Published in final edited form as:
J Immunol. 2015 February 1; 194(3): 929–939. doi:10.4049/jimmunol.1402168.

The V gene repertoires of classical and atypical memory B cells in malaria-susceptible West African children[1]

Severin Zinöcker[*], **Christine E. Schindler**[*], **Jeff Skinner**[*], **Tobias Rogosch**[†], **Michael Waisberg**[*,‡], **Jean-Nicolas Schickel**[§], **Eric Meffre**[§], **Kassoum Kayentao**[¶], **Aïssata Ongoïba**[¶], **Boubacar Traoré**[¶], and **Susan K. Pierce**[*]

[*]Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD, USA

[†]Laboratory for Neonatology and Pediatric Immunology, Department of Pediatrics, Philipps-University, Marburg, Germany

[‡]Department of Pathology, University of Virginia School of Medicine, Charlottesville, VA, USA

[§]Department of Immunobiology, Yale University School of Medicine, New Haven, CT, USA

[¶]Mali International Center for Excellence in Research, University of Sciences, Technique and Technology of Bamako, Bamako, Mali

Abstract

Immunity to *Plasmodium falciparum* malaria is naturally acquired in individuals living in malaria-endemic areas of Africa. Abs play a key role in mediating this immunity, however, the acquisition of the components of Ab immunity, long-lived plasma cells and memory B cells (MBCs), is remarkably inefficient, requiring years of malaria exposure. Although long-lived classical MBCs ($CD19^+/CD20^+/CD21^+/CD27^+/CD10^-$) are gradually acquired in response to natural infection, exposure to *P. falciparum* also results in a large expansion of what we have termed atypical MBCs ($CD19^+/CD20^+/CD21^-/CD27^-/CD10^-$). At present, the function of atypical MBCs in malaria is not known nor are the factors that drive their differentiation. To gain insight into the relationship between classical and atypical IgG^+MBCs we compared the Ab heavy and light chain variable (V) gene repertoires of children living in a malaria endemic region in Mali. We found that these repertoires were remarkably similar by a variety of criteria including V gene usage, rate of somatic hypermutation and CDR-H3 length and composition. The similarity in these repertoires suggests that classical MBCs and atypical MBCs differentiate in response to similar Ag-dependent selective pressures in malaria exposed children and that atypical MBCs do not express a unique V gene repertoire.

[1]This work was supported by the Intramural Research Program of the NIAID, NIH.

Correspondence: Dr. Severin Zinöcker (severin.zinocker@gmail.com) or Dr. Susan K. Pierce (spierce@nih.gov); Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Twinbrook II, 12441 Parklawn Dr, MSC 8180, Rockville, MD, 20852.

Introduction

Malaria is an infectious disease caused by mosquito-borne parasites of the genus *Plasmodium*, the most deadly species of which, *P. falciparum* (*Pf*), prevails in Africa. *Pf* malaria results in the deaths of nearly one million children each year in Africa alone (1) and at present there is no vaccine to combat malaria. Individuals who live in malaria endemic areas of Africa acquire resistance to clinical malaria but this process is remarkably slow requiring years of repeated exposure to *Pf* (2, 3). Although the age at which immunity is acquired varies depending on the intensity of transmission (4), in areas of intense malaria transmission children gradually become resistant to the most severe and lethal forms of malaria by the age of five or so, however, they remain susceptible to uncomplicated malaria until late childhood or early adolescence. Adults seldom experience clinical malaria symptoms but resistance to *Pf* infection is rarely, if ever, achieved (5).

Abs play a key role in naturally acquired immunity to malaria as demonstrated in passive IgG transfer studies (6) showing that Abs from malaria-resistant adults, when transferred to either children in Africa or to semi-immune adults in Thailand (7) with clinical malaria, reduced both the levels of parasitemia and fever. We do not yet know the nature or specificities of the Abs that confer protection to malaria nor do we understand why the acquisition of Ab immunity in malaria is so inefficient. Ab-mediated immunity requires the generation of both long-lived Ab-secreting plasma cells and memory B cells (MBCs). Several recent studies have provided evidence that *Pf*-specific MBCs are indeed acquired in response to natural infection but this acquisition is inefficient, reflecting the delayed acquisition of malaria immunity (8-13). For example, in a longitudinal study in Mali where malaria transmission is highly seasonal and intense, resulting in as many as 60 infectious mosquito bites *per* individual *per* month at the peak of the season, the frequency of MBCs to two malaria blood stage antigens, apical membrane antigen 1 (AMA-1) and merozoite surface protein 1 (MSP-1), increased only incrementally each year in children and reached adult levels only during adolescence (8). Moreover, the prevalence of AMA-1- and MSP-1-specific MBCs were relatively low among adults with specific MBCs only detectable in ∼30-50% of adults, consistent with the findings of others (9, 11, 12). Nonetheless, once acquired, *Pf*-specific MBCs appear to be long-lived and may persist longer than serum Abs in the absence of continued *Pf* exposure (10).

In addition to the acquisition of classical MBCs in malaria-exposed children and adults, a phenotypically distinct subset of MBCs are greatly expanded representing up to 30-40% of all circulating B cells (14). These MBCs are defined by their cell surface phenotype, $CD10^{-}$/$CD19^{+}$/$CD20^{+}$/$CD21^{-}$/$CD27^{-}$, and by the expression of several inhibitory receptors. Atypical MBCs are phenotypically similar to the previously described tissue-based MBCs in healthy individuals in the U.S. (15) and are also observed in the peripheral blood in individuals with chronic viral infections including, HIV (16), hepatitis C (17) and cytomegalovirus (18) but are rare in the circulation of healthy individuals living in malaria-free countries. In HIV-infected individuals atypical MBCs are hypo-responsive and have been suggested to contribute to the B cell deficiencies associated with HIV infections (16). In the context of malaria, we refer to this subset as atypical MBCs because we do not yet know if they play a beneficial or a detrimental role in this disease (14, 19).

Atypical MBCs have been described in the peripheral blood of children and adults who are exposed to *Pf* in a variety of geographically diverse malaria-endemic regions in Gabon (20), The Gambia (11), Ghana (21), Kenya (22), Mali (14, 23, 24) and Peru (23), as well in adults receiving experimental malaria infections in clinical trials (25). The number of atypical MBCs appears to correlate with the intensity of malaria transmission and declines in the absence of natural or experimental *Pf* exposure (25, 26). In addition, atypical MBCs tend to be more frequent in children who asymptomatically carry *Pf* as compared to children who are parasite-free (14). Together these findings suggest that *Pf* may drive the observed expansion of atypical MBCs. Of interest, Muellenbeck *et al*. (20) recently provided evidence that both classical and atypical MBCs in semi-immune adults living in Gabon express pairs of V region genes of the H chain (VH) and L chain (VL) that encode broadly neutralizing *Pf*-specific Abs and the authors speculated that atypical MBCs may expand due to chronic activation.

To gain insight into the relationship between classical and atypical MBCs we compared the expressed IgH and IgL chain V genes in single B cells from peripheral blood of children living in malaria-endemic Mali. We found that the classical and atypical MBC compartments have remarkably similar Ig V region repertoires and thus do not appear to be derived from different precursors or to be under different Ag-selective pressure in malaria.

Materials and Methods

Ethics statement

The Ethics Committee of the Faculté de Médecine, de Pharmacie et d'Odento-Stamologie at Université de Bamako, Mali, and the Institutional Review Board of the National Institute of Allergy and Infectious Diseases (NIAID) at the National Institutes of Health (NIH), MD, USA, approved this study. Written informed consent was obtained from the parents or guardians of children participating in the study.

Study cohort

All subjects in this study were participants in an observational cohort study in Kambila, Mali, a village in which *Pf* transmission occurs seasonally between July and December. Details about the study site, disease transmission and cohort design have been reported elsewhere (27). Briefly, 225 healthy children and young adults who were residents of a small ($\sim$1 km^2) well circumscribed rural village in Mali with a population of 1,500 were enrolled in a multi-year longitudinal observational study of the acquisition and maintenance of malaria immunity beginning in 2006. Case detection was passive although participants were encouraged to report symptoms of malaria at the village health center which was staffed 24 h a day by a study physician. The research definition of malaria was an auxiliary temperature of $\geq$ 37.5°C, *Pf* asexual parasitemia $\geq$ 5,000 parasites/red blood cell and no other obvious cause of fever. Malaria cases were treated with a regimen of artesunate and amodiaquine. The present study focused on the characterization of peripheral blood B cells obtained from three unrelated children 4-7 years old. Biometric data of the study participants are summarized in Table 1. Venous blood was drawn on multiple occasions between May 2006 and November 2008 on site. PBMCs were isolated and stored at −80°C following

standard protocols at the Mali International Center for Excellence in Research, Université de Bamako. PBMC samples were shipped frozen on dry ice to the NIAID where they were stored in liquid nitrogen until further processing.

Isolation of atypical and classical MBCs by single-cell FACS

PBMCs from several blood draws were pooled for each subject after thawing and washing in cold RPMI with decreasing concentrations of FBS (from 50% to 10% v/v) by repeated centrifugation (6-10 min at 400×*g*). PBMCs were enriched for B cells by negative selection using AutoMacs separation (Miltenyi Biotec, Bergisch Gladback, Germany) following the manufacturer's guidelines. The negative fraction contained 98 % $CD19^+$ cells by FACS and the positive fraction was largely devoid of $CD19^+$ cells. Cells were stained in PBS using appropriate concentrations of mouse anti-human CD10:APC (CB-CALLA; eBioscience, San Diego, CA), mouse anti-human CD19:PE-Cy5.5 (SJ25C1; eBioscience), mouse anti-human CD21:PE-Cy5 (B-ly4, BD Biosciences, Franklin Lakes, NJ), mouse anti-human CD27:Brilliant Violet 421™ (O323; BioLegend, San Diego, CA) and mouse anti-human IgG:FITC (G18-145; BD Biosciences) Abs and sorted by phenotype as either IgG-expressing conventional ($IgG^+/CD10^-/CD19^+/CD21^+/CD27^+$) or atypical ($IgG^+/CD10^-/CD19^+/CD21^-/CD27^-$) MBCs on a FACS Aria instrument (BD Biosciences). The strategy for gating classical and atypical MBCs is shown in Fig. 1. Cell purities of approximately 83-88% and 95-97% of the parent population were achieved for classical and atypical MBCs, respectively. Single cells were directly deposited into Hard-Shell® 96-well PCR plates WHT/CRL (Bio-Rad, Hercules, CA) without capture buffer and immediately frozen on dry ice.

