Lymphoma-associated mutations in autoreactive memory B cells of patients with Sjögren's syndrome

Bende, Richard J.; Slot, Linda M.; Kwakkenbos, Mark J.; Wormhoudt, Thera A.M.; Jongejan, Aldo; Verstappen, Gwenny M.; van Kampen, Antoine C.M.; Guikema, Jeroen E.J.; Kroese, Frans G.M.; van Noesel, Carel J.M.

Published in:
Journal of Pathology

DOI:
10.1002/path.6039

Journal of Pathology
J Pathol March 2023; **259:** 264–275
Published online 12 January 2023 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/path.6039

ORIGINAL ARTICLE

Lymphoma-associated mutations in autoreactive memory B cells of patients with Sjögren's syndrome

Richard J Bende[1,2,3*†], Linda M Slot[1,2,3†], Mark J Kwakkenbos[4], Thera AM Wormhoudt[1,2,3], Aldo Jongejan[5], Gwenny M Verstappen[6], Antoine CM van Kampen[5,7], Jeroen EJ Guikema[1,2,3], Frans GM Kroese[6] and Carel JM van Noesel[1,2,3*]

[1] *Department of Pathology, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands*
[2] *Lymphoma and Myeloma Center (LYMMCARE), Amsterdam, The Netherlands*
[3] *Cancer Center Amsterdam (CCA), Amsterdam, The Netherlands*
[4] *AIMM Therapeutics, Amsterdam, The Netherlands*
[5] *Bioinformatics Laboratory, Epidemiology & Data Science, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands*
[6] *Department of Rheumatology and Clinical Immunology, UMC Groningen, University of Groningen, Groningen, The Netherlands*
[7] *Biosystems Data analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands*

**Correspondence to: CJM van Noesel or RJ Bende, Department of Pathology, Amsterdam University Medical Centers, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. E-mail: c.j.vannoesel@amsterdamumc.nl or r.j.bende@amsterdamumc.nl*

†These authors contributed equally to this work.

Abstract

We recently demonstrated that normal memory B lymphocytes carry a substantial number of *de novo* mutations in the genome. Here, we performed exome-wide somatic mutation analyses of bona fide autoreactive rheumatoid factor (RF)-expressing memory B cells retrieved from patients with Sjögren's syndrome (SS). The amount and repertoire of the *de novo* exome mutations of RF B cells were found to be essentially different from those detected in healthy donor memory B cells. In contrast to the mutation spectra of normal B cells, which appeared random and non-selected, the mutations of the RF B cells were greater in number and enriched for mutations in genes also found mutated in B-cell non-Hodgkin lymphomas. During the study, one of the SS patients developed a diffuse large B-cell lymphoma (DLBCL) out of an RF clone that was identified 2 years earlier in an inflamed salivary gland biopsy. The successive oncogenic events in the RF precursor clone and the DLBCL were assessed. In conclusion, our findings of enhanced and selected genomic damage in growth-regulating genes in RF memory B cells of SS patients together with the documented transformation of an RF-precursor clone into DLBCL provide unique novel insight into the earliest stages of B-cell derailment and lymphomagenesis.

Keywords: blood; B cell lymphoma; Sjogren's syndrome; B Lymphocyte; MALT

Received 14 June 2022; Revised 1 November 2022; Accepted 23 November 2022

No conflicts of interest were declared.

Introduction

Tumorigenesis is a multi-step process based on the combination of introduction of random gene alterations and selection of clones harboring mutations that provide growth advantage. In the past decade, it has become clear that all normal cells of our body accumulate somatic gene alterations over our lifetime [1–4]. Due to the astronomical size of the genome and the fact that only 1% of the nucleotides are protein-encoding, the vast majority of the acquired base substitutions are without functional consequence. However, once a driver mutation has been introduced and/or the frequency of mutations introduced per cell division becomes orders of magnitudes greater due to inactivation of gatekeeper genes, like, for example, DNA repair genes, the process of clonal competition and tumor evolution accelerates and finally becomes unleashed [5].

