo* somatic selection. *Eur. J. Immunol.* 24:2941.
- Dorner, T., Brezinschek, H. P., Brezinschek, R. I., Foster, S. J., Domiati-Saad, R. and Lipsky, P. E. 1997. Analysis of the frequency and pattern of somatic mutations within nonproductively rearranged human variable heavy chain genes. *J. Immunol.* 158:2779.
- Brezinschek, H. P., Brezinschek, R. I. and Lipsky, P. E. 1995. Analysis of the heavy chain repertoire of human peripheral B cells using single-cell polymerase chain reaction. *J. Immunol.* 155:190.
- Volpe, J. M., Cowell, L. G. and Kepler, T. B. 2006. SoDA: implementation of a 3D alignment algorithm for inference of antigen receptor recombinations. *Bioinformatics* 22:438.
- Pocock, M., Down, T. and Hubbard, T. 2000. BioJava: open source components for bioinformatics. *ACM SIGBio Newsl.* 20:10.
- Taki, S., Meiering, M. and Rajewsky, K. 1993. Targeted insertion of a variable region gene into the immunoglobulin heavy chain locus. *Science* 262:1268.
- Dal Porto, J. M., Haberman, A. M., Shlomchik, M. J. and Kelsoe, G. 1998. Antigen drives very low affinity B cells to become plasmacytes and enter germinal centers. *J. Immunol.* 161:5373.
- Hannum, L. G., Haberman, A. M., Anderson, S. M. and Shlomchik, M. J. 2000. Germinal center initiation, variable gene region hypermutation, and mutant B cell selection without detectable immune complexes on follicular dendritic cells. *J. Exp. Med.* 192:931.
- Chen, J., Trunstine, M., Alt, F. W. et al. 1993. Immunoglobulin gene rearrangement in B cell deficient mice generated by targeted deletion of the JH locus. *Int. Immunol.* 5:647.
- Hershberg, U. and Shlomchik, M. J. 2006. Differences in potential for amino acid change following mutation reveals distinct strategies for κ and λ light chain variation. *Proc. Natl Acad. Sci. USA* 103:15963.
- Lefranc, M. P., Giudicelli, V., Kaas, Q. et al. 2005. IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res.* 593.

- 31 Shlomchik, M. J., Euler, C. W., Christensen, S. C. and William, J. 2003. Activation of rheumatoid factor (RF) B cells and somatic hypermutation outside of germinal centers in autoimmune-prone MRL/lpr mice. *Ann. N. Y. Acad. Sci.* 987:38.
- 32 Oprea, M., Cowell, L. G. and Kepler, T. B. 2001. The targeting of somatic hypermutation closely resembles that of meiotic mutation. *J. Immunol.* 166:892.
- 33 Allen, D., Cumano, A., Dildrop, R. *et al.* 1987. Timing, genetic requirements and functional consequences of somatic hypermutation during B-cell development. *Immunol. Rev.* 96:5.
- 34 Storey, J. D. and Tibshirani, R. 2003. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA.* 100:9440.
- 35 Lenze, D., Berg, E., Volkmer-Engert, R. *et al.* 2006. Influence of antigen on the development of MALT lymphoma. *Blood* 107:1141.
- 36 Volkheimer, A. D., Weinberg, J. B., Beasley, B. E. *et al.* 2007. Progressive immunoglobulin gene mutations in chronic lymphocytic leukemia: evidence for antigen-driven intraclonal diversification. *Blood* 109:1559.
- 37 Margolin, D. H., Saunders, E. H., Bronfin, B. *et al.* 2006. Germinal center function in the spleen during simian HIV infection in rhesus monkeys. *J. Immunol.* 177:1108.
- 38 Yang, F., Waldbieser, G. C. and Lobb, C. J. 2006. The nucleotide targets of somatic mutation and the role of selection in immunoglobulin heavy chains of a teleost fish. *J. Immunol.* 176:1655.
- 39 Ghosh, S. and Huber, B. T. 2007. Clonal diversification in OspA-specific antibodies from peripheral circulation of a chronic Lyme arthritis patient. *J. Immunol. Methods* 321:121.
- 40 Odegard, V. H. and Schatz, D. G. 2006. Targeting of somatic hypermutation. *Nat. Rev. Immunol.* 6:573.
- 41 Unniraman, S. and Schatz, D. G. 2007. Strand-biased spreading of mutations during somatic hypermutation. *Science* 317:1227.