RNA isolation and Ig-specific PCR amplification

mRNA was isolated from single cells and expressed Ig genes were amplified and sequenced as described by Tiller *et al*. (28) and Wardemann and Kofer (29). Cells were lysed directly in the PCR plates used for capture by adding 10.8 μL of lysis buffer containing:NP-40 Alternative (0.74% v/v final concentration; EMD Millipore, Darmstadt, Germany); random hexamer primers (2.7 μM final concentration; Applied Biosciences/Life Technologies, Carlsbad, CA); rRNasin ribonuclease inhibitor (0.74 U/μL final concentration; Promega, Madison, WI) and diethylpyrocarbonate (DEPC)-treated water (Quality Biological, Inc., Gaithersburg, MD). Cell lysates were incubated for 1 min at 65°C in 10.5 μL of RT-PCR buffer containing: dNTP (each at a 0.4 mM final concentration; Invitrogen/Life Technologies); DTT (7.0 μM final concentration; Invitrogen); rRNasin (0.6 U/μL final concentration); SuperScript III RT-polymerase (3.8 U/μL final concentration; Invitrogen); 5× First Strand buffer (Invitrogen) and DEPC water. RNA was transcribed to cDNA by RT-PCR at 42°C for 5 min, 25°C for 10 min, 50°C for 60 min, and 94°C for 5 min.

cDNA products from RT-PCR were used without purification as templates in three separate gene-specific nested PCR (96-well PCR plates from Eppendorf, Hamburg, Germany) to amplify human Ig heavy (IgH), kappa (Igκ) or lambda (Igλ) chain transcripts as described by Tiller *et al*. (28). Multiple Igκ- and Igλ-specific primers were used for nested PCR as described (29). For PCR amplification of the IgH locus only Igγ-specific primers were used (29). First-round PCR were performed in 40 μL total volume using 10× PCR Buffer

(Qiagen, Venlo, The Netherlands) containing: forward and reverse primer mix (each 0.16 μM final concentration; Integrated DNA Technologies, Coralville, IA; Eurofins/MWG Operon, Huntsville, AL or Invitrogen); dNTP (1 mM final concentration; Qiagen); HotStarTaq (0.02 U/μL final concentration; Qiagen), DEPC water and 3 μL of template cDNA at 94°C for 15 min; 50 cycles at: 94°C for 30 s, 58°C (IgH and Igκ) or 60°C (Igλ) for 30 s, 72°C for 55 s and 72°C for 10 min. The second PCR was prepared in 40 μL total volume using reagents at the same concentrations as for the first-round PCR except using second-round IgH-, Igκ- and Igλ-specific primers (29). First-round amplified PCR product (3.5 μL) was used as template. Second-round PCR products were analyzed by capillary gel electrophoresis using a 5k HT RNA/DNA chip (PerkinElmer, Waltham, MA) in a Caliper GX instrument (PerkinElmer) following the manufacturer's guidelines. PCR showing positive signals (bands at expected sizes of approximately 300-500bp) were selected for DNA sequencing.

DNA sequencing

Aliquots (0.5 μL) of second-round PCR products and specific primers (PCR1-H forward mix, 5′ pan-Vκ and 3′ Xho1-Cλ for IgH, Igκ and Igλ, respectively, at 6 μM final concentration) were added to MicroAmp® Optical 96-well PCR plates (P/N 4306737, Applied Biosystems/Life Technologies) in DEPC water. Samples were sequenced at the Genomics Unit at Rocky Mountain Laboratories, NIAID, MT, by Sanger sequencing. Sequencing reactions were prepared as recommended by Applied Biosystems for BigDye® Terminator v3.1 Cycle Sequencing Kit by adding 1 μL ABI BigDye® Terminator Ready Reaction Mix v3.1 (P/N 4336921; Applied Biosystems), 1.5 μL 5× ABI Sequencing Buffer (P/N 4336699), 1 μL betaine solution (P/N B0300; all from Sigma-Aldrich, St. Louis, MO) and 2.5 μL of water for a final volume of 10 μL. Sequencing was performed on either a Bio-Rad Tetrad2 (Bio-Rad) or ABI 9700 (Applied Biosystems) thermal cycler by cycling for 27 cycles at: 96°C for 10 s; 50°C for 5 s; 60°C for 4 min. Fluorescently-labeled extension products were purified following the BigDye® XTerminator™ Purification protocol and were processed on an ABI 3730xL DNA Analyzer (Applied Biosystems). Data was stored on the GeneSifter data management systems (Geospiza, Seattle, WA) and downloaded (http://ai-rmlgslep.niaid.nih.gov/) with restricted access through the NIH. Nucleotide sequence data were trimmed (Phred score > 20) (30) and sequences shorter than 50 nt were excluded from the analysis.

Calculation of the sequencing error

A subset of sequences was replicated by repeating the nested PCR protocol using the same aliquot of original template cDNA and Sanger sequencing as described above. Replicate nucleotide sequence reads were aligned and compared using ClustalW (https://www.ebi.ac.uk/Tools/msa/clustalw2/) after blunting sequence overhangs. Base calls that were divergent (non-identical) between sequence pairs over the length of the Ig V region were attributed to non-template incorporation of nucleotides by DNA Taq polymerase (Taq error) and were counted as sequencing errors.

An in-built function of the Immunoglobulin Analysis Tool (IgAT; release 1.14 downloaded from the Philipps-Universität, Marburg, Germany webserver at http://www.uni-marburg.de/

neonat/igat) (31) was used as an alternative method to calculate the sequencing error by aligning sequences derived from IgH PCR amplification to a 50-nt segment of the constant region of each of the 9 human IgH isotypes. Of 680 sequences aligned, we were unable to unambiguously annotate 197, of which were 79 classical MBC (11.6%) and 118 were atypical MBC (17.4%). We also noticed a recurring 12-nt sequence pattern that deviated from the known human IgH isotype sequences (Supplementary Table 1 and Supplementary Table 2), which we ascribe to a hitherto unknown allelic polymorphism. We excluded these sequences from the sequencing error calculation. 471 sequences aligned to the IgG isotype (69.2 %), 7 (1.0 %) were IgM, 5 (0.7 %) were IgA, and none were IgD or IgE (Supplementary Table 1 and Supplementary Table 2). We then used all 483 sequences that were unambiguously annotated to their respective isotype to calculate the average sequencing error rate.

DNA sequence data curation

Unintentional duplication of nucleotide sequences due to potential PCR cross-contamination may falsely appear as clonally related sequences. Such sequences were eliminated using the following criteria: (i) sequences occurred twice or more often on the same PCR plate; and (ii) sequences encoded the same V gene and J gene; and (iii) CDR3 predicted amino acid sequences were of equal length; and (iv) aligned nucleotide sequences were identical or nearly identical. Comparisons were done after sequence alignment using ClustalW (https://www.ebi.ac.uk/Tools/msa/clustalw2/). Out of a total of 3181 reads, 66 (2.1%) potential duplications that met the defined criteria were detected and removed.

Ig sequence annotation and data analysis

Single nucleotide sequence reads in FASTA format were submitted online and annotated to human Ig germline sequences stored in the international ImMunoGeneTics information system (IMGT)®(32) using the V-QUEST analysis tool version 3.3.0 (33, 34) (http://www.imgt.org/IMGT_vquest/share/textes/). Output data after IMGT/V-QUEST annotation were stored locally and further analyzed using in-house custom-made scripts in the Perl programming language and running a local Perl© version 5.16.3 (http://www.perl.org/) on a Windows©7 operating system. Annotations that had lower than 70% identity with the identified Ig germline nucleotide sequence or that were annotated by IMGT/V-QUEST with no identifiable rearrangement were excluded from further analysis.

All DNA sequence reads were compiled in multi-FASTA format sorted by research subject (KAM098, KAM111, KAM116), Ig locus (H, κ, λ), and cell type (classical, atypical) and were annotated to human Ig germline sequences using the online HighV-QUEST tool version 1.2.0 (35) (http://www.imgt.org/HighV-QUEST/index.action). Comma-separated value text files from the IMGT/HighV-QUEST output were then imported into IgAT (31) for further deconstruction. The IgAT does not compute sequence reads which do not match a known Ig germline sequence or are annotated as unproductive by IMGT/HighV-QUEST.

Quantification of antigen-driven selection

To assess to which extent somatic mutation patterns of expressed Ig genes were due to Ag-driven selection, we used Bayesian estimation of Antigen-driven SELectIoN (BASELINe)

version 1.3 (36). In brief, the BASELINe tool quantifies the frequency of replacement mutations relative to total mutations including synonymous mutations in both CDR and framework regions (FWR) of Ig germline-aligned sequences. It then uses a Bayesian model of a binomial likelihood function and β prior distribution to calculate a probability distribution function of replacement frequency, which is then transformed to a measure of selection strength (Σ) normalized for the germline sequence and a numerical convolution of multiple sequences into a single probability distribution function.

DNA sequence reads for input were transformed to BASELINe-compatible germline Ig-aligned sequences in IMGT/HighV-QUEST format with masked CDR3 using custom-made bioinformatical code at the Kleinstein lab, Yale School of Medicine (unpublished; Mohamed Uduman, personal communication). We applied the default settings of BASELINe (http://selection.med.yale.edu/baseline/) selecting focused statistics, human S5F model for somatic hypermutation (SHM) targeting and IMGT-compatible input format without CDR3.

Statistical Analysis

All data were exported to and analyzed with Prism version 6.01 (GraphPad Software, La Jolla, CA) where original graphs were plotted. Statistical analysis was performed using paired and unpaired Student's *T*-tests except for comparison of sequencing error data which showed non-normal distribution. Results from paired Student's *T*-tests for Figs. 5B and 6F were verified using JMP version 10.0.0, 64-bit addition (SAS, Cary, NC) to compute nested linear mixed models for split-plot data. *P* values smaller than 0.05 were deemed significant throughout.