We have previously demonstrated that normal tetanus toxoid (TT)-specific memory B lymphocytes from healthy donors harbor acquired mutations beyond the immunoglobulin loci [6]. In B-cell clones expressing immunoglobulin with different affinities for TT, we observed a statistically significant correlation between the numbers of *de novo* mutations present in the exome and those in the immunoglobulin heavy variable regions (IGHV), respectively. This finding, which was recently confirmed by others [7], suggested that the majority of

the genomic mutations arise in an antigen-dependent fashion most likely during clonal expansion in germinal centers. The non-IG mutations seemed randomly distributed without evidence for a selection bias and in the majority did not display hallmarks of activity of the B-cell-specific mutator activation-induced cytidine deaminase (AID) [8]. By extrapolation of our findings, we calculated the number of genome-wide somatic mutations in memory B cells in adults to range between 500 and 3,000 per cell. This number is concordant with a recent single-cell whole-genome sequencing study, which revealed that the average numbers of genomic mutations per B cell, in newborns and centenarians, are <500 and >3,000, respectively [9].

We and others have previously shown that tumor cells of a number of low-grade B-cell lymphoma entities, in particular salivary gland mucosa-associated lymphoid tissue (MALT) lymphoma, frequently express autoreactive B-cell receptors (BCRs) with high-affinity rheumatoid factor (RF) activity (binding to IgG-Fc) [10,11]. At least 75% of salivary gland MALT lymphomas that occur in patients with Sjögren's syndrome (SS) were found to express RFs [12]. The majority of RFs expressed by salivary gland MALT lymphomas are encoded by recurrent IGHV/IGKV rearrangements with stereotypic VH-CDR3 structures. Stereotyped RF BCRs are also frequently expressed by *Helicobacter pylori* (*Hp*)-associated gastric MALT lymphoma and hepatitis C virus (HCV)-related B-cell lymphoma, and more rarely by ocular adnexal MALT lymphoma, splenic marginal zone B-cell lymphoma, diffuse large B-cell lymphoma (DLBCL), and chronic lymphocytic leukemia (CLL) [10,13–17].

As the majority of MALT lymphomas in SS patients express RFs, we set out to explore the non-IG gene somatic mutation profiles of RF memory B cells retrieved from SS patients.

Materials and methods

Patient material

Peripheral blood (PB) mononuclear cells (PBMCs) were obtained from six SS patients: SG4, SG12, SG13, SG20, SG22, and SG33. A cell suspension of a parotid salivary gland biopsy was obtained from one SS patient, SG39. The salivary gland focus scores, serological data, ages, and follow-up of the seven SS patients are provided in supplementary material, Table S1. The control donors, TT3 and TT4, from whom we isolated tetanus toxoid (TT)-specific normal memory B cells and which were analyzed by whole-exome sequencing, have been described previously [6]. The follow-up of the SS patients was >10 years for SG12, SG13, SG20, SG22, and SG33; >6 years for SG4; and >4 years for SG39 (supplementary material, Table S1).

The samples were from the Department of Pathology, Amsterdam University Medical Centers and from the Department of Rheumatology and Clinical Immunology, University Medical Center Groningen, The Netherlands. The study was conducted in accordance with the ethical standards of our institutional medical ethics committee on human experimentation, as well as in agreement with the 1975 Declaration of Helsinki, as revised in 1983.

