The Journal of Immunology

Intraclonal Diversity in Follicular Lymphoma Analyzed by Quantitative Ultradeep Sequencing of Noncoding Regions

Janice M. Spence, Andrew Abumoussa, John P. Spence and W. Richard Burack

J Immunol published online 13 October 2014
http://www.jimmunol.org/content/early/2014/10/10/jimmunol.1401699

The Journal of Immunology is published twice each month by The American Association of Immunologists, Inc., 9650 Rockville Pike, Bethesda, MD 20814-3994.

Print ISSN: 0022-1767 Online ISSN: 1550-6606.

Published October 13, 2014, doi:10.4049/jimmunol.1401699

Intraclonal Diversity in Follicular Lymphoma Analyzed by Quantitative Ultradeep Sequencing of Noncoding Regions

Janice M. Spence,* Andrew Abumoussa,† John P. Spence,‡ and W. Richard Burack*

Cancers are characterized by genomic instability, and the resulting intraclonal diversity is a prerequisite for tumor evolution. Therefore, metrics of tumor heterogeneity may prove to be clinically meaningful. Intraclonal heterogeneity in follicular lymphoma (FL) is apparent from studies of somatic hypermutation (SHM) caused by activation-induced deaminase (AID) in *IGH*. Aberrant SHM (aSHM), defined as AID activity outside of the *IG* loci, predominantly targets noncoding regions causing numerous "passenger" mutations, but it has the potential to generate rare significant "driver" mutations. The quantitative relationship between SHM and aSHM has not been defined. To measure SHM and aSHM, ultradeep sequencing (>20,000-fold coverage) was performed on *IGH* (~1650 nt) and nine other noncoding regions potentially targeted by AID (combined 9411 nt), including the 5′ untranslated region of *BCL2*. Single-nucleotide variants (SNVs) were found in 12/12 FL specimens (median 136 SHMs and 53 aSHMs). The aSHM SNVs were associated with AID motifs ($p < 0.0001$). The number of SNVs at *BCL2* varied widely among specimens and correlated with the number of SNVs at eight other potential aSHM sites. In contrast, SHM at *IGH* was not predictive of aSHM. Tumor heterogeneity is apparent from SNVs at low variant allele frequencies; the relative number of SNVs with variable allele frequency < 5% varied with clinical grade, indicating that tumor heterogeneity based on aSHM reflects a clinically meaningful parameter. These data suggest that genome-wide aSHM may be estimated from aSHM of *BCL2* but not SHM of *IGH*. The results demonstrate a practical approach to the quantification of intratumoral genetic heterogeneity for clinical specimens. ***The Journal of Immunology*****, 2014, 193: 000–000.**

Follicular lymphoma (FL), particularly the low-grade subtype, is the most prevalent non-Hodgkin lymphoma in the United States and Europe. Although almost all patients with low-grade FL initially experience an indolent course, about a third will suffer a sudden and aggressive transformation of their disease. The management approaches for patients with low-grade (grade 1 or 2) FL are variable, reflecting the profound uncertainty about what the future holds for each of these patients. Therefore, better prognostic approaches are needed for patients with low-grade histology. Although the approach to patients with grade 3 FL is more standardized, the histologic line is not distinct between the lower grades and grade 3. Patients with low-grade FL are often considered incurable, whereas those with grade 3 are candidates for intensive therapies with curative intent. Therefore, understanding the biology that distinguishes grade 3 from lower grades, as well as the biology that drives transformation, could directly affect selection of therapy (1, 2).

*Department of Pathology and Laboratory Medicine, University of Rochester Medical Center, Rochester NY 14642; †Department of Computer Science, University of Rochester, Rochester NY 14642; and ‡NGS Tech, Webster, NY 14580

ORCID: 0000-0002-5584-584X (W.R.B.).

Received for publication July 3, 2014. Accepted for publication September 12, 2014.

This work was supported in part by a pilot grant (to W.R.B.), administered through the University of Rochester Clinical Translational Science Institute under Award UL1 RR024160, from the National Center for Research Resources and the National Center for Advancing Translational Sciences of the National Institutes of Health and by Award P50 CA130805 from the National Cancer Institute.

Address correspondence and reprint requests to Dr. W. Richard Burack, Department of Pathology and Laboratory Medicine, University of Rochester Medical Center, Box 628, 601 Elmwood Avenue, Rochester, NY 14642. E-mail address: Richard_Burack@urmc.rochester.edu

Abbreviations used in this article: AID, activation-induced cytidine deaminase; aSHM, aberrant SHM; DDiMAP, deep-drilling iterative mapping; FL, follicular lymphoma; NGS, next-generation sequencing; SHM, somatic hypermutation; SNV, single-nucleotide variant; VAF, variant allele frequency.

The hallmark genetic lesion of FL is the t(14;18) translocation of the antiapoptosis *BCL2* gene into the *IGH* locus, placing expression of the native BCL2 under the influence of the strong *IGH* enhancer. However, this translocation is not sufficient for lymphomagenesis, and additional driver mutations are required (3–5). The subsequent transformation of indolent FL to aggressive lymphoma is due to acquisition of yet more driver mutations, a number of which were defined recently through in-depth genomic-scale analyses of paired low-grade and subsequent specimens from the same patient obtained at progression or histologic transformation (6–8). In general, driver mutations are a tiny fraction of the mutations present in any tumor. The vast majority of mutations are passenger mutations (9–11). By definition, passenger mutations do not enhance the evolutionary fitness, and any given passenger mutation has, at most, a limited effect on tumor biology (12). We hypothesize that the greater the number of passenger mutations in a tumor, the more mutagenetic stress or less effective DNA repair. This suggests the hypothesis that FL patients with more passenger mutations are at highest risk for acquisition of additional, clinically significant driver mutations.

