High-throughput sequencing of the B-cell receptor in African Burkitt lymphoma reveals clues to pathogenesis

Katharine A. Lombardo,^{1,2} David G. Coffey,^{1,3} Alicia J. Morales,¹ Christopher S. Carlson,⁴ Andrea M. H. Towler,¹ Sarah E. Gerds,⁵ Francis K. Nkrumah,⁶ Janet Neequaye,⁷ Robert J. Biggar,⁸ Jackson Orem,⁹ Corey Casper,^{1,3-5,10,11} Sam M. Mbulaiteye,¹² Kishor G. Bhatia,¹² and Edus H. Warren^{1,3}

¹Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA; ²Molecular and Cellular Biology Program and ³Department of Medicine, School of Medicine, University of Washington, Seattle, WA; ⁴Division of Public Health Sciences and ⁵Division of Vaccine and Infectious Diseases, Fred Hutchinson Cancer Research Center, Seattle, WA; ⁶Noguchi Institute for Medical Research and ⁷Department of Child Health, Medical School, University of Ghana, Accra, Ghana; ⁸Institute of Health and Biotechnical Innovation, Queensland University of Technology, Brisbane, QLD, Australia; ⁹Uganda Cancer Institute, Kampala, Uganda; ¹⁰Department of Global Health, School of Medicine, and ¹¹Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA; and ¹²Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD

Key Points

- High-throughput sequencing of primary African Burkitt lymphoma tumors suggests disrupted immunoglobulin rearrangements in BL progenitors.
- Extensive mutation of expressed and nonexpressed *IGH* rearrangements suggests multiple active mutational processes in BL tumors.

Burkitt lymphoma (BL), the most common pediatric cancer in sub-Saharan Africa, is a malignancy of antigen-experienced B lymphocytes. High-throughput sequencing (HTS) of the immunoglobulin heavy (*IGH*) and light chain (*IGK/IGL*) loci was performed on genomic DNA from 51 primary BL tumors: 19 from Uganda and 32 from Ghana. Reverse transcription polymerase chain reaction analysis and tumor RNA sequencing (RNAseq) was performed on the Ugandan tumors to confirm and extend the findings from the HTS of tumor DNA. Clonal *IGH* and *IGK/IGL* rearrangements were identified in 41 and 46 tumors, respectively. Evidence for rearrangement of the second *IGH* allele was observed in only 6 of 41 tumor samples with a clonal *IGH* rearrangement, suggesting that the normal process of biallelic *IGHD* to *IGHJ* diversity-joining (DJ) rearrangement is often disrupted in BL progenitor cells. Most tumors, including those with a sole dominant, nonexpressed DJ rearrangement, contained many *IGH* and *IGK/IGL* sequences that differed from the dominant rearrangement by < 10 nucleotides, suggesting that the target of ongoing mutagenesis of these loci in BL tumor cells is not limited to expressed alleles. *IGHV* usage in both BL tumor cohorts revealed enrichment for *IGHV* genes that are infrequently used in memory B cells from healthy subjects. Analysis of publicly available DNA sequencing and RNAseq data revealed that these same *IGHV* genes were overrepresented in dominant tumor-associated *IGH* rearrangements in several independent BL tumor cohorts. These data suggest that BL derives from an abnormal B-cell progenitor and that aberrant mutational processes are active on the immunoglobulin loci in BL cells.

Introduction

Burkitt lymphoma (BL) is the most common pediatric malignancy in sub-Saharan Africa, with an incidence as high as 4.7 cases per year for males and 3.0 cases per year for females (per 100 000 children <15 years of age).¹ Long-term survival for BL patients in sub-Saharan Africa is only 30% to 50%.² African BL is associated with Epstein-Barr virus (EBV)³ and holoendemic *Plasmodium falciparum* infection.⁴ BL tumors carry a *c-MYC*;immunoglobulin (Ig) chromosomal translocation, which is necessary, but not sufficient,⁵ for malignant transformation.

Submitted 25 August 2016; accepted 30 January 2017. DOI 10.1182/bloodadvances.2016000794.

The data reported in this article have been deposited in the Sequence Read Archive database (accession number SRP099346).

The full-text version of this article contains a data supplement.

BL tumor cells derive from germinal center centroblasts and express B-cell receptors (BCRs) of the IgM and IgD isotype that demonstrate a high level of inferred somatic hypermutation (SHM).⁶⁻¹¹ Improved understanding of the molecular etiology of African BL, in particular the role of the BCR, could lead to better treatment and inform strategies for prevention. Most prior studies of Ig rearrangements in BL tumors used capillary sequencing of tumor RNA, which generated a limited number of sequence reads, and detected only productive, expressed Ig rearrangements.^{7,9,11} In this study, we used high-throughput sequencing (HTS) of tumor genomic DNA (gDNA) to analyze the complete ensemble of Ig gene rearrangements in primary BL tumor samples from 2 distinct patient cohorts. Our results comprehensively define the tumor-associated Ig rearrangements in these African BL patient cohorts with unprecedented resolution and provide novel insights that have been missed by RNA sequencing (RNAseq), but have important implications for models of BL pathogenesis.

Methods

Study populations

BL tumor biopsies were acquired from 2 cohorts of African children after obtaining verbal or written informed consent using institutional review board–approved protocols. One cohort comprised 19 children who presented to the Uganda Cancer Institute (UCI) in Kampala, Uganda between August 2013 and March 2014. All tumors underwent primary pathologic review in Uganda and secondary pathologic review at a central US site. Biopsies were obtained from facial tumors involving the mandible and/or maxilla. Histologically confirmed, cryopreserved tumor biopsies from each patient were shipped from the UCI to the Fred Hutchinson Cancer Research Center in Seattle, WA in the vapor phase of liquid nitrogen. A second cohort of 32 cryopreserved, archival BL tumor samples, collected between 1975 and 1992 at Korle Bu Hospital in Accra, Ghana and stored long-term at the Frederick National Cancer Laboratory in Frederick, MD, were shipped to the Fred Hutchinson Cancer Research Center on dry ice. Samples from the Ghanaian cohort were primarily obtained from abdominal masses, most frequently from the ovaries, kidney, or spleen and were diagnosed based on clinical and cytological criteria; no histology was performed on these samples.