Isolation of RF-expressing B-cell clones

Cell suspensions were incubated with antibodies for CD19-APC (Becton Dickinson, Franklin Lakes, NJ, USA; #555415, 1:50), CD27-PE (Becton Dickinson, #555441, 1:50), and IgD-FITC (Southern Biotech, Birmingham, AL, USA; #2030-02, 1:100). Marginal zone (MZ)-like B cells ($CD19^+CD27^+IgD^+$) and memory B cells ($CD19^+CD27^+IgD^-$) were FACS-sorted. From the parotid gland of SG39, total $CD19^+CD27^+$ memory B cells were sorted. The B cells were cultured in IMDM (Thermo Fisher Scientific, Breda, The Netherlands) with irradiated CD40L-expressing L cells and IL-21 (AIMM Therapeutics, Amsterdam, The Netherlands) for 36 h. Next, the cells were transduced with BCL-6/BCL-XL, as described previously [18,19]. The polyclonal B cells were cultured for 8–12 days under continuous CD40L and IL-21 stimulation. From the cultured polyclonal populations, RF-producing B cells were antigen-specifically FACS-sorted, using AF647-labeled human IgG (AF647 labeling kit, Thermo Fisher Scientific, #A20173), and cultured at one cell per well in 96-well culture plates. After 21–28 days, the B-cell clones were checked for production of IgM RFs by ELISA using human IgG (Sanquin, Amsterdam, The Netherlands) [10]. For patients SG12, SG13, and SG33, RF clones could only be isolated from $CD27^+IgD^+$ MZ-like B cells. For patient SG22, RF clones were exclusively isolated from the $CD27^+IgD^-$ memory B-cell population, and for patient SG39, the RF clones were isolated from $CD27^+$ B cells that had not been analyzed for IgD expression (supplementary material, Table S2). No RF-expressing clones were isolated from SG4 and SG20, even though both were positive for serum IgM-RF. In contrast, for patient SG13, RF-expressing clones were isolated, although the serum IgM-RF level (at the time of salivary gland biopsy) was reported to be below detection level (supplementary material, Tables S1 and S2). For exome sequencing as well as IGHV and IGKV/IGLV sequencing, the IgM-RF clones were expanded in culture up to a total of 1.5–2 $\times$ 10^6 cells.

Immunoglobulin sequencing

RNA was isolated from the B-cell clones using TRI Reagent (Sigma-Aldrich, Merck, Amsterdam, The Netherlands) and cDNA was synthesized using Pd(N)6 random primers and Moloney murine virus RT (Thermo Fisher Scientific, #28025013). Rearranged *IGHV* and *IGKV/IGLV* genes were amplified using IG family-specific primers in combination with a reverse primer, specific for the matching constant region of IgM (Cμ), IgK (Cκ), or IgL (Cλ), respectively. The resulting IGHV, IGKV, and IGLV PCR products were

sequenced (Thermo Fisher Scientific) and analyzed on the IMGT website with V-Quest (http://www.imgt.org/) [20].

Whole-exome sequencing

Genomic DNA was isolated using the QIAamp DNA mini kit (Qiagen, Venlo, The Netherlands; #51306). DNA concentration was measured fluorometrically using a Qubit 4 apparatus (Thermo Fisher Scientific). Whole-exome sequencing using an Illumina platform and sequence analyses were conducted, as we have described previously [6]. Patient-specific SNPs were identified in DNA retrieved from suspensions of corresponding polyclonal memory B cells. Somatic mutations were called when present in >30% of the sequence reads. The raw reads, reads that passed quality control, aligned reads, and mean exome coverage are given in supplementary material, Table S3.

To determine whether a mutation was within 1.5 kb of the transcription start site of a particular gene, we used the annotation of the UCSC genome browser assembly of February 2009, GRCg37/hg19 (https://genome.ucsc.edu/) [21]. Expression of mutated genes by germinal center (GC) B cells was determined based on the publicly available microarray database Amazonia!, using the B cell to plasma cell analysis. A microarray signal cutoff of at least 100 in centrocytes and centroblasts was scored as positive [22,23]. All the genes which we found to be mutated in the RF and TT clones were checked for expression in DLBCL using RNA-seq data on DLBCL cell lines of the Cancer Cell Line Encyclopedia (Broad Institute, Cambridge, MA, USA) (https://portals.broadinstitute.org/gpp/public). A mean gene-expression of at least −0.5 was scored as positive [24].

Targeted next-generation sequencing

A multiplexed PCR (Ion AmpliSeq) NGS panel (Thermo Fisher Scientific) was designed to cover the 30 missense and splice site mutations (genes marked in red in Figure 2B) found in the two RF subclones (groups 1 and 2) of the SG22 RF. Frozen sections of salivary gland and DLBCL tissues of SG22 were incubated in proteinase K lysis buffer, and DNA concentrations were measured fluorometrically using a Qubit 4 apparatus (Thermo Fisher Scientific). Each library was prepared using 10 ng of DNA.

Results

Isolation of RF B-cell clones

Starting with BCL-6/BCL-XL-immortalized polyclonal memory B cells, single RF B cells were FACS-sorted using AF647-labeled human IgG. From five SS patients (SG12, SG13, SG22, SG33, and SG39), we established long-term cultures of seven unique RF clones, from six of which multiple independent clone members were isolated (Table 1).