In many cancers, intraclonal heterogeneity has been suggested as a key feature allowing tumor evolution and therapy resistance; however, in most cases, the mechanism(s) of ongoing mutagenesis to generate additional driver mutations is unclear (13–22). DNA from FL cells, similar to DNA from germinal center B cells, shows evidence of the activity of activation-induced cytidine deaminase (AID), a member of the APOBEC family. APOBECs appear to be responsible for a significant fraction of the mutations in several solid tumor types (9–11). AID produces point mutations and indels in the coding region of the Ig genes, particularly *IGH*, giving rise to Ab diversification; however, these coding region mutations may represent a small fraction of the total mutations induced by AID. AID activity is limited to ssDNA and generally targets the first 1–2 kb downstream from the start of transcription, which, in most human genes, is noncoding (23). The 5′ untranslated

portions of a number of proto-oncogenes in lymphoma cells were shown to be affected by AID (24). Further downstream, driver mutations may be the result of this AID-induced collateral damage (25). The greater abundance of AID-induced passenger mutations compared with driver mutations suggests that the role of AID in creating tumor heterogeneity may be better estimated by analysis of passenger mutations rather than by analysis of driver mutations.

The possible clinical application of molecular measures of tumor heterogeneity is a relatively new area in need of investigation. In FL, it is known that AID remains active, as previously demonstrated by ongoing somatic hypermutation (SHM) in rearranged IgH, raising the likely potential for ongoing AID-mediated aberrant SHM (aSHM) to drive intraclonal heterogeneity (26, 27). A critical question in FL is whether the tumor-specific *IGH*, as the primary target of AID, is the best indicator of overall AID activity and resultant genomic damage. On one hand, the unique *IGH* chromosomal rearrangement makes the tumor-specific IgH molecule readily identifiable and eliminates the influence of nontumor cells on overall mutation assessments. On the other hand, estimating AID activity by sequencing *IGH* is confounded by the requirement to produce functional IgH, eliminating cells with irreparably mutated *IGH*, and specific mutations in the *IGH* may be positively selected in the germinal center reaction (28–30). Additionally, SHM of *IGH* represents the normal physiologic role of the tightly regulated AID; thus, it may not be the best indicator of the aberrant process responsible for genome-wide damage. Although the ultimate issue is whether intraclonal diversity marked by the number of passenger mutations predicts prognosis, there are preliminary questions that must be addressed. First, it is unclear whether the amount of SHM of the *IGH* locus predicts the aSHM at non-*IGH* sites. Second, there is little or no data that the subclonal structure of FL populations correlates with any biologic or clinical feature. Therefore, we also asked whether an assay designed to quantitatively characterize intraclonal diversity could allow any associations to be made with well-defined biologic or clinical features.

Materials and Methods

The primary human specimens were frozen cell pellets from lymph node biopsies from 12 FLs and three hyperplastic lymph nodes as a source of nonmalignant polyclonal B cells. The specimens were mechanically separated, minced, and passed through a 70-μm nylon mesh cell strainer under sterile conditions. The resultant single-cell suspensions were washed with RPMI 1640 medium, counted, and cryopreserved for future analysis. No enrichment of malignant cells was performed. All specimens were obtained from the Human Hematological Malignancy Tissue Bank at the University of Rochester James P. Wilmot Cancer Institute, in accordance with Institutional Review Board–approved protocols and the Declaration of Helsinki principles. A human cell line, HEK 293, was used as a clonal nonlymphoid control. DNA was extracted (QIAamp DNA Mini Kit, QIAGEN, Valencia, CA) from $\sim 5 \times 10^6$ cells and quantified by spectrophotometry (NanoDrop, Wilmington, DE). All lymphoma specimens contained >80% tumor cells, as estimated by flow cytometry, and fulfilled standard diagnostic criteria, including histologically nodular pattern, clonality detection by flow cytometry showing bright CD20, L-chain restriction, and CD10 expression (31). All FL specimens showed the nodular pattern of BCL2 overexpression that is typical of FL by standard clinical immunohistochemical methods. All FL specimens had an IgH:BCL2 translocation demonstrated by a PCR-based system (32).

Regions for targeted resequencing were selected on the basis of published reports indicating or suggesting that these may be targets of aSHM in various forms of lymphoma or in the B cells of mice with DNA repair deficiencies (Table I). These include 5′ portions of *BCL2*, *BCL6*, *PIM1*, *PAX5*, *RHOH*, *CD83*, *MYC*, and *SYK* (24, 33, 34). Although the PCR amplicon for *BCL2* is 1078 nt, there is a guanine and cytosine–dense region beginning at position 850 that precluded reliable sequencing of the 3′ region; all results for *BCL2* refer to the 5′ 850-nt region and exclude the last 228 nt of the amplicon.

The *IGH* region studied varied slightly, depending on the details of the productively rearranged, clonal *IGH* in each specimen. In all cases, the sequence included the entire coding region of the *IGHV* from the ATG to junctional *IGHJ* and extending into the remaining, unused portions of the *IGHJ* to include all of JH6. The clonal *IGHV* sequences from each FL specimen were identified through Sanger sequencing of homoduplexed, gel-purified PCR amplicons using BioMed2 primers (35). The final *IGH* amplicons were generated with *IGHV*-specific primers and *IGHJ* primer downstream of JH-6. See Table I for a list of primer locations, sequences, and amplicon sizes. PCR was performed according to the manufacturer's instructions with Phusion Hot Start High-Fidelity DNA Polymerase (NEB,