Results

Sequencing Ig V gene regions from single MBCs

PBMCs were obtained from three children (Table 1) participating in a longitudinal study of the acquisition of immunity to malaria in Kambila, Mali as described earlier (27). Malaria transmission in Kambila is seasonal, occurring June through December, and intense with the entomological inoculation rate estimated to be approximately 50-60 infective bites *per* person *per* month at the peak of the transmission season (27). One child, KAM116, had sickle trait (HbAS) that has been demonstrated to be associated with a decreased risk of malaria (37). We recently showed in this cohort that HbAS is also associated with a delay in the median time to first malaria episode (27). However, we did not observe differences in the malaria experience between the children as each of the three children studied had at least four cases of clinical malaria during the period that peripheral blood samples were obtained. As the blood draws in children are limited to a volume of 5mL, four times during a calendar year, to obtain sufficient numbers of B cells for analyses PBMCs collected at several time points over three malaria transmission seasons (between May and November of 2006, 2007 and 2008) were pooled for each child. Single, class-switched IgG^+, classical and atypical MBCs were sorted into 96-well PCR plates using the gating strategy shown in Fig. 1. Cells were lysed and cDNAs were synthesized from cellular RNAs by RT-PCR using random hexamer primers and RT-DNA polymerase (29). V genes of the heavy (VH), kappa (Vκ) and lambda (Vλ) Ig loci were amplified from template cDNA in a separate multiplex, two-

round nested PCR using specific primer mixes (16). The nucleotide sequences were determined by Sanger sequencing.

The DNA sequences obtained were curated as summarized in Fig. 2. Also indicated are the data sets on which the results displayed in Figs. 3-7 were based. DNA sequences were quality trimmed using a Phred quality score of 20 or higher and sequences shorter than 50 nt after trimming were removed. A total of 3,181 DNA sequence reads were obtained (1576 for classical MBCs and 1605 for atypical MBCs) from the three children. The data were further curated by excluding potential PCR contaminations as detailed in Materials and Methods reducing the number of classical and atypical MBC sequences by 1.3% and 2.9%, respectively. The nucleotide sequences were annotated to known germline Ig genes using IMGT/V-QUEST resulting in a total of 2925 annotated sequences: 1448 for classical MBCs and 1477 for atypical MBCs, including 926 IgH, 1235 Igκ and 764 Igλ. Ig sequences that matched the annotated V gene with less than 70% nucleotide sequence identity and annotations that lacked an identifiable rearrangement were excluded from further analysis resulting in a decrease in classical and atypical MBC sequences of 3.5% and 0.8%, respectively. All sequences were also annotated by IMGT/HighV-QUEST to make the output compatible with IgAT. Ig sequences that were identified by IMGT/HighV-QUEST as unproductive due to stop codons or out-of-frame mutations, likely due to sequencing errors, were removed at this step reducing the number of classical and atypical MBC sequences by 4.9% and 5.8%, respectively leaving 1192 classical and 1228 atypical MBC sequences. The curation reduced the number of sequences similarly in the repertoires of the three children and for VH, Vκ and Vλ.

The rate at which DNA Taq polymerase incorporated random errors (the Taq error) was determined by alignment of all IgH sequence reads to a 50-nt section in the constant domain (C_H) 1 for all 9 known human IgH isotypes (Supplementary Table 1 and Supplementary Table 2) using the IgAT. The Taq error was 3.29 ± 8.15 nt (mean ± SD) *per* 1,000 nt for classical MBC-derived sequences ($n = 226$) and 2.42 ± 7.25 nt *per* 1,000 nt for those derived from atypical MBCs ($n = 257$) which were not significantly different from one another (Wilcoxon rank sum test: $p = 0.174$). The Taq error rate was also calculated by comparing the entire portion of overlapping V region sequences from FWR1 through CDR3 in a subset of independently replicated sequence pairs using ClustalW for alignment. The average Taq error rates were determined to be 5.51 ± 9.20 nt *per* 1000 nt for classical MBC ($n = 48$) and 9.09 ± 13.71 nt *per* 1000 nt for atypical MBCs ($n = 80$) which were not significantly different (Wilcoxon rank sum test: $p = 0.080$).

The use of V gene segments is similar in classical and atypical MBCs

The frequencies at which VH, Vκ and Vλ gene families were expressed in the classical and atypical MBCs repertoires were determined for each child and then averaged for the three children. Our analysis showed that the VH, Vκ and Vλ gene families were expressed at similar frequencies in atypical and classical MBCs (Fig. 3A) with no statistically significant differences (unpaired, Student's *T*-test after correcting for multiple comparisons). We detected no statistically significant differences in the use of *D* gene families of IgH chains between the two MBC populations (Fig. 3B). The reading frame usage of D_H genes was also

equivalent in the three individuals (Fig. 3C). JH, Jκ and Jλ gene families were also used at approximately the same frequencies (Fig. 3D). Also given are the frequencies of the expression of individual VH, Vκ, Vλ genes (Fig. 3E) and D genes (Fig. 3F) showing no significant differences in the frequency of their use between classical and atypical MBCs (unpaired, Student's *T*-test after correcting for multiple comparisons). We also determined the recombinations of V and J genes of the H, κ and λ loci and displayed these as Circos plots that summarize the occurrences of V-J gene rearrangements (Fig. 4). In each circular plot, the V-J pairings are represented as ribbons connecting a V gene segment with its J gene segment as rearranged in an expressed clonal Ig sequence, with the width of the ribbons corresponding to the relative frequency at which a particular V-J rearrangement was used in the respective MBC repertoire. V-J rearrangements were overall highly similar between atypical and classical MBCs.

The composition of the CDR-H3 of classical and atypical MBCs

To gain insight into the composition and characteristics of the Ag-binding regions of expressed Iggenes, we characterized the CDR3s of IgH, Igκ and Igλ. The contributions of the V, D and J gene segments as well as P nucleotides and N additions to the entire length of the CDR-H3, CDR-K3, CDR-L3 nucleotide sequences were analyzed for classical and atypical MBCs. The overall contributions of these elements to the CDR3s were not significantly different between classical and atypical MBCs by unpaired, Student's *T*-test after correcting for multiple comparisons (Fig. 5A). We also determined that the mean nucleotide lengths of the CDR3sofclassical and atypical MBCs were similar for the VH (45.99 nt *vs*. 47.82 nt), Vκ (26.18 nt *vs*. 26.25 nt) and Vλ (30.44 nt *vs*. 30.36 nt).

The lengths of the predicted amino acid sequences of the CDR-H3sanalyzed using the IgAT were similar for atypical and classical MBCs (Fig. 5B). The mean CDR-H3 length of 15.74 amino acids for atypical MBCs and of 15.16 amino acids for classical MBCs, were not significantly different (paired Student's *T*-test: $p = 0.544$, $n = 3$). A nested linear mixed model analysis of the same data also showed no significant differences ($p = 0.433$). A retrospective power analysis of that data in Fig. 5A and B indicated that with three children we had only a 10% chance of detecting significant differences in the mean lengths of the V gene segments if these existed. However, the mean lengths were so similar that even if a much larger sample size showed significant differences the biological meaning of such differences would be unclear.

We analyzed several biochemical features of the predicted amino acid sequences of the CDR-H3s of classical and atypical MBCs. The frequencies at which amino acids were used in the CDR-H3s were identical when displayed as averages over the entire length of the CDR-H3 (Fig. 6A). When comparing only sequences of the median CDR-H3 length of 15 amino acids (Fig. 5), we also observed similar amino acid usage within the CDR-H3s, (Fig. 6B). Analyses of the Shannon entropy as a measure of amino acid variability at a given position of aligned protein sequences (38), and Kabat-Wu variability, the number of different amino acids observed at a position divided by the frequency of the most common amino acid (39), confirmed the consistency in amino acid usage between classical MBCs and atypical MBCs in the CDR-H3 (Fig. 6C, D).

Lastly, the charge of the predicted amino acids of the CDR-H3s were analyzed for each child, and showed that most CDR-H3s were neutrally charged with the net charges ranging between -1.0 and +1.0 and were indistinguishable for classical and atypical MBCs (p = 0.281) (Fig. 6E). The Kyte-Doolittle values (40) showed an equal distribution of hydrophobicity in the CDR-H3s of classical and atypical MBCs (paired Student's T-test: p = 0.423, n = 3) (Fig. 6F). A nested linear mixed model analysis of the same data also showed no significant differences (p = 0.275).

The somatic hypermutation (SHM) rate of V genes expressed by classical and atypical MBCs

The relative frequency at which SHM occurred in each V region sequence was determined (Fig. 7) using IgAT which requires batch annotation of sequence reads by IMGT/HighV-QUEST as detailed in Materials and Methods. A total of 2557 annotated, unique sequences were analyzed (705 VH, 1193 Vκ, 659 Vλ, from 1253 classical and 1304 atypical MBCs) (Fig. 2). The rate of SHM in the VH, Vκ and Vλ was greater in CDRs than FWRs as expected (Fig. 7A). The averages of SHM rates for atypical and classical MBCs VH, Vκ and Vλ over the V gene region spanning CDR1to FWR3 were analyzed for each child (Fig. 7B). As shown, there was a trend toward a lower SHM rate in the atypical as compared to classical MBCs that reached significance only for Vλ. However, a retrospective power analyses indicated that with three subjects we only had a 30-50% chance of detecting significant differences if they existed. Thus, it might be anticipated that with an analysis of three to four additional subjects the trend may become statistically significant. We also counted the absolute number of mutations over the whole V region of IgH, Igκ and Igλ sequences (data not shown) and found that classical and atypical MBCs had similar numbers of SHMs *per* VH, Vκ and Vλ sequence. The mean of the SHM of three children was 16.94 for VH, 13.22 for Vκ and 16.03 for Vλ of classical MBCs compared to 14.92 for VH, 11.42 for Vκ and 13.69 for Vλ of a typical MBCs.