Based on V(D)J rearrangements and VH-CDR3 amino acid sequences, we previously defined five groups of stereotypic RFs, i.e. encoded by two distinct *IGHV1-69/JH4* rearrangements, designated as V1-69-RF and WOL-RF (both known as Wa idiotype RFs), and by *IGHV3-7/JH3*, *IGHV4-59/JH2*, and *IGHV4-59/JH5* rearrangements, designated as V3-7-RF, V4-59-RF, and V4-59/JH5-RF, respectively [10,13]. The RF clones SG13, SG22, SG39 CT12, and SG39 CT13 express the stereotypic V1-69-RF *IGHV/IGKV* combination of *IGHV1-69/IGHJ4* and *IGKV3-20*. Of these, SG13, SG22, and SG39 CT13 also displayed homology with VH-CDR3s of stereotypic V1-69-RFs. SG39 CT5 expresses a stereotypic V3-7-RF encoded by the combination of *IGHV3-7/IGHJ3* and *IGKV3-15* (Table 1, Figure 1A, and supplementary material, Table S4).

The secreted IgM antibodies produced by the immortalized RF clones were tested in ELISA for binding IgG (i.e. RF activity), actin, insulin, and myosin. All seven RF clones displayed strong monoreactive RF activity in ELISA (Figure 1B,C).

Mutation analysis of cloned RF B cells

Whole-exome sequencing (WES) analysis was performed on all seven RF clones of the SS patients. The number of exome mutations that we found varied between 16 and 65 mutations per RF clone (mean 44 mutations) (Figure 2A and supplementary material, Table S4). For SG12, SG13, SG22, SG39 CT5, and SG39 CT12, more than one clone member was sequenced. Of note, the exome mutations found in the different members of each clone, except SG22, were identical or near-identical (Figure 2A and supplementary material, Table S5), underscoring that the mutations detected are indeed generated *in vivo* and that the *in vitro* mutation rate, if present, is very low. Mutation analyses performed on four of the ten clone members of SG22 (CL1, M3, 2B2, and M20) demonstrated that in spite of the fact that they all expressed the same *IGHV/IGKV* rearrangement with identical VH-CDR3, they represented members of two major subclones, group 1 (CL1 and M3) and group 2 (2B2 and M20), respectively harboring 35 and 38 exome mutations. A total of 16 mutations were shared between all four members of the SG22 clone (Figure 2B).

We detected a statistically significant difference ($p = 0.002$) between the numbers of exome mutations of the seven RF clones (mean 44, range 16–65) and the nine normal TT-specific memory B cells (mean 16, range 6–35), as previously reported (Figure 2C) [6]. Moreover, in the RF clones, no correlation was found between the numbers of somatic IGHV mutations and non-IG mutations (Figure 2D). In total, we found 346 and 153 mutations in the seven RF clones and the nine TT clones, respectively, of which 147 and 80 were non-synonymous mutations (i.e. missense, splice site, and small insertions or deletions) in the RF and TT clones, respectively.

The collective exome mutation spectrum of the RF clones, as in TT memory cells, did not for the majority display the hallmarks of involvement of the B-cell-specific mutator AID, i.e. preference for cytidine (C) nucleoside

Table 1. IGHV and IGKV sequence analysis of the cloned RF B cells.

Clone	No. of clone members	IGHV/IGKV rearrangement	No. of somatic mutations	VH/VK-CDR3 sequence	CDR3 length
SG12	6	V3-15/D5-18/JH4	3	C (T/S)TDKDTAMTQFDY WGQG*	13
		VL1-44/JL2	5	C AAWDDSLNGVV FGGG	11
SG13	4	V1-69/D2-15/JH4	21	C ARESGDGTQQVRPFVY WGQG	16
		VK3-20/JK1	15	C HQYGSSPQT FGQG	9
SG22	10	V1-69/D6-13/JH4	9–11	C AREGKQQIASSFDY WGQG	14
		VK3-20/JK1	5	C QQYGSSPQT FGQG	9
SG33	4	V7-4.1/D6-19/JH4	10	C AREHDSGWSYFDY WGQG	13
		VK4-1/JK1	7	C QQYYTTPRT FGQG	9
SG39 CT5	4	V3-7/D3-22/JH3	9	C ARGDYHDSSGSFIDAFDI WGQG	18
		VK3-15/JK1	6	C QQYNNWPPWT FGQG	10
SG39 CT12	10	V1-69/D1-14/JH4	14–16	C VRSWDHLTPWNY WGQG	12
		VK3-20/JK5	5–6	C QQYGSSPIT FGQG	9
SG39 CT13	1	V1-69/D4-23/JH4	9	C AREGRATVVNPFDY WGQG	14
		VK3-20/JK4	4	C QQYGSSPLT FGGG	9

*One clone member of SG12 (H9) expressed a serine (S) at position 105.