Table I. Primers and regions studied

Gene	Chr	Primer	Chromosome Location[a]	Sequence	Size (bases)
BLC2	18q21	Pv295	18:60986652	5′-GCTCTTGAGATCTCCGGTTGGGATTCC-3′	1078
		Pv265	18:60985575	5′-GGTAGCGGCGGGAGAAGTCGTCG-3′	
BCL6	3q27	Pv239	3:187463235	5′-GGTGATGCAAGAAGTTTCTAGGAAAGG-3′	948
		Pv238	3:187462288	5′-CGGCTCTCATTAGGAAGATCACG-3′	
CD83	6p23	Pv299	6:14117875	5′-CCCCGGCCTAAGCGGGACTAGGAG-3′	1558
		Pv302	6:14119432	5′-CAGAGCACCTTTGCAATTTATAGGG-3′	
MYC	8q24	Pv244	8:128748420	5′-AAAGAACGGAGGGAGGGATCG-3′	1104
TSS-1		Pv247	8:128749532	5′-ACCCAACACCACGTCCTAACACCT-3′	
MYC	8q24	Pv246	8:128750451	5′-CTTTAACTCAAGACTGCCTCCCG-3′	660
TSS-2		Pv245	8:128751110	5′-TCGTTGAGAGGGTAGGGGAAGAC-3′	
PAX5	9p13	Pv248	9:37026233	5′-TACGGGTCCAAGCCAGGGTTCTCC-3′	888
		Pv249	9:37027120	5′-GCCCCAGAGACTCGGAGAAGCAGA-3′	
PIM1	6p21	Pv241	6:37138166	5′-CAGCGCCCTCAGTTGTCCTCCG-3′	1106
		Pv240	6:37139271	5′-TCACCATCGAAGTCCGTGTAGAC-3′	
RHOH	4p13	Pv243	4:40198904	5′-TCGGCATTCTGCAACAGGTAAGG-3′	971
		Pv242	4:40199874	5′-CCTTCTCCTTCTACCGACACTTC-3′	
SYK	9q22	Pv290	9:93564046	5′-GCGTTAAGGAAGTTGCCCAAAATGAG-3′	1098
		Pv292	9:93565143	5′-CCAAGAATCAGATATGGCACAACATC-3′	
IGHJ	14q32	Pv235	14:106329305	5′-CGCCCAGGTCCCCTCGGAACATGCC-3′	
*IGHV*3-48	14q32	Pv272	14:106994331	5′-GCCTTAGCCCTGGATTCCAAGGCATT-3′	
*IGHV*1-18	14q32	Pv276	14:106642227	5′-GGTAGGGGATGCGTGGCCTCTAAC-3′	
*IGHV*4-39	14q32	Pv280	14:106878110	5′-GTCCAACTCATAAGGGAAATGCTTTCTG-3′	
*IGHV*1-8	14q32	Pv282	14:106539605	5′-GGTAAATATAGGTATATTGGTGCCCTG-3′	
*IGHV*1-46	14q32	Pv273	14:106967471	5′-GAGGGTCTTCTGCTTGCTGGCTGTAGC-3′	
*IGHV*3-7	14q32	Pv274	14:106518864	5′-GTCTCAGAGAGGAGCCTTAGCCCTGGACTC-3′	
IgHV4 family	14q32	Pv251	Many sites	5′-CCCAGATGGGTCCTGTCCCAGGTGCAG-3′	
IgHV3 family	14q32	Pv252	Many sites	5′-AAGGTGTCCAGTGTGAGGTGCAG-3′	

[a]Location according to GRCh37.p13 (http://www.ncbi.nlm.nih.gov, accessed 10/09/2013).

Table II. SNVs identified in FL specimens and controls

Targets	FL Specimens													LN Controls and HEK				
	1	2	3	4	5	6	7	8	9	10	11	12	Median	LN A	LN B	LN C	Median	HEK
BCL2	106	65	38	38	35	26	18	18	9	6	4	2	22	0	0	0	0	0
BCL6	39	13	7	5	3	5	9	5	6	2	3	2	5	0	0	0	0	4
RHOH	16	3	4	2	3	3	1	1	1	1	1	3	2.5	1	0	0	0	1
PAX5	6	5	7	4	0	2	2	2	1	1	1	1	2	0	0	0	0	1
MYC TSS-1	0	1	2	3	1	4	2	1	1	0	0	2	1	1	2	2	2	2
PIM1	11	1	4	0	0	0	2	0	0	2	0	1	0.5	0	2	0	0	1
CD83	7	0	2	0	0	1	1	0	1	2	1	1	1	1	0	0	0	0
MYC TSS-2	1	1	2	0	1	2	0	2	0	1	0	4	1	0	2	0	0	0
SYK	0	0	0	1	1	1	0	1	0	0	1	0	0	2	2	0	2	0
Total (not *BCL2*)	80	24	28	15	9	18	17	12	10	9	7	14	14.5	5	8	2	5	9
IGH	50	367	194	193	123	237	8	111	164	149	92	81	136	NA	NA	NA		NA
Sequenced nt	11,853	11,927	11,952	12,630	11,752	11,756	11,727	11,389	11,719	11,449	11,789	12,233		10,129	10,129	10,129		10,129

HEK, human epithelial kidney cell line 293 (tetraploid); LN, lymph node (reactive lymph node controls: A, B, and C); NA, not applicable; TSS, transcription start site.

Ipswich, MA) using HF buffer and 3% DMSO. The template amount (250 ng) was chosen to ensure sampling of sufficient cells (~40,000) for statistically valid estimation of low-frequency events. A total of 2 μg DNA (equimolar for each amplicon) with 0.1 equimolar spike of pBlue-Script II KS (Agilent Technologies, Santa Clara, CA) fragment as an internal negative control were used for library preparation and sequencing on a SOLiD 4 platform (Life Technologies, Carlsbad, CA) by the University of Rochester Genome Research Center. The length of sequence obtained for the *IGH* locus varied (range: 1260–2501; median, 1643), depending on *IGHJ* segment usage (where JH6 usage results in the shortest region assessed due to loss of all JH-intronic regions).

To quantify intraclonal heterogeneity based on regions of high-density mutagenesis, we developed a novel next-generation sequencing (NGS) analytic pipeline, deep-drilling iterative mapping (DDiMAP) (J.M. Spence, J.P. Spence, A. Abumoussa, and W.R. Burack, manuscript in preparation). A common limitation of many NGS aligners (or mappers) is their inability to align reads from genetic regions with dense mutations to reference sequences, as observed in *IGH*. Reads with multiple single-nucleotide variants (SNVs) on a single allele are either not mapped or are "soft-clipped" to exclude read regions containing the high density of SNVs from the alignment. In both cases, the data on regions with high mutations loads would be lost. DDiMAP overcomes these limitations and recovers the highly mutated sequence reads through an iterative, remapping process that augments the original reference sequences with additional reference sequences that contain the previously identified, high-confidence SNV calls. This enables the mapping of reads to mutational hot spots through an iterative approach. The relative efficacy of DDiMAP was demonstrated by its detection of many SNVs occurring at high variant allele frequencies (VAFs; and apparent by Sanger sequencing) that were undetected by NGS when the data were analyzed using several established and widely available approaches (36–38). These failures were uniformly due to the inability to map reads from highly mutated genetic regions to their corresponding reference. More than 80% of the SNVs in the *IGH* gene found by the iteration inherent in DDiMAP were undetected by standard, single-pass mapping. A legitimate concern with iterative mapping is the potential to include false-positive calls due to NGS platform errors. Inclusion parameters were set to prevent this positive feedback from arising, and the efficacy of this approach can be observed in the many regions that have no SNV calls (Table II), even with >20,000-fold coverage.