HTS of *IGH* and *IGK/IGL* chains

HTS of the Ig heavy (*IGH*) and Ig light (*IGK/IGL*) loci was performed on 1 to 3 μ g of gDNA from each tumor using the ImmunoSEQ platform from Adaptive Biotechnologies (Seattle, WA).^{12,13} In brief, libraries of rearranged Ig loci were generated from tumor gDNA by multiplex polymerase chain reaction (PCR) using sense primers specific for all variable (V) (and diversity [D]) gene segments and antisense primers specific for all joining (J) gene segments for the *IGH* and *IGK/IGL* loci. The inclusion of primers specific for *IGHD* enabled the capture of incomplete *IGHD* to *IGHJ* (DJ) rearrangements. The PCR products were ligated to adapters, and 130-nucleotide (nt) sequence reads encompassing the 3' portion of framework region (FR) 3 and the entirety of complementarity-determining region (CDR) 3 of the *IGH* and *IGK/IGL* loci were generated on the Illumina MiSeq platform. Amplification bias from multiplexed primer sets was removed by direct sequencing of synthetic DNA templates.¹⁴ Each molecule of amplified DNA was sequenced at least 10 times. The loading density on the MiSeq flow cell affected the total number of sequencing reads per sample. Adaptive Biotechnologies performed the initial analysis of raw sequence reads, including filtering and decomposition of reads into their component V, [D], and J segments^{15,16} and nontemplated junctional nucleotide insertions. Tumors were defined as clonal if a unique Ig rearrangement that comprised $\geq 15\%$ of the repertoire was detected. Sequencing data files are available on the Adaptive Biotechnologies Web site (<https://clients.adaptivebiotech.com/pub/lombardo-2017-bloodadvances>). Subsequent sequence analyses were performed using the LymphoSeq R package (<http://www.bioconductor.org/packages/LymphoSeq>) created by D.G.C.¹⁷ Analysis of SHM was performed using IMGT/V-Quest^{15,16} and IgBlast.¹⁸

HTS of *TRB*

HTS of the T-cell receptor β (*TRB*) locus was performed on tumor gDNA using the hsTCRB Kit (Adaptive Biotechnologies). Sequencing and analysis of *TRB* sequencing data were performed in the same manner as for the *IGH* and *IGK/IGL* loci.

RNAseq

RNAseq was performed on the 18 Ugandan tumors with an RNA integrity number >7 (no Ghanaian tumors were included due to low-quality RNA). RNAseq data are available in the Sequence Read Archive (accession number SRP099346). Polyadenylation-selected sequencing libraries were prepared with the Illumina TruSeq v2.0 Kit. Paired-end, 50-nt sequencing was performed on the Illumina HiSeq platform at a depth of 100 million reads per sample. Variants were called by the Broad Institute's Genome Analysis Tool Kit.¹⁹ Normal tissue samples were not available to filter germ line variants. Nonsynonymous mutations were called if they were: (1) predicted to be deleterious by metaSVM,²⁰ (2) not present in a database of known single nucleotide polymorphisms defined by the 1000 Genomes Project and a panel of nonmalignant samples from Sanger Cancer Genome Project sequencing, (3) observed in >1 tumor, and (4) previously reported in the Catalogue of Somatic Mutations in Cancer v77²¹ database.

Statistical analyses

Fisher's exact test was used to compare the frequency of *IGHV* utilization, and a Bonferroni correction was used to adjust for multiple comparisons. A Student *t* test was used to evaluate the frequency of SHM in *IGHV* gene regions based on EBV status. The Wald test was used to assess differential gene expression, with a Benjamini-Hochberg adjustment was used for multiple comparisons. Clonal relatedness was defined as the fraction of unique sequences with an edit distance <10 from the most common sequence, where 1 indicates all sequences are related to the most frequent sequence and 0 indicates none of the sequences are related. Clonal relatedness considers both unproductive rearrangements and sequence relatedness, and so it was used preferentially over the clonality metric, which only measures the repertoire diversity of productive rearrangements.

See supplemental Methods for additional methods.

Results

Study populations

From August 2013 to March 2014, diagnostic tumor samples were obtained from 19 children with histologically confirmed BL at the UCI in Kampala, Uganda. One of the patients was HIV-seropositive, and all 4 Ziegler disease stages were represented (Table 1). Only 5 patients (26%) had a sustained complete response to standard therapy (6 cycles of cyclophosphamide, vincristine, and methotrexate) and were alive 1-year after study enrollment. One-year overall survival for the Ugandan cohort was 42% (median = 253 days; supplemental Figure 1).

Archival, cytologically confirmed BL tumor samples from 32 children, collected between 1975 and 1992 at Korle Bu Hospital in Accra, Ghana, were received from the National Cancer Institute sample repository. Clinical information was only available for a subset of patients. Seventy-five percent of the children in the Ghanaian cohort had Ziegler stage C disease ($n = 20$; Table 1), and their 1-year survival rate was 51% (median = 458 days; $n = 29$; supplemental Figure 1). The Ugandan and Ghanaian cohorts comprised a total of 51 BL patients, which form the focus of this study.

Assessment of EBV DNA and RNA

The EBV status of each BL tumor was determined by PCR detection of EBV DNA sequences in tumor gDNA. EBV-encoded *EBER1*, *EBER2*, and *EBNA1* were detected in 47 of the 51 tumors (92%) (supplemental Figure 2).