Bayesian estimation of Ag-driven SELectIoN (BASELINe) is an online tool that analyzes the observed frequencies of replacement mutations in Ig germline gene aligned sequences, compares them to the calculated expected frequencies of directed SHM in CDR and FWR, and quantifies the selection strength, Σ, exerted on the Ig as a measure of Ag-dependent selection during germinal center B cell maturation (36). An analysis of IgH, Igκ and Igλ sequences from atypical and classical MBCs showed that the FWR had on average consistently negative Σ values, indicating weak selection pressure, whereas the CDRs were on average close to neutral selections, that is; slightly positive or, less often, slightly negative Σ values (Fig. 7C). CDR- Σ values were always more positive than FWR Σ values. Classical and atypical MBC-derived VH, Vκ and Vλ sequences showed the same selection trends in the three individuals and their mean Σ values when comparing their CDR and FWR were not significantly different (Fig. 7C).

Discussion

Two distinctive features of the natural acquisition of malaria Ab immunity are the inefficiency with which long-lived plasma cells and classical MBCs are acquired, requiring

years of malaria exposure, and the early expansion of atypical MBCs in children (19). We carried out a detailed analysis of the V gene repertoires of classical and atypical MBCs to gain insight into the relationship between classical and atypical MBC populations and how the selective pressures exerted by *Pf* infections and malaria may have influenced their acquisition. We observed overall highly similar patterns of V gene usage, V-J rearrangements, and the extent, quality and location of SHMs and highly similar predicted amino acid sequences and biochemical profiles of the CDR-H3, suggesting similar developmental histories of the classical and atypical MBC populations.

The results presented here analyzing all V genes in malaria susceptible children showed that the classical and atypical MBC repertoires were highly similar, suggesting that these two populations are not under different Ag-driven selective pressures. However, in malaria, we do not yet have the tools to investigate the B cell repertoire that functions to control disease. The *Pf* genome encodes over 5,400 genes and we have yet to identify unambiguous targets of protective Abs. It is possible that differences in the classical and atypical MBC repertoires may become apparent when particular *Pf*-specific protective V gene repertoires are analyzed. It was recently reported that HIV-specific B cells in HIV viremic individuals were enriched in abnormal B cell subsets including atypical MBCs and activated MBCs (41). In contrast, tetanus-specific and influenza-specific B cells were enriched in the normal resting MBC subset. This finding raises the possibility that *Pf*-specific B cells may be enriched in the atypical MBC population, a possibility that will be of interest to investigate as probes become available for the isolation of *Pf*-specific B cells.

Muellenbeck *et al*. (20) recently reported Ig sequences derived from classical and atypical MBC from three semi-immune adults living in an area of high *Pf* transmission in Gabon. The authors characterized the V gene repertoires of classical and atypical MBCs specific for a current blood stage vaccine candidate, GMZ2, a fusion protein of an immunodominate repeat region of *Pf* glutamate-rich protein and a conserved domain of *Pf* merozoite surface protein 3. Of the 560 VH, 355 Vκ and 175 Vλ genes analyzed no major differences were found in the V segments usage for classical and atypical MBCs nor in the features of the CDRH3, similar to what we observed in our analysis of the total V gene repertories in non-immune Malian children. The numbers of SHM for classical versus atypical MBCs for VH (25 vs. 28), Vκ (17 vs. 21) and Vλ (20 vs 20) in these semi-immune Gabon adult V genes appeared somewhat higher than the number of SHM we observed for the VH, Vκ and Vλ of classical and atypical MBCs in children (mean SHM rates between 12.5 and 17.5 for VH, 10.5 and 13.8 for Vκ, and 12.3 and 16.8 for Vλ), a difference that may reflect the accumulation of SHMs in MBCs with age or with the acquisition of resistance to clinical malaria. Indeed, our preliminary analysis of Ig sequences from one 20 year old semi-immune adult from the Kambila cohort showed greater numbers of SHMs as compared to the children in that same cohort (VH, 21.0 vs.18.9, $p = 0.3973$; Vκ, 20.8 vs. 14.6, $p = 0.0012$; Vλ, 20.6 vs. 11.8, $p = 0.0027$; means compared by unpaired Student's *T*-test). Interestingly, in the Muellenbeck *et al*. study the atypical MBCs pooled from the three adult stended to have a greater number of SHMs as compared to classical MBCs in their VH and Vκ whereas the trend we observed appeared to be toward fewer SHMs in the atypical

MBCs. Because both studies had a sample size of three individuals these differences will need to be confirmed with larger numbers of samples.

Important unanswered questions remain concerning conventional and atypical MBCs including what drives the expansion of atypical MBCs in children living in malaria endemic areas and what function do atypical MBCs provide. Muellenbeck *et al*. (20) concluded that atypical MBCs contribute directly to the production of *Pf*-specific IgG. However, with only a limited amount of data available we may not yet want to exclude the possibility that atypical MBCs may also function in a regulatory capacity. Clearly, we need to better understand the biology and specificity of this greatly expanded MBC population in order to better understand their role in the immune response to *Pf* malaria.

The results presented here and those presented by Muellenbeck *et al*. (20) are to our knowledge the first V gene sequences of Africans from malaria endemic regions. It will be of interest to determine if the V gene repertoires of individuals living in malaria endemic areas are skewed as compared to those of individuals living without malaria and if the African V gene repertoires resemble non-African V gene repertoires. Several MBC V gene repertoires have been reported recently from U.S. individuals (42-47) and there appears to be some preferential V gene usage that differs from what we and Muellenbeck *et al*. (20) observed in children and adults living in malaria endemic Africa. However, with the limited data sets available no clear patterns have yet emerged.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to the Malian adults and children of Kambila for their gracious participation in this study. We also wish to thank Peter Crompton and Silvia Portugal at NIAID Laboratory of Immunogenetics, MD, for their thoughtful discussions of the biology and functions of classical and atypical MBCs. We would also like to thank Tom Moyer, Carol Henry, Cal Eigsti, David Stephany and Kevin Holmes at NIAID Flow Cytometry Core Facility, MD, for their help with cell sorting and expert advice; Kent Barbian, Jennifer Hashimoto and Michael Walker at NIAID Rocky Mountain Laboratories Genomics Unit, MT, for their help with Sanger sequencing, curating sequence data and expert advice; Colleen Bolin and Harold Wang for helping design Perl scripts for automated Ig sequence data analysis; Constantine Chrysostomou and George Georgiou at University of Texas at Austin, TX, for sharing code for automated human CDR3 analysis; Christian Busse at Max Planck Institute for Infection Biology, Germany, for helpful input and feedback on curating and analyzing sequence data; Mohamed Uduman and Steven Kleinstein at Yale School of Medicine, CT, for advice and help with sequence data formatting for BASELINe and Amie D. Moody and Tara Burke with the NIH Fellows Editorial Board for their editorial comments.

References

1. Murray CJ, Rosenfeld LC, Lim SS, Andrews KG, Foreman KJ, Haring D, Fullman N, Naghavi M, Lozano R, Lopez AD. Global malaria mortality between 1980 and 2010: a systematic analysis. Lancet. 2012; 379:413–431. [PubMed: 22305225]
2. Langhorne J, Ndungu FM, Sponaas AM, Marsh K. Immunity to malaria: more questions than answers. Nature immunology. 2008; 9:725–732. [PubMed: 18563083]
3. Crompton PD, Moebius J, Portugal S, Waisberg M, Hart G, Garver LS, Miller LH, Barillas-Mury C, Pierce SK. Malaria immunity in man and mosquito: insights into unsolved mysteries of a deadly infectious disease. Annu Rev Immunol. 2014; 32:157–187. [PubMed: 24655294]

4. Okiro EA, Al-Taiar A, Reyburn H, Idro R, Berkley JA, Snow RW. Age patterns of severe paediatric malaria and their relationship to Plasmodium falciparum transmission intensity. Malar J. 2009; 8:4. [PubMed: 19128453]
5. Tran TM, Li S, Doumbo S, Doumtabe D, Huang CY, Dia S, Bathily A, Sangala J, Kone Y, Traore A, Niangaly M, Dara C, Kayentao K, Ongoiba A, Doumbo OK, Traore B, Crompton PD. An Intensive Longitudinal Cohort Study of Malian Children and Adults Reveals No Evidence of Acquired Immunity to Plasmodium falciparum Infection. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America. 2013; 57:40–47. [PubMed: 23487390]
6. Cohen S, Mc GI, Carrington S. Gamma-globulin and acquired immunity to human malaria. Nature. 1961; 192:733–737. [PubMed: 13880318]
7. Bouharoun-Tayoun H, Attanath P, Sabchareon A, Chongsuphajaisiddhi T, Druilhe P. Antibodies that protect humans against Plasmodium falciparum blood stages do not on their own inhibit parasite growth and invasion in vitro, but act in cooperation with monocytes. J Exp Med. 1990; 172:1633–1641. [PubMed: 2258697]
8. Weiss GE, Traore B, Kayentao K, Ongoiba A, Doumbo S, Doumtabe D, Kone Y, Dia S, Guindo A, Traore A, Huang CY, Miura K, Mircetic M, Li S, Baughman A, Narum DL, Miller LH, Doumbo OK, Pierce SK, Crompton PD. The Plasmodium falciparum-specific human memory B cell compartment expands gradually with repeated malaria infections. PLoS Pathog. 2010; 6:e1000912. [PubMed: 20502681]
9. Wipasa J, Suphavilai C, Okell LC, Cook J, Corran PH, Thaikla K, Liewsaree W, Riley EM, Hafalla JC. Long-lived antibody and B Cell memory responses to the human malaria parasites, Plasmodium falciparum and Plasmodium vivax. PLoS Pathog. 2010; 6:e1000770. [PubMed: 20174609]
10. Ndungu FM, Olotu A, Mwacharo J, Nyonda M, Apfeld J, Mramba LK, Fegan GW, Bejon P, Marsh K. Memory B cells are a more reliable archive for historical antimalarial responses than plasma antibodies in no-longer exposed children. Proc Natl Acad Sci U S A. 2012; 109:8247–8252. [PubMed: 22566630]
11. Nogaro SI, Hafalla JC, Walther B, Remarque EJ, Tetteh KK, Conway DJ, Riley EM, Walther M. The breadth, but not the magnitude, of circulating memory B cell responses to P. falciparum increases with age/exposure in an area of low transmission. PLoS One. 2011; 6:e25582. [PubMed: 21991321]
12. Dorfman JR, Bejon P, Ndungu FM, Langhorne J, Kortok MM, Lowe BS, Mwangi TW, Williams TN, Marsh K. B cell memory to 3 Plasmodium falciparum blood-stage antigens in a malaria-endemic area. J Infect Dis. 2005; 191:1623–1630. [PubMed: 15838788]
13. Ndungu FM, Lundblom K, Rono J, Illingworth J, Eriksson S, Farnert A. Long-lived Plasmodium falciparum specific memory B cells in naturally exposed Swedish travelers. Eur J Immunol. 2013; 43:2919–2929. [PubMed: 23881859]
14. Weiss GE, Crompton PD, Li S, Walsh LA, Moir S, Traore B, Kayentao K, Ongoiba A, Doumbo OK, Pierce SK. Atypical memory B cells are greatly expanded in individuals living in a malaria-endemic area. J Immunol. 2009; 183:2176–2182. [PubMed: 19592645]
15. Ehrhardt GR, Hsu JT, Gartland L, Leu CM, Zhang S, Davis RS, Cooper MD. Expression of the immunoregulatory molecule FcRH4 defines a distinctive tissue-based population of memory B cells. J Exp Med. 2005; 202:783–791. [PubMed: 16157685]
16. Moir S, Ho J, Malaspina A, Wang W, DiPoto AC, O'Shea MA, Roby G, Kottilil S, Arthos J, Proschan MA, Chun TW, Fauci AS. Evidence for HIV-associated B cell exhaustion in a dysfunctional memory B cell compartment in HIV-infected viremic individuals. J Exp Med. 2008; 205:1797–1805. [PubMed: 18625747]
17. Doi H, Tanoue S, Kaplan DE. Peripheral CD27-CD21- B-cells represent an exhausted lymphocyte population in hepatitis C cirrhosis. Clin Immunol. 2014; 150:184–191. [PubMed: 24434272]
18. Dauby N, Kummert C, Lecomte S, Liesnard C, Delforge ML, Donner C, Marchant A. Primary human cytomegalovirus infection induces the expansion of virus-specific activated and atypical memory B cells. J Infect Dis. 2014; 210:1275–1285. [PubMed: 24795470]
19. Portugal S, Pierce SK, Crompton PD. Young lives lost as B cells falter: what we are learning about antibody responses in malaria. J Immunol. 2013; 190:3039–3046. [PubMed: 23526829]