DDiMAP results in highly precise estimates of the VAF, allowing for the identification and quantification of subclonal populations within the tumor. When applied to a predetermined set of target genes, the pipeline is sensitive (80% sensitive for SNVs occurring at a frequency of 0.4% and >99% sensitive for SNVs at 1% frequency) and specific (false-discovery rate of 2.8% for all SNVs, with a maximum false-discovery rate of 4.2% for the SNVs with VAF < 5%).

Results

A number of proto-oncogenes and highly expressed transcripts were previously shown to be subjected to aSHM, resulting in numerous passenger mutations in diffuse large B cell lymphoma (24). We sought to quantify the extent of aSHM variation across cases of FL by examining these loci and others suggested by these prior studies (34). More specifically, we sought to assess whether the amount of physiologic SHM at the *IGH* locus predicted the extent of genome-wide aSHM. To characterize aSHM, both Sanger sequencing and ultradeep NGS (average read depth of 20,000), followed by DDiMAP analysis, were obtained for the 5′ noncoding regions from nine non-Ig genes in 12 cases of FL, three nonneoplastic control specimens (hyperplastic lymph nodes), and an epithelial cell line HEK 293. The regions studied were the 5′ transcribed (but largely untranslated) portions of *BCL2*, *BCL6*, *PIM1*, *PAX5*, *RHOH*, *CD83*, *MYC* (two transcription start sites), and *SYK*. All SNVs with a VAF > 20% by NGS were detected by Sanger methods; for *BCL2*, allele-specific PCR showed that SNVs detected by Sanger sequencing were in a single allele (data not shown), most likely the highly transcribed t(14;18). To characterize SHM, the clonal productively rearranged *IGH* locus of the 12 FL specimens was sequenced by Sanger methods for use as a patient-specific *IGH* reference (Table I). Therefore, *IGH* SHM refers to SNV variation around the clone-specific reference sequence.

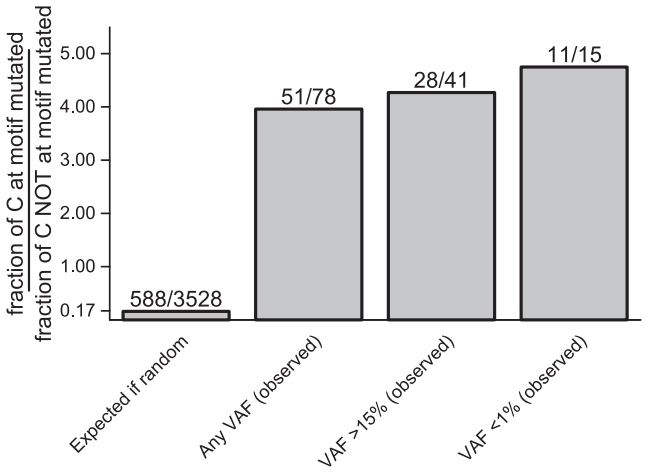

FIGURE 1. SNVs in the noncoding region of *BCL2* are attributable to AID. AID preferentially targets C residues in the motif 5′-WRCY (W is A or T; Y is C or T; R is A or G). The number of SNVs at C within WRCY compared with SNVs at C not in the motif provides an indicator of AID activity (labels on each bar: number within WRCY/number not within WRCY). If mutation of C residues were random, they would be expected to occur at the ratio of 0.17, reflecting the number C within WRCY and outside of WRCY. For the first bar, the numbers show the total count of C residues within (588) and outside WRCY (3528). The numbers of observed mutations at C are shown in the remaining bars, demonstrating a profound skewing of C mutations to occur within the WRCY motif. The skewing is apparent for SNVs, regardless of VAF, indicating that the low-frequency SNVs identified by DDiMAP are valid ($p < 0.0001$ in each subgroup; χ^2 for 2×2 based on each group versus expected). Similar results were obtained for the WA motif associated with *POLH* repair of AID-induced damage (data not shown). Data are for the 850-nt 5′ untranslated region of *BCL2* from all 12 FL specimens combined.

A median of 53 passenger mutations was found in the nine non-*IGH* regions of the FL cases (Table II; germline single-nucleotide polymorphisms are omitted). Note that the assay detects a much lower number of mutations in control specimens (median: 5). Furthermore, the SNVs showed a highly significant skewing to AID motifs ($p < 0.0001$); this skewing was true for SNVs at both high VAF (≥0.15) and low VAF (<0.01), indicating that SNVs contributing to ongoing diversification appear to be largely attributable to AID (Fig. 1).

The amount of AID-induced damage to *BCL2* and *IGH* was highly variable among specimens, and no correlation between the number of SNVs in *BCL2* and *IGH* was apparent (Fig. 2). Note that the density of mutations in the 5′ noncoding region of *BCL2* is remarkably high in some specimens; for example, the specimen with the highest mutation burden has ~100 mutations within the 850-nt region evaluated. The number of SNVs in *BCL2*, but not the number of SNVs in *IGH*, correlated with the combined SNV burden at the eight other noncoding regions scattered throughout the genome (Fig. 3).

Ongoing intratumoral diversification, based on the *IGH* locus, was seen, with the majority of the observed variation due to SNVs in the *IGHJ* intron (Fig. 4A). These SNVs were identified using the patient-specific clonal (rearranged and somatically mutated) *IGH* sequence as reference (rather than germline as reference); therefore, they reflect variation within the population around the most common sequence at the time of the biopsy, which we term the "clonal" sequence. Every SNV represents a branch point in the evolutionary tree, and the VAF for each SNV is the fraction of the population that is on the evolutionary limb beyond that branch point. To summarize these data, the mutations are ranked from the lowest to highest VAF and presented as the cumulative sum distribution (Fig. 4B).