Table 1. Clinical information for the Ugandan and Ghanaian BL patient cohorts

Characteristic	Ugandan cohort (n = 19)	Ghanaian cohort (n = 29*)
Sex, no. (%)		
Female	7 (37)	13 (45)
Male	12 (63)	16 (55)
Age at enrollment, y		
Median	7	9
Range	4-12	4-13
HIV status, no. (%)		
Negative	18 (95)	29 (100)
Positive	1 (5)	0 (0)
Ziegler disease stage, no. (%)†		
A	7 (37)	2 (10)
B	3 (16)	0 (0)
C	3 (16)	15 (75)
D	6 (31)	3 (15)

*Clinical data not available for all 32 patients.

†Disease stage data only available for 20 of the Ghanaian patients.

Reverse transcription polymerase chain reaction (RT-PCR) detection of *EBER* and *EBNA1* transcripts was used to evaluate the expression of EBV genes in a subset of tumors with quality RNA (RNA integrity number >5). *EBER1*, *EBER2*, and *EBNA1* RNA was detected in all tumors in which EBV DNA was detected (supplemental Figure 2). These results were confirmed by RNAseq analysis of EBV transcripts in the Ugandan cohort (n = 18). All 3 of the Ugandan BL patients with EBV-negative tumors had died 1 year after study enrollment.

HTS of Ig gene rearrangements in BL tumors

HTS was performed on gDNA from each of the 51 tumors to investigate the repertoire of *IGH* and *IGK/IGL* gene rearrangements. The sequencing strategy captured both incomplete, nonproductive DJ rearrangements, as well as complete V to D to J (VDJ) rearrangements of the *IGH* locus and V to J rearrangements of the *IGK/IGL* loci, using 130-nt read (supplemental Figure 3; supplemental Table 1).

HTS of the *IGH* and *IGK/IGL* loci revealed a clonal rearrangement, in which the dominant rearrangement comprised $\geq 15\%$ of the repertoire, in 41 and 46 of the 51 tumor samples, respectively (Figures 1A-B and 2A-B). An *IGH* VDJ rearrangement was detected in 31 of the 41 clonal tumors, and a sole DJ rearrangement was detected in the remaining 10 tumors. A clonal pattern of *IGK/IGL* V to J rearrangements was detected in all 41 of these cases. Five tumors carried a clonal *IGK/IGL* rearrangement, but a polyclonal *IGH* repertoire, in which neither a clonal VDJ or DJ rearrangement was observed. In 5 tumors, all from the Ghanaian cohort, both the heavy and light chain repertoires were polyclonal. In contrast to the overwhelmingly clonal nature of the *IGH* and *IGK/IGL* repertoires in BL tumors, the repertoires of B cells from bone marrow and peripheral blood of healthy donors were highly polyclonal, with no sequences comprising $\geq 2.25\%$ of the reads (supplemental Figure 4).

IGH and *IGK/IGL* sequence variation

Analysis of the *IGH* and *IGK/IGL* repertoires in clonal tumors revealed, in most cases, a large number of Ig sequence variants. Using neighbor-

joining tree estimation from edit distance matrices, phylogenetic trees were created for each tumor to visualize the relationships between all unique sequences (Figure 3A). Most of the unique sequences in each tumor were closely related to the dominant sequence, and clustered within the trees. A small fraction of highly dissimilar sequences in each tumor likely derived from nonmalignant B cells captured in the tumor biopsy. We compiled density plots of the edit distance, defined as the number of nucleotide differences between the most frequent sequence and all other unique sequences in the tumor, and found that the BL tumors were characterized by many related, low-edit distance (<10 nt) sequences (Figure 3B-C). In contrast, edit distance density plots from control peripheral blood mononuclear cell (PBMC)¹⁷ and bone marrow²² samples contained very few unique sequences that were closely related to, and therefore have low edit distances from, the most frequent sequence (Figure 3E-F). Furthermore, HTS of the *TRB* locus in all 51 tumors revealed very few sequences with a low edit distance, demonstrating that this phenomenon is restricted to the Ig loci in tumor cells (Figure 3D). The most frequent sequence and all variants with an edit distance <10 likely define the malignant population.

In tumors with clonal Ig rearrangements, 3 to 2091 unique *IGH* sequences with an edit distance <10 from the most frequent sequence were detected. Surprisingly, this rich sequence variation was observed in tumors with a dominant, nonproductive DJ rearrangement, as well as in those with a dominant, productive VDJ rearrangement. The median number of variants with an edit distance <10 was 262 for VDJ sequence families and 192 for DJ sequence families ($P = .54$). Sequence variation within VDJ and DJ families appeared to be uniformly distributed throughout the length of the sequencing reads (supplemental Figure 5). The high degree of sequence diversity within both VDJ and DJ sequence families demonstrates that mutagenesis of *IGH* in BL tumor cells is not limited to productively rearranged, expressed alleles.

We calculated the clonal relatedness of each tumor, defined as the fraction of all unique sequences with an edit distance <10 from the most frequent sequence. Most tumors had high *IGH* clonal relatedness (median 0.48; n = 51) (Figure 3H). To evaluate whether the sequence variation observed in BL tumors might be attributable to PCR or sequencing error, we compared the sequence diversity in BL tumors with that observed in clonal but nonmalignant samples (Figure 3G). The median clonal relatedness of the *TRB* locus in 14 primary CD8⁺ T-cell clones (generated using the same sequencing strategy and platform as used for the BL tumors) was 0.08. Thus, the high degree of BL-associated sequence variation is not likely due to sequencing error, but rather to active mutational processes within the tumor cells (Figure 3H). We propose that clonal relatedness has utility to evaluate ongoing hypermutation in BL tumors.