20. Muellenbeck MF, Ueberheide B, Amulic B, Epp A, Fenyo D, Busse CE, Esen M, Theisen M, Mordmuller B, Wardemann H. Atypical and classical memory B cells produce Plasmodium falciparum neutralizing antibodies. J Exp Med. 2013; 210:389–399. [PubMed: 23319701]
21. Ampomah P, Stevenson L, Ofori MF, Barfod L, Hviid L. Kinetics of B cell responses to Plasmodium falciparum erythrocyte membrane protein 1 in Ghanaian women naturally exposed to malaria parasites. J Immunol. 2014; 192:5236–5244. [PubMed: 24760153]
22. Illingworth J, Butler NS, Roetynck S, Mwacharo J, Pierce SK, Bejon P, Crompton PD, Marsh K, Ndungu FM. Chronic exposure to Plasmodium falciparum is associated with phenotypic evidence of B and T cell exhaustion. J Immunol. 2013; 190:1038–1047. [PubMed: 23264654]
23. Weiss GE, Clark EH, Li S, Traore B, Kayentao K, Ongoiba A, Hernandez JN, Doumbo OK, Pierce SK, Branch OH, Crompton PD. A positive correlation between atypical memory B cells and Plasmodium falciparum transmission intensity in cross-sectional studies in Peru and Mali. PLoS One. 2011; 6:e15983. [PubMed: 21264245]
24. Portugal S, Doumtabe D, Traore B, Miller LH, Troye-Blomberg M, Doumbo OK, Dolo A, Pierce SK, Crompton PD. B cell analysis of ethnic groups in Mali with differential susceptibility to malaria. Malar J. 2012; 11:162–168. [PubMed: 22577737]
25. Scholzen A, Teirlinck AC, Bijker EM, Roestenberg M, Hermsen CC, Hoffman SL, Sauerwein RW. BAFF and BAFF receptor levels correlate with B cell subset activation and redistribution in controlled human malaria infection. J Immunol. 2014; 192:3719–3729. [PubMed: 24646735]
26. Ayieko C, Maue AC, Jura WG, Noland GS, Ayodo G, Rochford R, John CC. Changes in B Cell Populations and Merozoite Surface Protein-1-Specific Memory B Cell Responses after Prolonged Absence of Detectable Infection. PLoS One. 2013; 8:e67230. [PubMed: 23826242]
27. Crompton PD, Traore B, Kayentao K, Doumbo S, Ongoiba A, Diakite SA, Krause MA, Doumtabe D, Kone Y, Weiss G, Huang CY, Doumbia S, Guindo A, Fairhurst RM, Miller LH, Pierce SK, Doumbo OK. Sickle cell trait is associated with a delayed onset of malaria: implications for time-to-event analysis in clinical studies of malaria. J Infect Dis. 2008; 198:1265–1275. [PubMed: 18752444]
28. Tiller T, Meffre E, Yurasov S, Tsuiji M, Nussenzweig MC, Wardemann H. Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. Journal of immunological methods. 2008; 329:112–124. [PubMed: 17996249]
29. Wardemann H, Kofer J. Expression cloning of human B cell immunoglobulins. Methods Mol Biol. 2013; 971:93–111. [PubMed: 23296959]
30. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 1998; 8:175–185. [PubMed: 9521921]
31. Rogosch T, Kerzel S, Hoi KH, Zhang Z, Maier RF, Ippolito GC, Zemlin M. Immunoglobulin analysis tool: a novel tool for the analysis of human and mouse heavy and light chain transcripts. Frontiers in immunology. 2012; 3:176. [PubMed: 22754554]
32. Lefranc MP. Immunoglobulin and T Cell Receptor Genes: IMGT ((R)) and the Birth and Rise of Immunoinformatics. Frontiers in immunology. 2014; 5:22. [PubMed: 24600447]
33. 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 Acids Res. 2008; 36:W503–508. [PubMed: 18503082]
34. 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 Harbor protocols. 2011; 2011:695–715. [PubMed: 21632778]
35. Alamyar E, Duroux P, Lefranc MP, Giudicelli V. IMGT ((R)) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. Methods Mol Biol. 2012; 882:569–604. [PubMed: 22665256]
36. Yaari G, Uduman M, Kleinstein SH. Quantifying selection in high-throughput Immunoglobulin sequencing data sets. Nucleic Acids Res. 2012; 40:e134. [PubMed: 22641856]
37. Williams TN, Mwangi TW, Wambua S, Alexander ND, Kortok M, Snow RW, Marsh K. Sickle cell trait and the risk of Plasmodium falciparum malaria and other childhood diseases. J Infect Dis. 2005; 192:178–186. [PubMed: 15942909]

38. Shannon CE. The mathematical theory of communication. 1963. MD computing: computers in medical practice. 1997; 14:306–317.
39. Wu TT, Johnson G, Kabat EA. Length distribution of CDRH3 in antibodies. Proteins. 1993; 16:1–7. [PubMed: 8497480]
40. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982; 157:105–132. [PubMed: 7108955]
41. Kardava L, Moir S, Shah N, Wang W, Wilson R, Buckner CM, Santich BH, Kim LJ, Spurlin EE, Nelson AK, Wheatley AK, Harvey CJ, McDermott AB, Wucherpfennig KW, Chun TW, Tsang JS, Li Y, Fauci AS. Abnormal B cell memory subsets dominate HIV-specific responses in infected individuals. J Clin Invest. 2014; 124:3252–3262. [PubMed: 24892810]
42. Briney BS, Willis JR, McKinney BA, Crowe JE Jr. High-throughput antibody sequencing reveals genetic evidence of global regulation of the naive and memory repertoires that extends across individuals. Genes Immun. 2012; 13:469–473. [PubMed: 22622198]
43. Mroczek ES, Ippolito GC, Rogosch T, Hoi KH, Hwangpo TA, Brand MG, Zhuang Y, Liu CR, Schneider DA, Zemlin M, Brown EE, Georgiou G, Schroeder HW Jr. Differences in the composition of the human antibody repertoire by B cell subsets in the blood. Frontiers in immunology. 2014; 5:96. [PubMed: 24678310]
44. Wu YC, Kipling D, Leong HS, Martin V, Ademokun AA, Dunn-Walters DK. High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. Blood. 2010; 116:1070–1078. [PubMed: 20457872]
45. Tian C, Luskin GK, Dischert KM, Higginbotham JN, Shepherd BE, Crowe JE Jr. Evidence for preferential Ig gene usage and differential TdT and exonuclease activities in human naive and memory B cells. Mol Immunol. 2007; 44:2173–2183. [PubMed: 17196657]
46. Arnaout R, Lee W, Cahill P, Honan T, Sparrow T, Weiand M, Nusbaum C, Rajewsky K, Koralov SB. High-resolution description of antibody heavy-chain repertoires in humans. PLoS One. 2011; 6:e22365. [PubMed: 21829618]
47. Boyd SD, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, Jones CD, Simen BB, Hanczaruk B, Nguyen KD, Nadeau KC, Egholm M, Miklos DB, Zehnder JL, Fire AZ. Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. Science translational medicine. 2009; 1:12ra23.