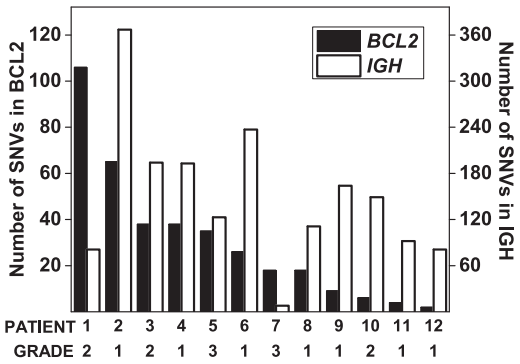

FIGURE 2. There is wide variation in the number SNVs detected in the 5′ noncoding region of *BCL2* and in *IGH*, with no correlation between these parameters (Pearson test, $r = +0.14$). Cases are ordered by number of SNVs in *BCL2* (*BCL2*, filled bars, left-hand axis; *IGH*, open bars, right-hand axis). The sequenced region in *BCL2* is 850 nt in all 12 cases. The amount of *IGH* region sequenced (coding and noncoding up to the 3′ end of JH6) varied among specimens, depending on the junctional JH used in the productively rearranged *IGH* allele. Even with correction for this variation, there was no significant correlation between the number of SNVs in the 5′ noncoding region of *BCL2* and in *IGH* (data not shown; Pearson $r = +0.26$).

In all FL cases, intraclonal diversity was seen based on the *IGH* coding and noncoding regions. The cumulative counts of SNVs as a function of VAF provide a comparison of the amount of ongoing diversification among cases (Fig. 5). In all cases, subclonal populations with VAF < 0.01 were detected. Furthermore, the relative greater slope of the cumulative curves at low VAF (<0.05) compared with high VAF shows that there are more small subclonal populations than major subclones within the tumor.

The plots of cumulative SNVs as a function of VAF for *BCL2* showed a distinctive pattern that distinguished SNVs present in the founder clone and SNVs reflecting subpopulations of the clone. For *BCL2*, germline sequence was used as the reference to allow identification of all SNVs in this locus that are associated with the neoplasm. In all specimens, a spike in the cumulative curves at VAF ~0.5 indicates a group of SNVs with uniform abundance and represent the clonal population; we refer to this component of the cumulative curve as the "clonal signal" (Fig. 6). The SNVs with lower VAF reflect descendants or "cousins" of the clonal population, and we refer to this component of the curve as the "subclonal signal."

All of the SNVs, both clonal and subclonal, appear to occur on the same allele of *BCL2*. The sequencing method that we used produced 50-bp reads. Therefore, we could assess whether each

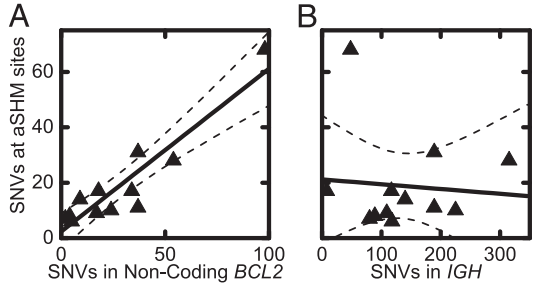

FIGURE 3. aSHM at eight non-IgH sites correlates with aSHM of *BCL2* (**A**) but not SHM of *IGH* (**B**). The same data used in Fig. 1 are the abscissa [*BCL2* in (A); *IGH* in (B)], and the aSHMs measured for eight non-*IGH* putative aSHM sites (*BCL6*, *PIM1*, *PAX5*, *RHOH*, *CD83*, *MYC* [two transcription start sites], and *SYK*) are the ordinate. Pearson correlation $r = +0.92$ in (A); exclusion of the single most extreme datum in (A) (farthest right-hand point) decreases r to +0.74, which is still a highly significant correlation. No correlation was detected between genome-wide aSHM and SHM of the *IGH* (B).

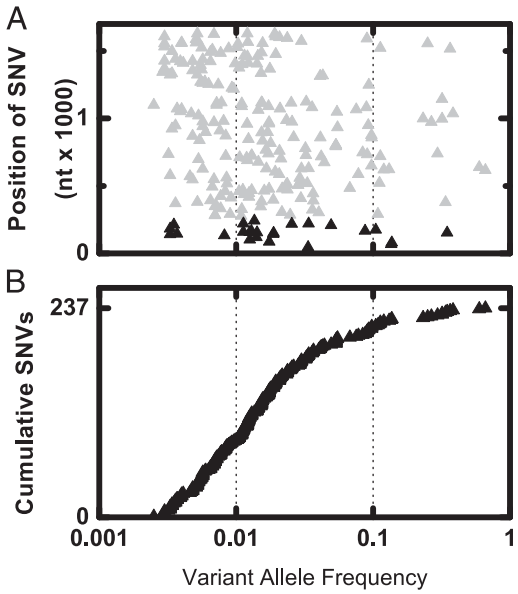

FIGURE 4. A representative plot of the cumulative SNVs ordered by VAF demonstrates the preponderance of SNVs are at low VAF. Each SNV in a single case is represented by an individual symbol. These data are for patient 6 in Fig 2. (**A**) These SNVs are scattered throughout the *IGH* coding (black triangles) and adjacent noncoding (gray triangles) regions. Furthermore, the vast majority of data are derived from the noncoding portion of *IGH*, a region in which cells carrying SNVs are presumably not subject to the selective pressures. (**B**) The count of these SNVs from lowest VAF to highest; the steepest portions of the cumulative SNV plot reflect the range of VAFs that are most highly represented in a specific specimen.