Monoallelic *IGH* rearrangements in BL tumor cells

At the pro-B-cell stage of development, DJ rearrangements occur on both *IGH* alleles, followed by sequential (if necessary) V to DJ rearrangements on each *IGH* allele (supplemental Figure 6).²³ If BL tumors develop from B cells that have followed this canonical developmental pathway, tumor cells should carry vestiges of 2 codominant *IGH* rearrangements, either 1 productive VDJ and 1 incomplete DJ rearrangement (Figure 2A, patient K00106900), or 2 VDJ rearrangements, 1 productive and 1 nonproductive (Figure 1A, patient 009-0231). However, only 6 of the 41 tumors with a clonal *IGH* repertoire appeared to carry 2 codominant *IGH* rearrangements. PCR, quantitative RT-PCR, and droplet digital PCR were employed to

Figure 1. HTS of gDNA and RNA from Ugandan BL tumors identifies the repertoire of Ig rearrangements and sequence variants in tumor cells. (A-B) Cumulative frequency plots of all unique *IGH* (A) or *IGK/IGL* (B) sequences identified by HTS of gDNA from the Ugandan BL cohort. Each segment in the bar plots represents a unique nucleotide sequence, and the color indicates the type of Ig rearrangement. Black lines separate the unique sequences, so highly polyclonal regions appear black in the figure. *Indicates that the tumor sample was EBV-negative; † indicates that the repertoire was classified as polyclonal. (C) Normalized *c-MYC* expression for each of the 18 Ugandan tumors on which RNAseq analysis was performed. RNAseq was not performed on sample 009-0210 due to poor quality RNA. (D) Genes with at least 1 single nucleotide variant (SNV) detected in RNAseq and predicted to be pathologic in each of the 18 Ugandan BL tumors. A red box indicates the presence of at least 1 SNV.

assess the rearrangement status of the *IGH* alleles in BL tumors. However, the presence of nonmalignant cells in the tumor biopsies, such as infiltrating T cells and stromal cells, which carry germ line *IGH* loci, prevented unambiguous resolution of this question.

IGHV gene segment utilization

To evaluate whether the BCRs expressed in African BL tumors exhibit stereotypy, manifested by biased Ig gene segment utilization, we

analyzed the *IGHV*-gene segment usage of the dominant clone in 30 tumor samples that contained a clonal VDJ rearrangement and for which an unambiguous gene segment assignment could be made. In addition, we examined the *IGHV* usage in 40 endemic and 33 sporadic tumor-associated BCRs of BL cases reported in the literature.²⁴⁻²⁶ We also used a novel computational method to identify *IGHV* utilization in RNAseq data from 28 sporadic BL tumors.²⁵ Combining the 30 evaluable cases from our 2 BL cohorts with these independent tumor

Figure 2. HTS of gDNA from Ghanaian BL tumors identifies the repertoire of Ig rearrangements in tumor cells. (A-B) Cumulative frequency plots of all unique *IGH* (A) or *IGK/IGL* (B) sequences identified by HTS of gDNA from the Ghanaian BL cohort. Each segment in the bar plots represents a unique nucleotide sequence, and the color indicates the type of Ig rearrangement. Black lines separate the unique sequences, so polyclonal regions appear black in the figure. *Indicates that the tumor sample was EBV-negative; †indicates that the repertoire was classified as polyclonal.

cohorts allowed us to examine *IGHV* usage in 131 BL cases: 70 endemic and 61 sporadic. BL-associated *IGHV* usage was compared with the *IGHV* usage in unfractionated PBMC, bone marrow, and purified B-cell populations from 33 adult and 9 pediatric control subjects^{17,22,27} (Figure 4A). *IGHV3-30*, *IGHV3-21*, *IGHV3-07*, *IGHV4-59*, and *IGHV4-34* were preferentially used in BL BCRs as compared with the control B-cell populations (P values: 1.1×10^{-12} , 1.5×10^{-6} , 5.7×10^{-6} , 8.9×10^{-6} , and 1.5×10^{-2} , respectively; Figure 4B). After correcting for multiple comparisons, no *IGHV* genes were differentially used in sporadic vs endemic BL BCRs.

Transcription of BL tumor *IGH* and *IGK/IGL* rearrangements

Analysis of BL tumor RNA by RT-PCR and RNAseq was performed to assess transcription of productive tumor-associated *IGH* and *IGK/IGL* rearrangements identified by HTS of tumor gDNA. This analysis focused on 18 Ugandan BL tumors from which RNA of sufficient quality was extracted. RNAseq was able to identify 7 of 12 clonal, productive *IGH* VDJ rearrangements and 13 of 15 clonal, productive *IGK/IGL* rearrangements (supplemental Figure 7). Failure to detect transcription of dominant Ig rearrangements predicted by HTS of tumor gDNA was closely associated with the length of the CDR3. The average CDR3 length of tumor-associated *IGH* rearrangements not detected by RNAseq was 51 nt, which is longer than the length of the RNAseq reads (50 nt). In contrast, the average CDR3 length of *IGH* and *IGK/IGL* rearrangements that were detected by RNAseq was 43.5 and 31 nt, respectively.

RT-PCR was used to confirm transcription of dominant *IGH* and *IGK/IGL* rearrangements and to determine the isotype of the putative tumor-associated rearrangement (supplemental Figure 3). Transcripts encoding the predicted heavy chain were identified in 13 Ugandan BL tumors with dominant, productive *IGH* rearrangements (supplemental Figure 8). Two tumors carried a dominant,

nonproductive *IGH* VDJ rearrangement, only 1 of which contained a detectable transcript. The inferred isotype of tumors with a dominant VDJ rearrangement was IgM⁺IgD⁺ in 12 cases and IgG⁺ in 2 cases (supplemental Figure 8). Moreover, 15 of the 16 evaluable tumors expressed a detectable *IGK/IGL* rearrangement, although 2 were predicted to be nonproductive (supplemental Figure 8).

RNAseq differential gene expression analysis demonstrated twofold higher *c-MYC* expression in tumors carrying a dominant productive *IGH* rearrangement, compared with those for which a productive rearrangement was not identified by HTS ($P = .037$, Figure 1C). No significant differences, however, were seen in the expression of BCR signaling pathway genes in tumors with and without a dominant, productive *IGH* rearrangement. RNAseq variant analysis revealed sequence variants in genes previously reported to be mutated in BL, including *ID3*, *TP53*, *SMARCA4*, *ZNF587*, and *FOXO1*,^{24,25,28-30} as well as 14 genes that are mutated in other cancers, including *NOTCH1*, *PAX5*, and *TFAP4* (Figure 1D; supplemental Table 2). Two tumors with a polyclonal *IGH* repertoire and a clonal *IGK/IGL* rearrangement by HTS of gDNA carried mutations in several genes mutated in other BL tumors, including *TP53*, *POMC*, *COL6A3*, and *BSCL2*, supporting the diagnosis of BL in these cases.