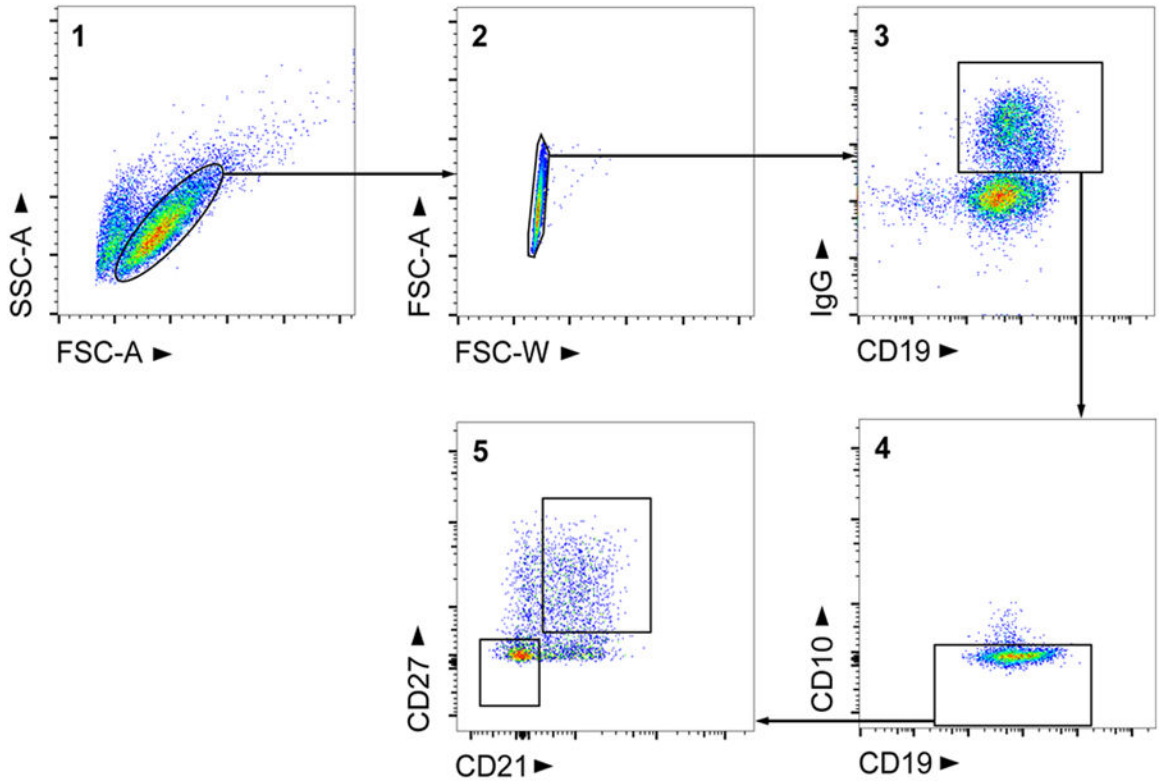

Figure 1. FACS gating to obtain peripheral blood classical and atypical MBCs

PBMCs were enriched for B cells by negative selection of non-B cells using magnetic beads and single cells of either classical or atypical MBC phenotype were deposited into wells of PCR plates by FACS. The sequential gating strategy of (**1**) lymphocytes, (**2**) single cells, (**3**) $IgG^+/CD19^+$ B cells, (**4**) $CD10^-$ mature B cells, (**5**) $CD21^+/CD27^+$ classical MBCs and $CD21^-/CD27^-$ atypical MBCs is shown on a representative plot. Events are displayed on linear, logarithmic or biexponential scales as appropriate.

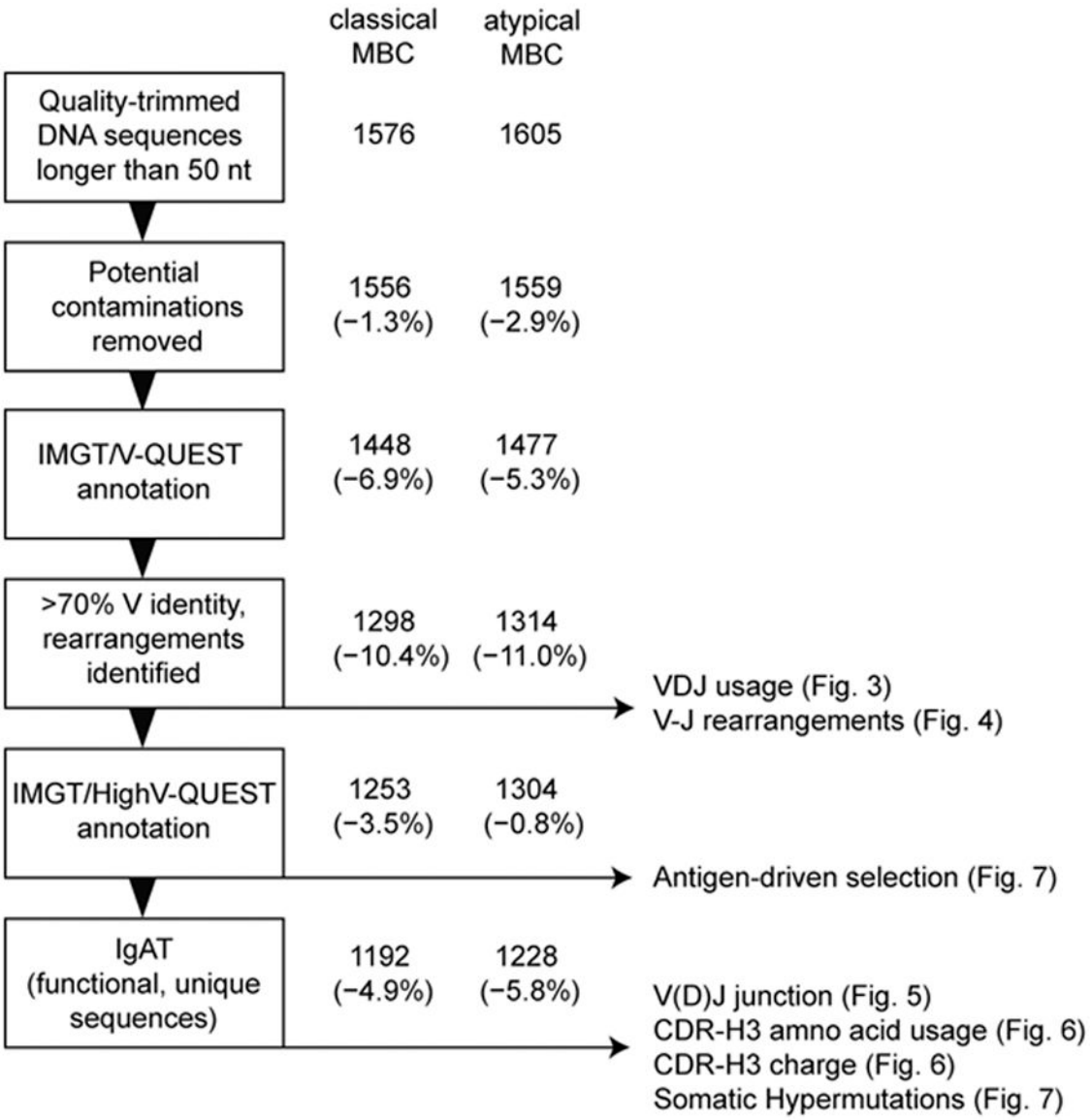

Figure 2. Curation of V gene sequence data

Depicted is the curation of nucleotide sequences obtained from classical and atypical MBCs. Shown are the number of sequences in each data set at the respective stage of the curation process, with percentages in parentheses showing the change in number of sequences from the previous step. Given on the right are the figures in which the results of the various analyses are presented.

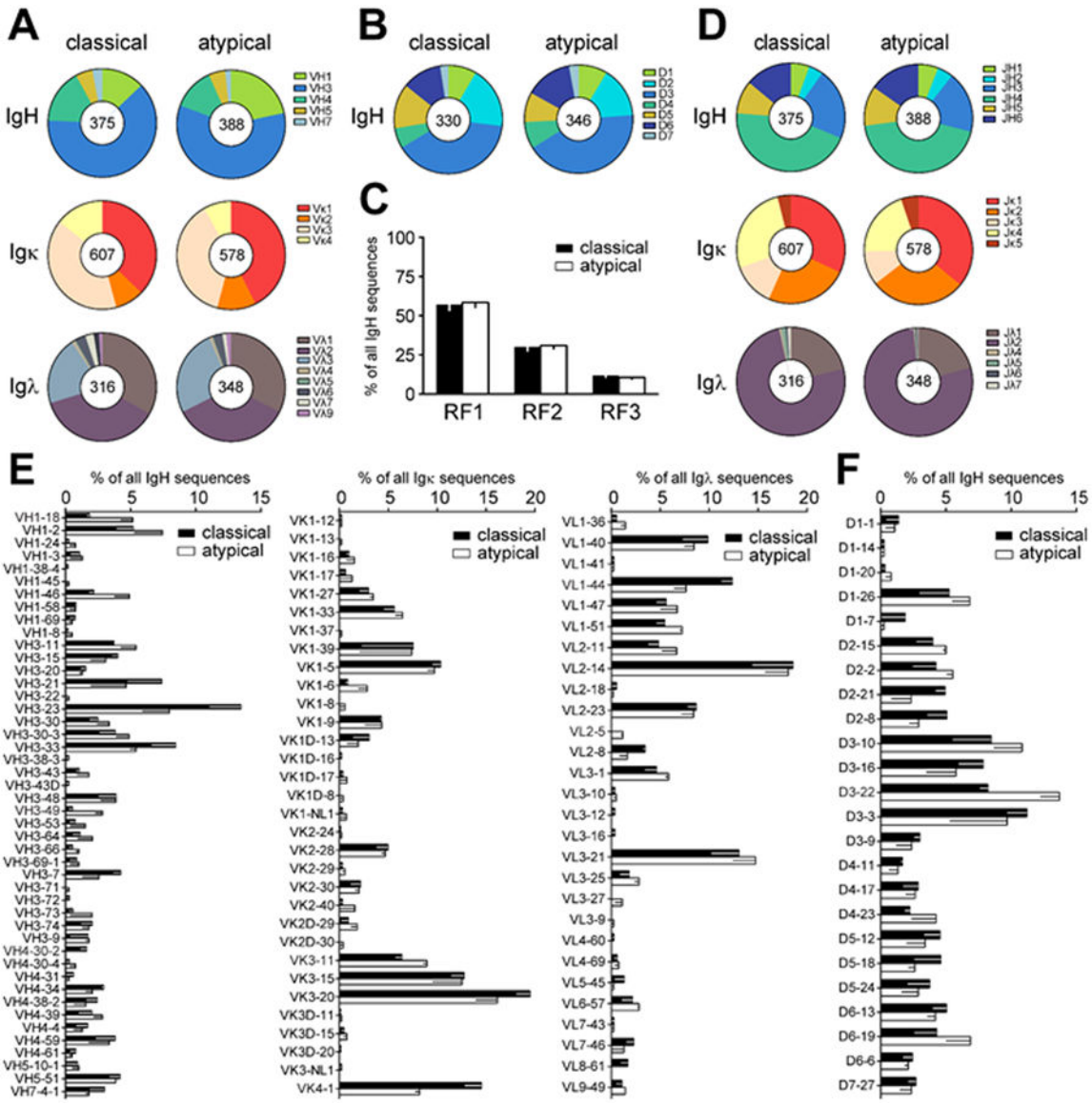

Figure 3. Classical and atypical MBCs express members of the variable (V), diversity (D), and joining (J) gene families at similar frequencies