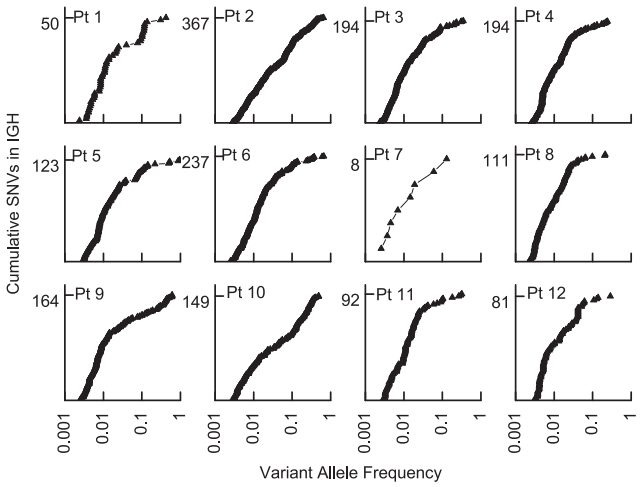

FIGURE 5. Variants of the *IGH* locus were detected in all 12 FL specimens (ordered as in Fig. 2). The total number of SNVs is shown for each patient. In general, the SNVs with VAFs < 0.05 were most numerous, and there was a monotonic decrease in the abundance of SNVs as the VAF increased (Fig. 3B is displayed as patient 6 for comparison). SNVs for the *IGH* are defined as relative to the productively rearranged *IGH* sequence obtained from the bulk tumor. Therefore, the SNVs for *IGH* reflect the current variation around the dominant population.

low-VAF (subclonal) SNV was on the same allele as the high-VAF (clonal) SNVs. For the 106 SNVs in *BCL2* in patient 1, 64 were clonal, and 42 were subclonal; all subclonal SNVs appear only in reads that contain clonal SNVs (data not shown). Therefore, SNVs with both high and low VAFs appear to occur on the same allele, which presumably is the translocated and overexpressed allele.

The relative contribution of the *BCL2*-derived subclonal signal to the total VAF varies strikingly among the specimens (Fig. 6). Two specimens, from patients 5 and 7, showed a highly significant lack of a subclonal phase based on the *BCL2* locus. These are the only grade 3 specimens in the collection.

Discussion

Ultradeep NGS of regions known to be aberrantly targeted by AID shows that the degree of aSHM is highly variable among specimens. However, in any single patient the relative amount of aSHM is consistent across multiple loci. Furthermore, although AID activity is high at the *IGH* locus, it appears that activity at *IGH* does not correlate with the genome-wide collateral damage induced by AID. Thus, sequencing of the 5′ *BCL2* region can allow a rapid assessment of aSHM in a variety of FL specimens.

Inclusion of large noncoding regions of the *IGH* locus, which have not been previously assessed in this context, allows the relationship between ongoing SHM and aSHM to be quantitatively assessed. No correlation was detected between the extent of ongoing SHM in *IGH* and aSHM in non-*IGH* genes. Explanations for this disparity include the possibility that there are distinct repair mechanisms targeted to *IGH* and non-*IGH* sites (34).

A recent study of paired low-grade and transformed FL specimens demonstrated abundant aSHM in transformed specimens but little to no aSHM in the precursor, low-grade specimens (6). Prior studies identified incremental additional aSHM in FL cases undergoing transformation (39). However, Halldórsdóttir et al. (33) noted that aSHM also accumulated between pairs of FL specimens, both of which showed low-grade histology, and that the relative accumulation with transformation was not demonstrably

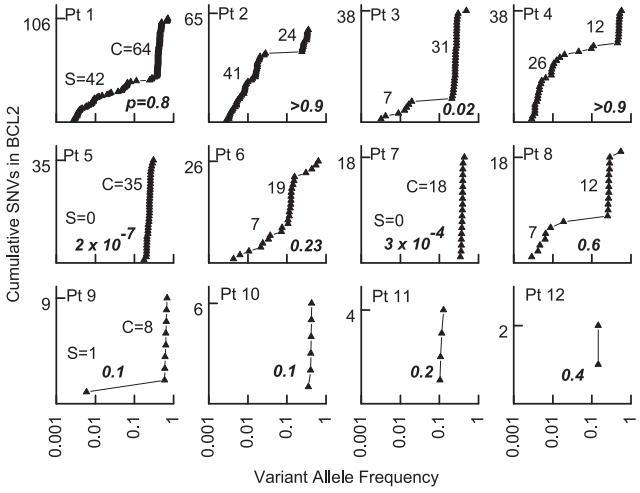

FIGURE 6. Variants in *BCL2* were detected in all 12 FL specimens, and the cumulative SNV curve allows distinction of SNVs present in founder (or the currently most dominant clone) and those reflecting newly derived subpopulations within the dominant clone (ordered as in Fig. 2). The total number of SNVs detected is shown for each patient. The vertical component of the cumulative curves at high VAF is the "clonal" phase, and the SNVs with lower VAF, reflecting the variation in the population, is the "subclonal" phase; the numbers of SNVs in the clonal (C) and subclonal (S) phases are shown. In contrast to the *IGH* data, SNVs for *BCL2* are defined as relative to germline sequence. Therefore, the SNVs for *BCL2* reflect all variation present in the founder cell for the neoplasm and those that have subsequently arisen during evolution of the neoplasm. To distinguish which specimens showed significant variation from the behavior of the aggregate of specimens, the aggregate number of SNVs in the clonal and subclonal phases for all specimens was determined (131 and 235; p values by binomial distribution are in bold). Specimens #5 and #7, the only grade 3 specimens in the collection, showed highly significant skewing from the aggregate ($p = 2 \times 10^{-7}$ and $p = 3.4 \times 10^{-4}$, respectively).

greater (i.e., aSHM increased with time and appeared independent of transformation). In this study, 8 of 12 low-grade FL specimens showed extensive aSHM. Histologic review showed that none of our specimens was transformed. Interestingly, patient 1, with the greatest number of SNVs in *BCL2* gene, experienced a particularly stormy clinical course and within a year had transformed to large cell lymphoma. This anecdotal observation, coupled with the extensive dataset from Pasqualucci et al. on the presence of aSHM in specimens of transformed FL, suggests that aSHM may precede and, perhaps lead to, transformation (6).