SHM in BL BCRs

A PCR-based strategy was used to sequence the complete V region of productive *IGH* and *IGK/IGL* rearrangements not covered by HTS (supplemental Figure 3). Nucleotide-level analysis revealed a median of 23 candidate sites of SHM per *IGH* V region (range, 4–43; $n = 29$).^{15,16,18} Nonsynonymous SHM clustered in the CDRs as compared with the FRs (P values: FR1: CDR1, 7.1×10^{-6} ; CDR1:FR2, 1.0×10^{-4} ; FR2:CDR2, 5.2×10^{-7} ; CDR2:FR3, 7.6×10^{-5} ; Figure 5A). An analogous approach was used on the clonal *IGK/IGL* rearrangements, and a

Figure 3. HTS of *IGH* in BL tumors reveals large families of closely related sequences. (A) Phylogenetic tree of all *IGH* sequences observed in a representative BL tumor sample. Unique *IGH* VDJ gene rearrangements are indicated by node color. The red arrow indicates the dominant unique sequence identified in the tumor. (B-G) Density plots of nucleotide edit distance (the number of nucleotide differences between a given unique sequence and the most frequent sequence in the tumor) for all sequences identified in all BL *IGH* samples (B), BL *IGK/IGL* samples (C), BL *TRB* samples (D), nonmalignant bone marrow and PBMC *IGH* control samples (E), bone marrow control *IGK/IGL* samples (F), and for CD8⁺ T-cell clones (G). (H) Clonal relatedness scores for each of the sample populations listed above (defined as the total number of distinct unique sequences with an edit distance <10 from the most frequent sequence, divided by the total number of unique sequences).

median of 13 candidate sites of SHM per *IGK/IGL* V region were identified (range, 0–33; n = 23).

The median number of SHM sites in EBV-positive and -negative tumors were 24 and 13, respectively ($P = .014$; Figure 5B). This disparity has been cited to support the hypothesis that EBV-positive and -negative tumors arise from distinct germinal center subpopulations.⁸ A higher number of candidate sites of SHM was found in patients with a favorable clinical outcome, but this difference was not statistically significant (supplemental Figure 9).

Enrichment for NLGS in the CDRs

We examined the putative tumor-associated Ig sequences in the Ugandan and Ghanaian BL tumors for the canonical N-linked glycosylation site (NLGS) motif: Asn-X-Ser/Thr. Of the 33 clonal BL tumors from which a complete *IGH* or *IGK/IGL* V region sequence was captured, we identified 22 NLGS across 18 samples (Figure 5C). Nineteen of the sites were likely introduced by mutation, and 3 were present in the germ line *IGHVH4-34* sequence.^{15,16,18} The NLGS cluster within and around the CDRs, which could enable interactions between BL tumor cells and stimulatory molecules in the tumor microenvironment.

Discussion

This study demonstrates that Ig gene rearrangements in African BL tumors are characterized by far more molecular heterogeneity than suggested by previous studies, most^{6-11,31} but not all²⁴ of which were performed using capillary sequencing. HTS of the Ig loci in gDNA from

BL tumors identified at least 1 clonal locus in 90% (46/51) of tumors. Thirty-one tumors (61%) harbored a clonal *IGH* rearrangement, comparable to the clonality detection rate obtained by the BIOMED-2 FR3 primers (63%).^{32,33} In 5 Ghanaian tumors cytologically classified as BL, no dominant Ig rearrangements were identified by HTS. It is possible that these tumors were misdiagnosed as BL, particularly given the clinical limitations >20 years ago when the samples were acquired. It is also possible that the 15% threshold used to define clonal tumors was set too high; however, this stringency seemed most appropriate due to the likelihood of reactive B-cell expansion within the tumors. In the Ugandan cohort, RT-PCR and RNAseq confirmed transcription of 1 or both of the predicted productive *IGH* and *IGK/IGL* rearrangements in most tumors. We hypothesize that malignant transformation occurred in antigen-experienced B cells that expressed a functional heterotetrameric BCR, but that 1 or both of the BCR components may have been rendered nonproductive by mutation during disease pathogenesis.

HTS demonstrates that most BL tumors carry many closely related *IGH* and *IGK/IGL* sequences with low edit distances that collectively define the malignant population. This rich sequence diversity is not likely due to PCR or sequencing error, because it was not observed within the *IGH* repertoires of B cells from the blood or marrow of healthy donors, nor in the *TRB* repertoire in any of the 51 tumors or in 14 nonmalignant CD8⁺ T-cell clones, all generated using the same sequencing strategy. Thus, the mutational mechanism that creates this diversity is not active on *IGH* in nonmalignant B cells, nor on *TRB* in BL tumor-infiltrating T cells. Comparable *IGH* sequence diversity was observed within families of both VDJ and DJ rearrangements. Selective pressure

Figure 4. Biased *IGHV* gene utilization observed in BL tumors. (A) Heat map illustrating *IGHV* gene usage in the 30 endemic BL samples from our cohorts, 40 published endemic BL samples, and 61 published sporadic BL samples, as compared with 42 healthy controls. Highly enriched *IGHV* genes are indicated in red, and genes with significant differential utilization ($P < .05$) in BL tumor and control B cells are indicated by an asterisk. Both *IGHV* gene utilization and all samples are clustered by similarity. (B) Frequency of *IGHV* gene utilization in the 5 significantly differentially used *IGHV* genes. *IGHV* gene frequencies in BL tumors are plotted in red, and *IGHV* gene frequencies of healthy controls are plotted in blue.