(**A**) The frequencies at which the Ig heavy (IgH), kappa (Igκ) and lambda (Igλ) V gene families are expressed by classical MBCs and atypical MBCs are shown in pie charts as the mean percentages of gene sequences obtained from the three children. The total number of all analyzed sequences is given in the center of each plot. (**B**) The mean frequencies at which members of the D_H gene families are expressed for IgH and (**C**). the reading frame usage in the VDJ junction are shown for classical MBCs and atypical MBCs. (**D**) The mean frequencies at which the J gene families are expressed for IgH, Igκ and Igλ are given as in (**A**). Note that in a number of IgH sequences the D_H gene was not uniquely identified, resulting in lower sample sizes compared to those of V and J genes. The mean frequency of individual V_H, Vκ, and Vλ (**E**) and D_H (**F**) gene segments in classical MBCs and atypical MBCs are shown as a percentage of all annotated genes. Error bars represent the negative standard error of the mean (n = 3).

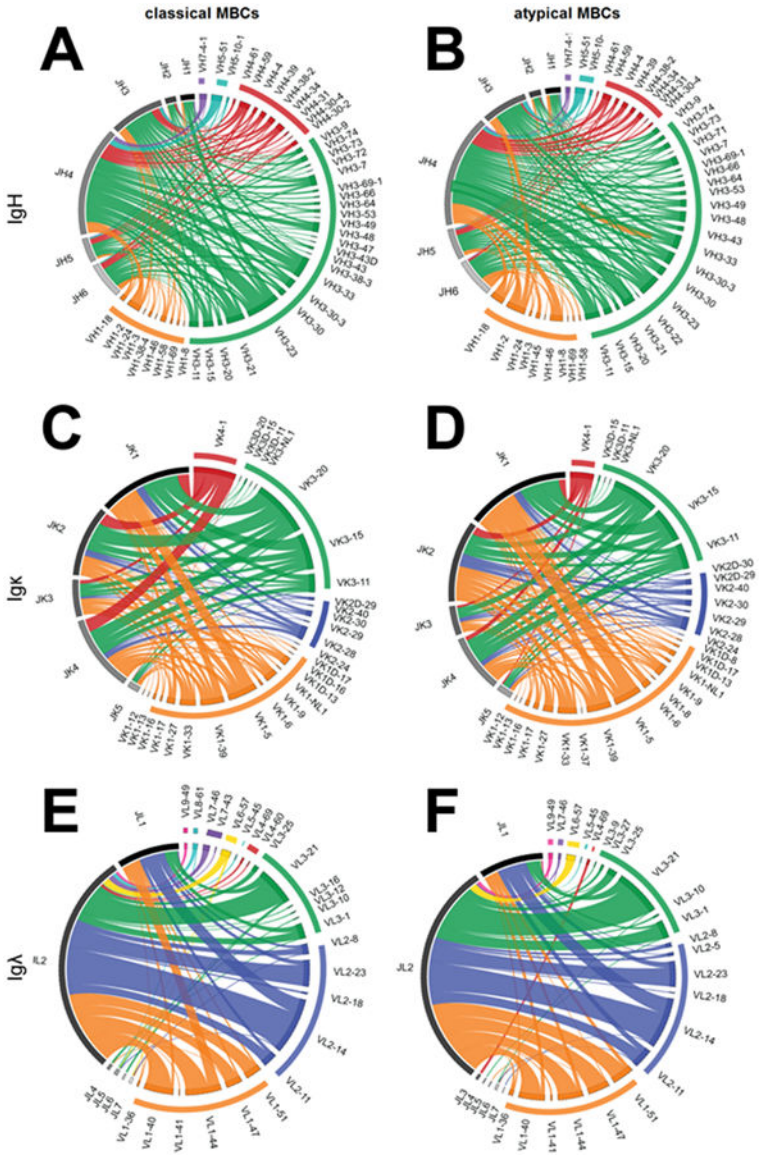

Figure 4. Rearrangements between individual V and J genes occur at similar frequencies in classical and atypical MBCs

Circos plots summarize the combinations of V and J genes used in the rearranged Ig genes expressed by classical MBCs and atypical MBCs for IgH (**A** and **B**), Igκ (**C** and **D**) and Igλ (**E** and **F**). For each plot, the right half depicts each segment of a V gene family (colored) and on the left side the J genes are shown (gray shades). The arc length of each segment denotes the relative frequency at which each gene segment was identified. Rearrangement of a J gene with a V gene segment in a clonal Ig sequence is represented by a ribbon (ribbons carry the color of the V family of the gene participating in the pairing). The width of the ribbons corresponds to the frequency at which each particular V-J rearrangement is used in the respective MBC repertoire.

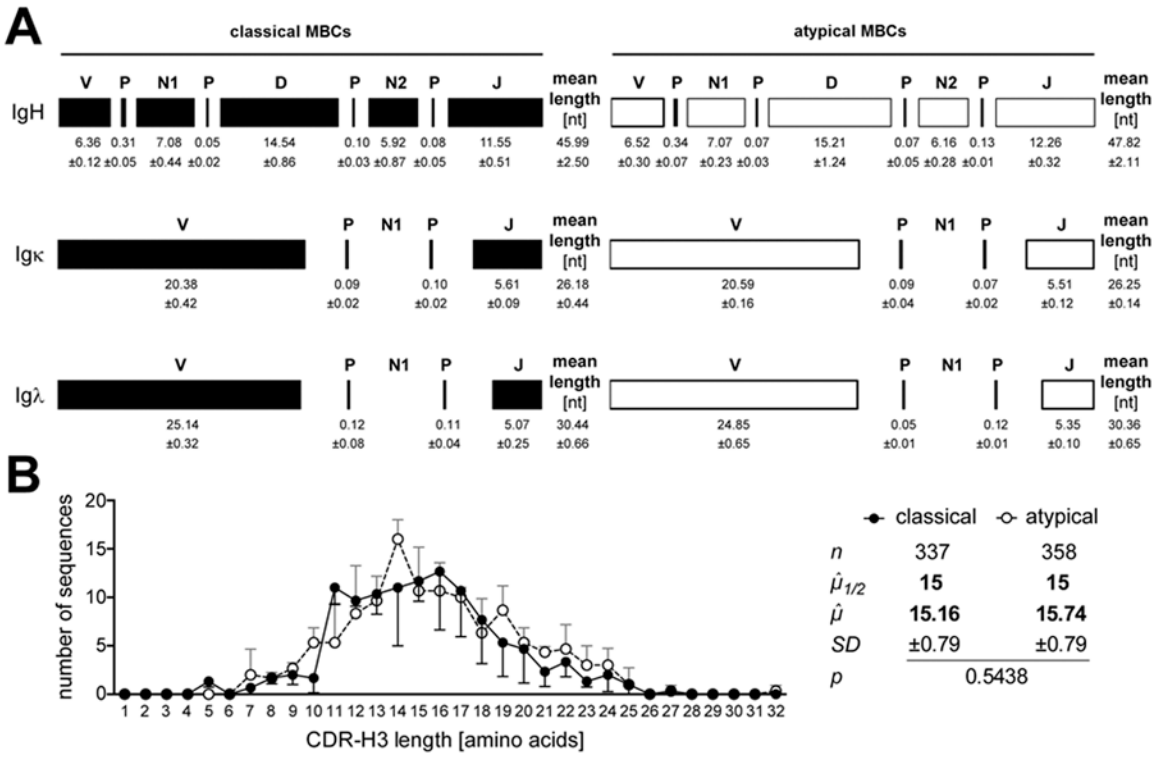

Figure 5. VDJ and VJ junctions in classical and atypical MBCs are similar in length and genetic makeup

(**A**) The contributions of nucleotides provided by V, D and J gene segments to the full length of the CDR3 of productively rearranged IgH, Igκ and Igλ sequences as well as by P junctions and N additions at the junctions are shown. The lengths of the components of CDR-H3 (top), CDR-K3 (middle) and CDR-L3 (bottom) nucleotide sequences were determined for classical and atypical MBCs. The number of nucleotides that V, P (3′ of V), N1, P (5′ of J), J genes, and, in case of IgH sequences, D genes and P (5′ and 3′ of D) contributed to its CDR3 is also given. The mean length of each component is based on the overall averages of all sequences analyzed for each of the three children ($n = 3$) with the SD are also given. Only CDR3 segments of productive, rearranged Ig sequences with an identifiable V, J, and, in case of heavy chain sequences, a D gene segment were analyzed. (**B**) The mean number of predicted amino acids in the CDR-H3 of classical MBCs (filled circles) and atypical MBC (open circles, dashed line) Ig sequences are given. The means and SDs are based on the averages of three individuals. Unpaired comparisons of the means of sequences with the same CDR-H3 length were completed. For comparisons of the two data sets n represents the total number of sequences analyzed for all three subjects, $\hat{\mu}_{1/2}$ the calculated average of the median and $\hat{\mu}$ the average of the mean CDR-H3 lengths. The resulting p value for comparisons of the means are also given.