Our data suggest that qualitative features of a tumor's evolutionary tree, defined by passenger mutations, reflect the grade of FL, a highly significant clinical parameter. Although we did not predict the lack of diversity in grade 3 FL specimens, this observation is consistent with a prior study (40). This lack of apparent diversity could reflect a true cessation of aSHM-driven diversification or, although aSHM continued, a failure of subpopulations to increase to detectable levels. The data obtained for the IgH locus do not clearly distinguish between these two options: although the near-lack of SHM of the IgH locus for one of the grade 3 patients (#7) suggests that AID activity may be diminished, the other grade 3 patient (#5) showed abundant evidence of ongoing AID activity at the IgH locus. Similarly, we could not demonstrate that the lack of diversity in the grade 3 specimens was due to the higher proliferation. Because grade 3 FL is associated with a relatively high mitotic rate and Ki67 labeling, we measured Ki67 labeling in all 12 specimens. As expected, the grade 3 specimens had the highest labeling; however, among the nongrade 3 cases, no correlation with SNV abundance was detected (data not shown). Further work will be required to determine whether proliferation rate or some other feature of grade 3 specimens accounts for the apparent relationship between diversity and grade.

We show that tumor heterogeneity due to AID activity can be demonstrated and quantified using passenger mutations in discrete areas of the genome. Although genetic variation typically has been difficult to assess in clinical samples, we demonstrate a tractable approach. Applying these methods to a clinical trial collection, it will be possible to test whether clinical behavior can be predicted by the number of AID-induced passenger mutations.

Disclosures

The authors have no financial conflicts of interest.

References

1. Casulo, C., W. R. Burack, and J. W. Friedberg. 2014. Transformed lymphoma: a therapeutic challenge. *Blood.* In press.
2. Bernstein, S. H., and W. R. Burack. 2009. The incidence, natural history, biology, and treatment of transformed lymphomas. *Hematology Am. Soc. Hematol. Educ. Program*: 532–541.
3. Limpens, J., R. Stad, C. Vos, C. de Vlaam, D. de Jong, G. J. van Ommen, E. Schuuring, and P. M. Kluin. 1995. Lymphoma-associated translocation t(14;18) in blood B cells of normal individuals. *Blood* 85: 2528–2536.
4. Liu, Y., A. M. Hernandez, D. Shibata, and G. A. Cortopassi. 1994. BCL2 translocation frequency rises with age in humans. *Proc. Natl. Acad. Sci. USA* 91: 8910–8914.
5. Roulland, S., M. Faroudi, E. Mamessier, S. Sungalee, G. Salles, and B. Nadel. 2011. Early steps of follicular lymphoma pathogenesis. *Adv. Immunol.* 111: 1–46.
6. Pasqualucci, L., H. Khiabanian, M. Fangazio, M. Vasishtha, M. Messina, A. B. Holmes, P. Ouillette, V. Trifonov, D. Rossi, F. Tabbò, et al. 2014. Genetics of follicular lymphoma transformation. *Cell Reports* 6: 130–140.
7. Okosun, J., C. Bödör, J. Wang, S. Araf, C. Y. Yang, C. Pan, S. Boller, D. Cittaro, M. Bozek, S. Iqbal, et al. 2014. Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. *Nat. Genet.* 46: 176–181.
8. Green, M. R., A. J. Gentles, R. V. Nair, J. M. Irish, S. Kihira, C. L. Liu, I. Kela, E. S. Hopmans, J. H. Myklebust, H. Ji, et al. 2013. Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma. *Blood* 121: 1604–1611.
9. Alexandrov, L. B., S. Nik-Zainal, D. C. Wedge, S. A. J. R. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A.-L. Børresen-Dale, et al.; Australian Pancreatic Cancer Genome Initiative.; ICGC Breast Cancer Consortium.; ICGC MMML-Seq Consortium.; ICGC PedBrain. 2013. Signatures of mutational processes in human cancer. [Published erratum appears in 2013 *Nature* 502: 258.] *Nature* 500: 415–421.
10. Burns, M. B., L. Lackey, M. A. Carpenter, A. Rathore, A. M. Land, B. Leonard, E. W. Refsland, D. Kotandeniya, N. Tretyakova, J. B. Nikas, et al. 2013. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 494: 366–370.
11. Burns, M. B., N. A. Temiz, and R. S. Harris. 2013. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.* 45: 977–983.
12. McFarland, C. D., K. S. Korolev, G. V. Kryukov, S. R. Sunyaev, and L. A. Mirny. 2013. Impact of deleterious passenger mutations on cancer progression. *Proc. Natl. Acad. Sci. USA* 110: 2910–2915.
13. Campbell, P. J., E. D. Pleasance, P. J. Stephens, E. Dicks, R. Rance, I. Goodhead, G. A. Follows, A. R. Green, P. A. Futreal, and M. R. Stratton. 2008. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc. Natl. Acad. Sci. USA* 105: 13081–13086.
14. Campbell, P. J., S. Yachida, L. J. Mudie, P. J. Stephens, E. D. Pleasance, L. A. Stebbings, L. A. Morsberger, C. Latimer, S. McLaren, M.-L. Lin, et al. 2010. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* 467: 1109–1113.
15. Diaz, L. A., Jr., R. T. Williams, J. Wu, I. Kinde, J. R. Hecht, J. Berlin, B. Allen, I. Bozic, J. G. Reiter, M. A. Nowak, et al. 2012. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* 486: 537–540.
16. Ding, L., T. J. Ley, D. E. Larson, C. A. Miller, D. C. Koboldt, J. S. Welch, J. K. Ritchey, M. A. Young, T. Lamprecht, M. D. McLellan, et al. 2012. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 481: 506–510.
17. Gerlinger, M., and C. Swanton. 2010. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *Br. J. Cancer* 103: 1139–1143.
18. Hunter, C., R. Smith, D. P. Cahill, P. Stephens, C. Stevens, J. Teague, C. Greenman, S. Edkins, G. Bignell, H. Davies, et al. 2006. A hypermutation phenotype and somatic MSH6 mutations in recurrent human malignant gliomas after alkylator chemotherapy. *Cancer Res.* 66: 3987–3991.
19. Marusyk, A., and K. Polyak. 2010. Tumor heterogeneity: causes and consequences. *Biochim. Biophys. Acta* 1805: 105–117.
20. Walter, M. J., D. Shen, L. Ding, J. Shao, D. C. Koboldt, K. Chen, D. E. Larson, M. D. McLellan, D. Dooling, R. Abbott, et al. 2012. Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* 366: 1090–1098.
21. Yachida, S., S. Jones, I. Bozic, T. Antal, R. Leary, B. Fu, M. Kamiyama, R. H. Hruban, J. R. Eshleman, M. A. Nowak, et al. 2010. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* 467: 1114–1117.
22. Yip, S., J. Miao, D. P. Cahill, A. J. Iafrate, K. Aldape, C. L. Nutt, and D. N. Louis. 2009. MSH6 mutations arise in glioblastomas during temozolomide therapy and mediate temozolomide resistance [Published erratum appears in 2013 *Clin. Cancer Res.* 19: 4543–4544.] *Clin. Cancer Res.* 15: 4622–4629.
23. Cenik, C., A. Derti, J. C. Mellor, G. F. Berriz, and F. P. Roth. 2010. Genome-wide functional analysis of human 5′ untranslated region introns. *Genome Biol.* 11: R29.
24. Pasqualucci, L., P. Neumeister, T. Goossens, G. Nanjangud, R. S. Chaganti, R. Küppers, and R. Dalla-Favera. 2001. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* 412: 341–346.
25. Pasqualucci, L., G. Bhagat, M. Jankovic, M. Compagno, P. Smith, M. Muramatsu, T. Honjo, H. C. Morse, III, M. C. Nussenzweig, and R. Dalla-Favera. 2008. AID is required for germinal center-derived lymphomagenesis. *Nat. Genet.* 40: 108–112.
26. Zelenetz, A. D., T. T. Chen, and R. Levy. 1992. Clonal expansion in follicular lymphoma occurs subsequent to antigenic selection. *J. Exp. Med.* 176: 1137–1148.
27. Zhu, D., R. E. Hawkins, T. J. Hamblin, and F. K. Stevenson. 1994. Clonal history of a human follicular lymphoma as revealed in the immunoglobulin variable region genes. *Br. J. Haematol.* 86: 505–512.
28. McCann, K. J., P. W. Johnson, F. K. Stevenson, and C. H. Ottensmeier. 2006. Universal N-glycosylation sites introduced into the B-cell receptor of follicular lymphoma by somatic mutation: a second tumorigenic event? *Leukemia* 20: 530–534.
29. McCann, K. J., C. H. Ottensmeier, A. Callard, C. M. Radcliffe, D. J. Harvey, R. A. Dwek, P. M. Rudd, B. J. Sutton, P. Hobby, and F. K. Stevenson. 2008. Remarkable selective glycosylation of the immunoglobulin variable region in follicular lymphoma. *Mol. Immunol.* 45: 1567–1572.
30. Ottensmeier, C. H., A. R. Thompsett, D. Zhu, B. S. Wilkins, J. W. Sweetenham, and F. K. Stevenson. 1998. Analysis of VH genes in follicular and diffuse lymphoma shows ongoing somatic mutation and multiple isotype transcripts in early disease with changes during disease progression. *Blood* 91: 4292–4299.
31. Harris, N. L., S. H. Swerdlow, E. S. Jaffe, C. H. Ottensmeier, B. N. Nathwani, D. De Jong, T. Yoshino, and D. Spagnolo. 2008. Follicular Lymphoma. In *WHO Classification of Tumours of the Haematopoietic and Lymphoid Tissues.* S. H. Swerdlow, E. Campo, N. L. Harris, E. S. Jaffe, S. A. Pileri, H. Stein, J. Thiele, and J. W. Vardiman, eds. International Agency for Research on Cancer, Lyon, France, p. 220–226.
32. Spence, J. M., P. G. Rothberg, N. Wang, and W. R. Burack. 2011. Demonstration of array-based analysis for highly multiplexed PCR assays application to detection of IGH@-BCL2 translocations in FFPE follicular lymphoma specimens. *J. Mol. Diagn.* 13: 252–262.
33. Halldórsdóttir, A. M., M. Frühwirth, A. Deutsch, A. Aigelsreiter, C. Beham-Schmid, B. A. Agnarsson, P. Neumeister, and W. Richard Burack. 2008. Quantifying the role of aberrant somatic hypermutation in transformation of follicular lymphoma. *Leuk. Res.* 32: 1015–1021.