from SHM-induced affinity maturation is only expected to occur on productive, actively transcribed Ig alleles.³⁴ These observations imply that a mutational mechanism that is not appropriately regulated by antigen-driven SHM and affinity maturation is active on both *IGH* and *IGK/IGL* in BL tumor cells. Although it is tempting to speculate that this mechanism involves activation-induced cytidine

deaminase (AID), particularly given that AID transcripts were detected in all tumors by RNAseq, this remains to be proved. *IGH* sequence evolution has also been reported by HTS of pediatric acute lymphoblastic leukemia, a malignancy of an early B-cell progenitor, in a strikingly similar magnitude as that discovered in BL.³⁵

Figure 5. Nonsynonymous sites of SHM and SHM-induced NLGS are enriched in CDRs. Analyses were performed on all clonal BL tumors for which the complete V region of the most frequent sequence could be determined by capillary sequencing ($n = 29$). (A) Sites of nonsynonymous SHM plotted per amino acid residue in each *IGH* gene region. Statistically significant differences in SHM incidence are indicated by asterisks. (B) Sites of nonsynonymous (NS) and synonymous (S) SHM plotted per nucleotide residue in EBV⁺ and EBV⁻ tumors by CDR or FR. (C) A schematic of an *IGH* or *IGK/IGL* gene rearrangement is shown and the location of all NLGS (Asn-X-Ser/Thr amino acid motifs) identified within a complete clonal BL tumor *IGH* and *IGK/IGL* sequences are indicated by red circles.

Because B cells undergo concurrent biallelic *IGH* DJ rearrangement early in development, both *IGH* alleles of a clonal population should be detected by HTS. However, HTS of tumor gDNA identified only 1 *IGH* rearrangement in 35 of the 41 (85%) clonal BL tumors. Identification of at least 1 rearrangement in 41 of the 51 (80%) tumors demonstrates that our HTS strategy could efficiently detect *IGH* rearrangements. Given the unique translocation involving the *IGH* locus that occurs in BL, one possible interpretation is that the DJ rearrangement on the second *IGH* allele may not have occurred in BL progenitors. DJ rearrangement may have been inhibited by the t(c-MYC;IGH) translocation that occurs in 80% of BL tumors (the remaining 20% of translocations are to the *IGK/IGL* loci).

Detection of only 1 *IGH* DJ rearrangement in BL tumors is consistent with a previous study that reported a high frequency of monoallelic DJ rearrangements in BL.³⁶ The consensus in the BL literature is that t(c-MYC;Ig) translocations occur in mature B cells due to recurrent antigenic stimulation and aberrant AID activity in the germinal center.³⁷⁻⁴⁰ However, our data raise the possibility that the c-MYC translocation may actually occur in a developing B cell, before that cell completes DJ rearrangement on both *IGH* alleles, suggesting that BL precursors may emerge much earlier in B-cell development. Another group recently identified a pre-B-cell population with concurrent expression of recombination activating genes and AID, providing additional support for the possibility that the t(c-MYC;Ig) translocation could occur early in development.⁴¹

Biased, or stereotyped, usage of particular *IGHV* genes was observed in BL tumors, suggesting that BL progenitors carrying particular Ig genes preferentially differentiate into the memory B-cell compartment. Thus, BL progenitors are preferentially selected based on their BCR, and likely their antigenic specificity. Chronic lymphocytic leukemia,⁴²⁻⁴⁶ mantle cell lymphoma,⁴⁷ and splenic marginal zone lymphoma⁴⁸ all demonstrate BCR stereotypy, strongly suggesting a role for antigens in the pathogenesis of a substantial subset of mature B-cell malignancies, including BL.

BL Ig sequences reportedly contain significantly more SHM-induced NLGS (82%) than nonmalignant antigen-experienced B cells (9%).³¹ In our BL cohorts, NLGS were highly enriched and clustered within CDRs, which could allow for interactions with molecules in the tumor microenvironment. Lectins on tumor-resident macrophages and dendritic cells can bind to NLGS and induce signaling through the BCR,⁴⁹ suggesting a mechanism by which the BCR could be activated in BL and promote tumor survival. Furthermore, despite extensive SHM, IgM, rather than IgG, dominated the BL BCR repertoire, suggesting an increased dependence on BCR signaling in BL tumors.^{50,51} BCR signaling has been implicated in the pathogenesis of a number of B-cell malignancies,^{43,44,52-54} against which BCR signaling pathway inhibitors have proven effective.^{55,56} These data suggest that these agents may also have utility for the treatment of African BL.

This study demonstrates that a distinct family of Ig rearrangements uniquely characterizes most African BL tumors and suggests that these sequences may have utility as a biomarker to diagnose and monitor disease progression. Given the frequency of nonproductive VDJ and DJ *IGH* rearrangements and the rich sequence variation detected in BL tumors, an RNA- or PCR-based approach would have limited utility. These findings support the potential use of a gDNA-based HTS approach for disease monitoring. Indeed, we have used HTS to detect BL-associated *IGH* sequences in matched blood, serum, and cerebrospinal fluid samples from the patients reported on in this study. Preliminary data suggest that the detection of these sequences may have prognostic value for BL patients (K.A.L. and E.H.W., manuscript in preparation).

Acknowledgments

The authors thank the M. J. Murdock Charitable Trust Partners in Science Program for support and Frederick Matsen for helpful discussions.

This work was supported by the National Institutes of Health National Cancer Institute (Cancer Center Support Grant P30 CA015704; and Chromosome Metabolism and Cancer training grant T32CA009657 [K.A.L.]), the Cancer Therapeutics Endowment, a Research Training Award for Fellows (D.G.C.) from the American Society of Hematology, and grant N01-CO-12400 from the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

Authorship

Contribution: K.A.L. and E.H.W. designed the study; K.A.L., A.J. M., and A.M.H.T. performed experiments; K.A.L. and D.G.C. performed data analysis; D.G.C. contributed analytical tools; K.A.L., D.G.C., C.S.C., S.M.M., K.G.B., and E.H.W. interpreted the data; S.E.G., F.K.N., J.N., R.J.B., J.O., C.C., S.M.M., and K.G.B. aided with sample acquisition; and K.A.L. and E.H.W. wrote the manuscript.