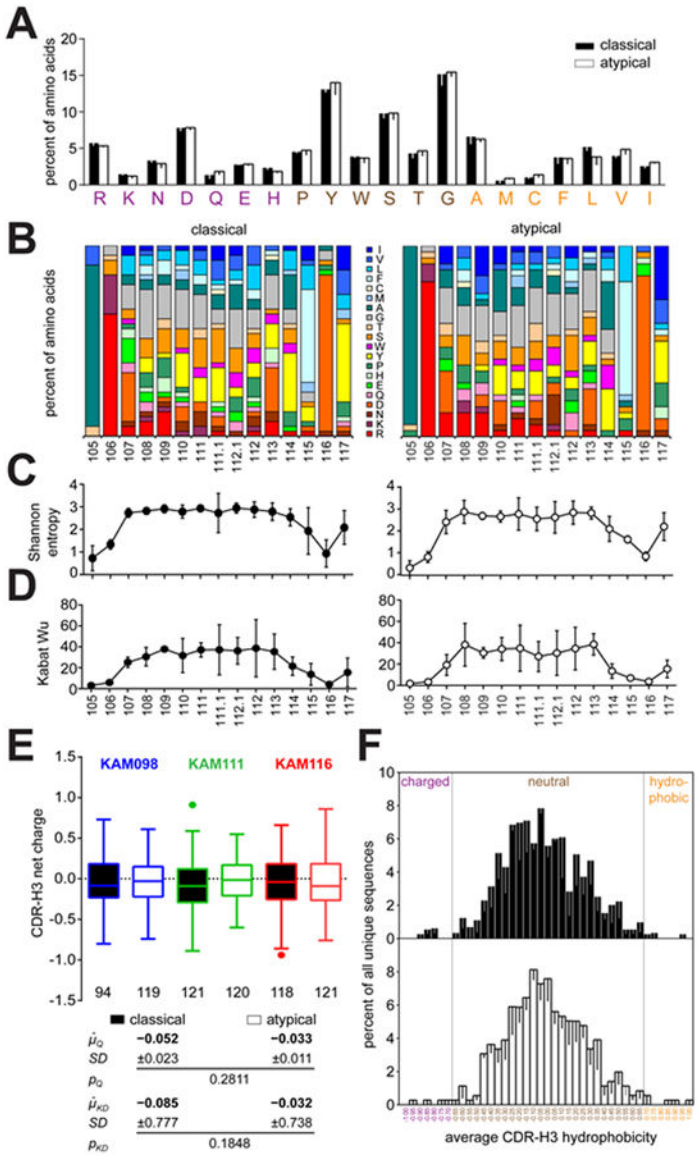

Figure 6. Biochemical characteristics of the predicted amino acid sequences of CDR-H3 expressed in classical and atypical MBCs

(**A**) The relative occurrences of the predicted amino acids abbreviated using standard one letter code in the CDR-H3s of unique IgH sequences expressed by classical (filled bars) and atypical (open bars) MBCs are shown. Amino acids are color-coded: charged (purple); neutral (brown) and hydrophobic (orange). Bars show the mean frequencies of amino acids as the percentage of all amino acids used in the CDR-H3 and error bars represent the negative SD (n = 3). (**B**) Predicted amino acid sequences of CDR-H3s of median length (15 amino acids Fig. 3) from classical (n = 35) and atypical (n = 32) MBCs, were selected to display graphically the relative occurrence of amino acids at each position (following the IMGT unique numbering system) (32) in the CDR-H3predicted amino acid sequence. Shannon entropy (**C**) and Kabat-Wu variability (**D**) numerically estimate the diversity/ variability of amino acids (38, 39) at CDR-H3 positions. Sequences of median CDR-H3 length were compared. (**E**) The average net charge over the entire length of the CDR-H3 of

unique sequences was calculated and classical MBCs (black boxes) and atypical MBCs (open boxes) were compared for each child. The data are displayed as Tukey plots with outliers (numbers under the boxes denote the number of total analyzed sequences). The mean net charge ($\hat{\mu}_Q$) and SD as well as the result of an unpaired, Student's *T*-test comparison of classical and atypical MBCs are shown. (**F**) The Kyte-Doolittle index is a measure of the hydrophobic character of proteins (40). The average hydrophobicity of sequences with a unique CDR-H3 was calculated for each of the three children. The overall means and SD are depicted as bars (classical MBCs, filled; atypical MBCs, open) with negative error bars.

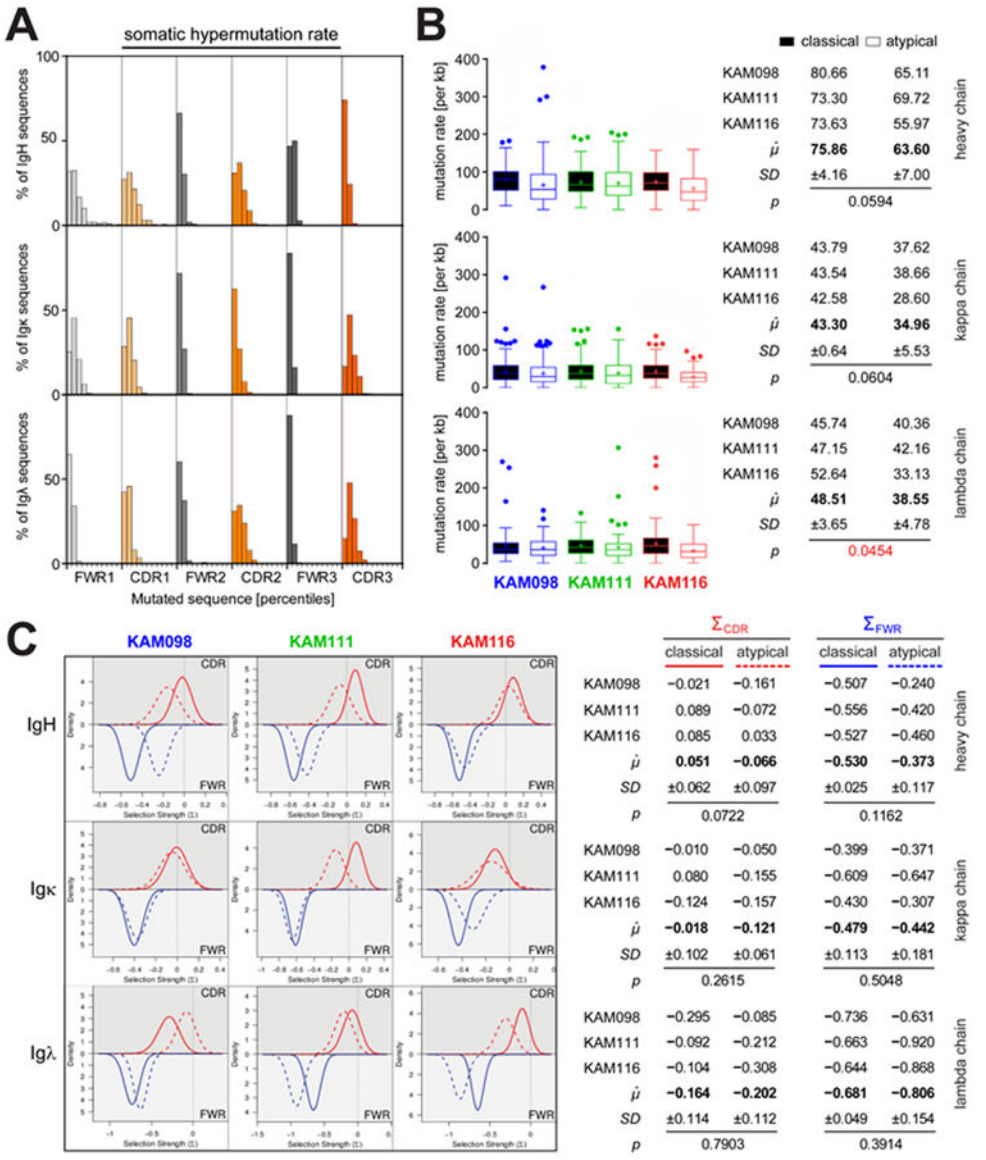

Figure 7. Somatic hypermutation and antigen-driven selection of classical and atypical MBCs

(**A**) The extent of SHM observed in FWR and CDR of all IgH (top), Igκ (middle) and Igλ (bottom) sequences are plotted as histograms depicting percentiles of mutated nucleotide sequence (from 0% to 100% in 10% increments) on the *x*-axis. The percent of total sequences that fall in each percentile is plotted on the *y*-axis. (**B**) The Immunoglobulin Analysis Tool (IgAT) (31) which takes into account SHM in CDR1 through FWR3 (indicated) following IMGT annotation (32) was used to calculate the SHM rate (the number of nucleotides divergent from the germline sequences *per* 1000 nt). Based on our calculations, the average Taq error rate contributed less than 10 mutations *per* 1,000 nt to the SHM rate. The SHM rates for each subject (KAM098, blue; KAM111, green, KAM116, red) for IgH, Igκ and Igλ are shown. Average mutation rates are shown as Tukey plots with outliers (left panel) and the mean values (+) are summarized together with the overall mean ($\hat{\mu}$) and SD for all subjects (n = 3, right). (**C**) Bayesian estimation of Ag-driven SELectIoN

(BASELINe) (36) is a statistical tool that allows the quantification of Ag-dependent B cell selection of large Ig sequence data sets. It calculates the probability distribution frequency as a measure of the likelihood that mutations of the germline IgH (left panel, top), Igκ (middle) and Igλ (bottom) locus are due to selection (positive or negative) in the CDR (red lines) and FWR (blue lines), respectively. We compared classical (solid lines) and atypical (dashed lines) MBC-derived Ig sequences of each of the three children and summarize the quantified average selection strength (Σ) as well as their means with SD (n = 3) and the results of paired Student's T-test comparisons (right panel).

Table 1

Demographic and clinical data of study participants

Subject ID	Gender	Ethnicity	Age at blood draws[1] [years]	Blood type, Rhesus factor	Hemoglobin genotype
KAM098	female	Sarakole	4-6	0+	AA
KAM111	male	Sarakole	5-7	A+	AA
KAM116	female	Sarakole	4-6	A+	AS

Subject ID	Malaria incidence[2]			classical MBC	atypical MBC
	2006	2007	2008	[% of B cells][3]	[% of B cells][3]
KAM098	2	5	2	25.5	23.0
KAM111	1	1	2	21.7	10.4
KAM116	3	2	0	19.9	11.2

[1] PBMC samples from several blood donations during May 2006 through November 2008 were pooled.

[2] A clinical malaria episode was defined as defined as a body temperature ≥ 37.5°C and blood parasitemia> 2500 *per* μL.

[3] B cells are peripheral blood $CD10^{-}/CD19^{+}$ lymphocytes.