34. Liu, M., J. L. Duke, D. J. Richter, C. G. Vinuesa, C. C. Goodnow, S. H. Kleinstein, and D. G. Schatz. 2008. Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 451: 841–845.
35. van Dongen, J. J., A. W. Langerak, M. Brüggemann, P. A. Evans, M. Hummel, F. L. Lavender, E. Delabesse, F. Davi, E. Schuuring, R. García-Sanz, et al. 2003. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 17: 2257–2317.
36. David, M., M. Dzamba, D. Lister, L. Ilie, and M. Brudno. 2011. SHRiMP2: sensitive yet practical SHort Read Mapping. *Bioinformatics* 27: 1011–1012.
37. Homer, N., B. Merriman, and S. F. Nelson. 2009. BFAST: an alignment tool for large scale genome resequencing. *PLoS ONE* 4: e7767.
38. Homer, N., B. Merriman, and S. F. Nelson. 2009. Local alignment of two-base encoded DNA sequence. *BMC Bioinformatics* 10: 175.
39. Rossi, D., E. Berra, M. Cerri, C. Deambrogi, C. Barbieri, S. Franceschetti, M. Lunghi, A. Conconi, M. Paulli, A. Matolcsy, et al. 2006. Aberrant somatic hypermutation in transformation of follicular lymphoma and chronic lymphocytic leukemia to diffuse large B-cell lymphoma. *Haematologica* 91: 1405–1409.
40. Adam, P., T. Katzenberger, M. Eifert, M. M. Ott, A. Rosenwald, H. K. Müller-Hermelink, and G. Ott. 2005. Presence of preserved reactive germinal centers in follicular lymphoma is a strong histopathologic indicator of limited disease stage. *Am. J. Surg. Pathol.* 29: 1661–1664.