Conflict-of-interest disclosure: C.S.C. holds stock in Adaptive Biotechnologies, Inc. The remaining authors declare no competing financial interests.

ORCID profiles: K.A.L., 0000-0002-3344-5669; D.G.C., 0000-0003-3544-8836; C.S.C., 0000-0003-0561-3538; J.O., 0000-0002-9824-346X; C.C., 0000-0002-3609-661X; E.H.W., 0000-0002-9570-2755.

Correspondence: Edus H. Warren, Division of Clinical Research, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, D4-100, P.O. Box 19024, Seattle, WA 98109-1024; e-mail: ehwarren@u.washington.edu.

References

1. Orem J, Mbidde EK, Lambert B, de Sanjose S, Weiderpass E. Burkitt's lymphoma in Africa, a review of the epidemiology and etiology. *Afr Health Sci*. 2007;7(3):166-175.
2. Schmitz R, Cerbelli M, Pittaluga S, Wright G, Staudt LM. Oncogenic mechanisms in Burkitt lymphoma. *Cold Spring Harb Perspect Med*. 2014;4(2):a014282.
3. Magrath I. Epidemiology: clues to the pathogenesis of Burkitt lymphoma. *Br J Haematol*. 2012;156(6):744-756.

4. Kafuko GW, Burkitt DP. Burkitt's lymphoma and malaria. *Int J Cancer*. 1970;6(1):1-9.
5. Müller JR, Janz S, Goedert JJ, Potter M, Rabkin CS. Persistence of immunoglobulin heavy chain/c-myc recombination-positive lymphocyte clones in the blood of human immunodeficiency virus-infected homosexual men. *Proc Natl Acad Sci USA*. 1995;92(14):6577-6581.
6. Chapman CJ, Wright D, Stevenson FK. Insight into Burkitt's lymphoma from immunoglobulin variable region gene analysis. *Leuk Lymphoma*. 1998;30(3-4):257-267.
7. Chapman CJ, Mockridge CI, Rowe M, Rickinson AB, Stevenson FK. Analysis of VH genes used by neoplastic B cells in endemic Burkitt's lymphoma shows somatic hypermutation and intraclonal heterogeneity. *Blood*. 1995;85(8):2176-2181.
8. Bellan C, Lazzi S, Hummel M, et al. Immunoglobulin gene analysis reveals 2 distinct cells of origin for EBV-positive and EBV-negative Burkitt lymphomas. *Blood*. 2005;106(3):1031-1036.
9. Klein U, Klein G, Ehlin-Henriksson B, Rajewsky K, Küppers R. Burkitt's lymphoma is a malignancy of mature B cells expressing somatically mutated V region genes. *Mol Med*. 1995;1(5):495-505.
10. Tamaru J, Hummel M, Marafioti T, et al. Burkitt's lymphomas express VH genes with a moderate number of antigen-selected somatic mutations. *Am J Pathol*. 1995;147(5):1398-1407.
11. Riboldi P, Ikematsu W, Brambilla B, Caprani C, Gerosa M, Casali P. Diversity and somatic hypermutation of the Ig VHDJH, V kappa J kappa, and V lambda J lambda gene segments in lymphoma B cells: relevance to the origin of the neoplastic B cell clone. *Hum Immunol*. 2003;64(1):69-81.
12. Wu D, Emerson RO, Sherwood A, et al. Detection of minimal residual disease in B lymphoblastic leukemia by high-throughput sequencing of IGH. *Clin Cancer Res*. 2014;20(17):4540-4548.
13. DeWitt WS, Lindau P, Snyder TM, et al. A public database of memory and naive B-cell receptor sequences. *PLoS One*. 2016;11(8):e0160853.
14. Carlson CS, Emerson RO, Sherwood AM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. *Nat Commun*. 2013;4:2680.
15. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acid Res*. 2008;36(suppl 2):W503-W508.
16. Giudicelli V, Brochet X, Lefranc MP. IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. *Cold Spring Harb Protoc*. 2011;2011(6):695-715.
17. Kanakry CG, Coffey DG, Towler AM, et al. Origin and evolution of the T cell repertoire after posttransplantation cyclophosphamide. *JCI Insight*. 2016;1(5):e86252.
18. Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acid Res*. 2013;41(W1):W34-W40.
19. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11.10.1-11.10.33.
20. Dong C, Wei P, Jian X, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*. 2015;24(8):2125-2137.
21. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*. 2015;43(D1):D805-D811.
22. Bertucci CRM. Bone marrow from healthy adults. Available at: <https://clients.adaptivebiotech.com/pub/bone-marrow-healthy-adults-control>. Accessed January 27, 2017.
23. Alt FW, Yancopoulos GD, Blackwell TK, et al. Ordered rearrangement of immunoglobulin heavy chain variable region segments. *EMBO J*. 1984;3(6):1209-1219.
24. Amato T, Abate F, Piccaluga P, et al. Clonality analysis of immunoglobulin gene rearrangement by next-generation sequencing in endemic Burkitt lymphoma suggests antigen drive activation of BCR as opposed to sporadic Burkitt lymphoma. *Am J Clin Pathol*. 2016;145(1):116-127.
25. Schmitz R, Young RM, Ceribelli M, et al. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature*. 2012;490(7418):116-120.
26. Baptista MJ, Calpe E, Fernandez E, et al. Analysis of the IGHV region in Burkitt's lymphomas supports a germinal center origin and a role for superantigens in lymphomagenesis. *Leuk Res*. 2014;38(4):509-515.
27. Kolhatkar NS, Brahmandam A, Thouvenel CD, et al. Altered BCR and TLR signals promote enhanced positive selection of autoreactive transitional B cells in Wiskott-Aldrich syndrome. *J Exp Med*. 2015;212(10):1663-1677.
28. Love C, Sun Z, Jima D, et al. The genetic landscape of mutations in Burkitt lymphoma. *Nat Genet*. 2012;44(12):1321-1325.
29. Abate F, Ambrosio MR, Mundo L, et al. Distinct viral and mutational spectrum of endemic Burkitt lymphoma. *PLoS Pathog*. 2015;11(10):e1005158.
30. Richter J, Schlesner M, Hoffmann S, et al; ICGC MMML-Seq Project. Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nat Genet*. 2012;44(12):1316-1320.
31. Zhu D, Ottensmeier CH, Du MQ, McCarthy H, Stevenson FK. Incidence of potential glycosylation sites in immunoglobulin variable regions distinguishes between subsets of Burkitt's lymphoma and mucosa-associated lymphoid tissue lymphoma. *Br J Haematol*. 2003;120(2):217-222.
32. van Krieken JH, Langerak AW, Macintyre EA, et al. Improved reliability of lymphoma diagnostics via PCR-based clonality testing: report of the BIOMED-2 Concerted Action BHM4-CT98-3936. *Leukemia*. 2007;21(2):201-206.
33. EuroClonality. Available at: www.euroclonality.org/faq/. Accessed 8 March 2017.
34. Fukita Y, Jacobs H, Rajewsky K. Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity*. 1998;9(1):105-114.

35. Gawad C, Pepin F, Carlton VE, et al. Massive evolution of the immunoglobulin heavy chain locus in children with B precursor acute lymphoblastic leukemia. *Blood*. 2012;120(22):4407-4417.
36. Bhatia K, Gutierrez M, Magrath IT. Burkitt's lymphoma cells frequently carry monoallelic DJ rearrangements. *Curr Top Microbiol Immunol*. 1992;182:319-324.
37. Robbiani DF, Deroubaix S, Feldhahn N, et al. Plasmodium infection promotes genomic instability and AID-dependent B cell lymphoma. *Cell*. 2015;162(4):727-737.
38. Robbiani DF, Bothmer A, Callen E, et al. AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell*. 2008;135(6):1028-1038.
39. Greisman HA, Lu Z, Tsai AG, Greiner TC, Yi HS, Lieber MR. IgH partner breakpoint sequences provide evidence that AID initiates t(11;14) and t(8;14) chromosomal breaks in mantle cell and Burkitt lymphomas. *Blood*. 2012;120(14):2864-2867.
40. Takizawa M, Tolarová H, Li Z, et al. AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. *J Exp Med*. 2008;205(9):1949-1957.
41. Swaminathan S, Klemm L, Park E, et al. Mechanisms of clonal evolution in childhood acute lymphoblastic leukemia. *Nat Immunol*. 2015;16(7):766-774.
42. Agathangelidis A, Darzentas N, Hadzidimitriou A, et al. Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. *Blood*. 2012;119(19):4467-4475.
43. Dühren-von Minden M, Übelhart R, Schneider D, et al. Chronic lymphocytic leukaemia is driven by antigen-independent cell-autonomous signalling. *Nature*. 2012;489(7415):309-312.
44. Zwick C, Fadle N, Regitz E, et al. Autoantigenic targets of B-cell receptors derived from chronic lymphocytic leukemias bind to and induce proliferation of leukemic cells. *Blood*. 2013;121(23):4708-4717.
45. Seiler T, Woelfle M, Yancopoulos S, et al. Characterization of structurally defined epitopes recognized by monoclonal antibodies produced by chronic lymphocytic leukemia B cells. *Blood*. 2009;114(17):3615-3624.
46. Chu CC, Catera R, Hatzl K, et al. Chronic lymphocytic leukemia antibodies with a common stereotypic rearrangement recognize nonmuscle myosin heavy chain IIA. *Blood*. 2008;112(13):5122-5129.
47. Hadzidimitriou A, Agathangelidis A, Darzentas N, et al. Is there a role for antigen selection in mantle cell lymphoma? Immunogenetic support from a series of 807 cases. *Blood*. 2011;118(11):3088-3095.
48. Zibellini S, Capello D, Forconi F, et al. Stereotyped patterns of B-cell receptor in splenic marginal zone lymphoma. *Haematologica*. 2010;95(10):1792-1796.
49. Coelho V, Krysov S, Ghaemmaghami AM, et al. Glycosylation of surface Ig creates a functional bridge between human follicular lymphoma and microenvironmental lectins. *Proc Natl Acad Sci USA*. 2010;107(43):18587-18592.
50. Horikawa K, Martin SW, Pogue SL, et al. Enhancement and suppression of signaling by the conserved tail of IgG memory-type B cell antigen receptors. *J Exp Med*. 2007;204(4):759-769.
51. Waisman A, Kraus M, Seagal J, et al. IgG1 B cell receptor signaling is inhibited by CD22 and promotes the development of B cells whose survival is less dependent on Ig alpha/beta. *J Exp Med*. 2007;204(4):747-758.
52. Davis RE, Ngo VN, Lenz G, et al. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature*. 2010;463(7277):88-92.
53. Hoogeboom R, van Kessel KP, Hochstenbach F, et al. A mutated B cell chronic lymphocytic leukemia subset that recognizes and responds to fungi. *J Exp Med*. 2013;210(1):59-70.
54. Satchell KL, Strohmman MJ, Singletary J, et al. Self-antigen recognition by follicular lymphoma B-cell receptors. *Blood*. 2012;120(20):4182-4190.
55. Gopal AK, Kahl BS, de Vos S, et al. PI3Kδ inhibition by idelalisib in patients with relapsed indolent lymphoma. *N Engl J Med*. 2014;370(11):1008-1018.
56. Wang ML, Rule S, Martin P, et al. Targeting BTK with ibrutinib in relapsed or refractory mantle-cell lymphoma. *N Engl J Med*. 2013;369(6):507-516.