A

	IGHV	IGHV-CDR3 region	Homology	Genbank
SG13	V1-69/JH4	C ARESGDGTQQVRPFVY WGQG	63%	
V1-69-RF, Ro7-RF	V1-69/JH4	C ARGSGEHTNMVVPFDY WGQG		AAA20178
SG22	V1-69/JH4	C AREGKQQIASSFDY WGQG	57%	
V1-69-RF, MR1-RF	V1-69/JH4	C AREGRRMAANPYDY WGQG		CAA84417
SG39 CT5	V3-7/JH3	C ARGDYHDSSGSFIDAFDI WGQG	94%	
V3-7-RF, MALT-L M6	V3-7/JH3	C ARGDYFDSSGSFIDAFDI WGQG		AAP34731
SG39 CT13	V1-69/JH4	C AREGRATVVNPFDY WGQG	79%	
V1-69-RF, BOR-RF	V1-69/JH4	C AREGRRMAINPFDY WGQG		1313976A

B

RF clone	IGHV	Stereotypic RF	Actin	Insulin	Myosin	Human IgG
SG12	V3-15/JH4					
SG13	V1-69/JH4	V1-69-RF				
SG22	V1-69/JH4	V1-69-RF				
SG33	V7-4.1/JH4					
SG39 CT5	V3-7/JH3	V3-7-RF				
SG39 CT12	V1-69/JH4					
SG39 CT13	V1-69/JH4	V1-69-RF				
CLL57 pos contr.	V3-30/JH5					
M8 neg contr.	V3-30.3/JH6					

C

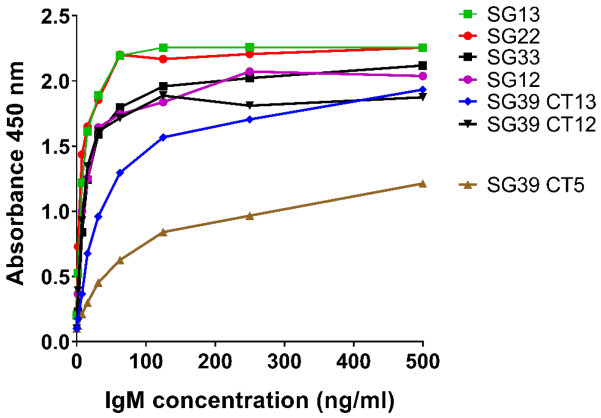

Figure 1. Cloned autoreactive B cells produce high-affinity rheumatoid factors. (A) Four out of seven RF clones identified shared VH-CDR3 amino acid sequence homology with stereotypic RFs. Identical and homologous residues are highlighted in red and blue, respectively. (B) All seven RF B-cell clones strongly bind coated human IgG and lack reactivity with actin, insulin, and myosin. The polyreactive CLL57 IgM strongly binds all tested antigens. M8 IgM is derived from a MALT lymphoma known to be unreactive for the tested antigens. Red-colored boxes indicate reactivity at all IgM concentrations tested (125, 250, 500, and 1,000 ng/ml). Green boxes indicate binding at none of the concentrations. (C) Binding profiles of titrated IgM RF to coated human IgG as measured in ELISA.

in WRC (W = A/T and R = A/G) hotspot motif in a region 1–2 kb downstream of the promotor regions of expressed genes (Table 2). In the RF and TT clones, only 25% and 26% of the G/C mutations were present within the AID hotspot motif WRC and in both groups, not more than 7% and 8% of these hotspot G/C mutations were in genes known to be expressed by GC B cells. For comparison, in the IGHV/IGKV/IGLV regions as much as 64% and 61% of the mutations at G/C were located in the WRC, hotspot motif for the RF and TT clones, respectively (Table 2).

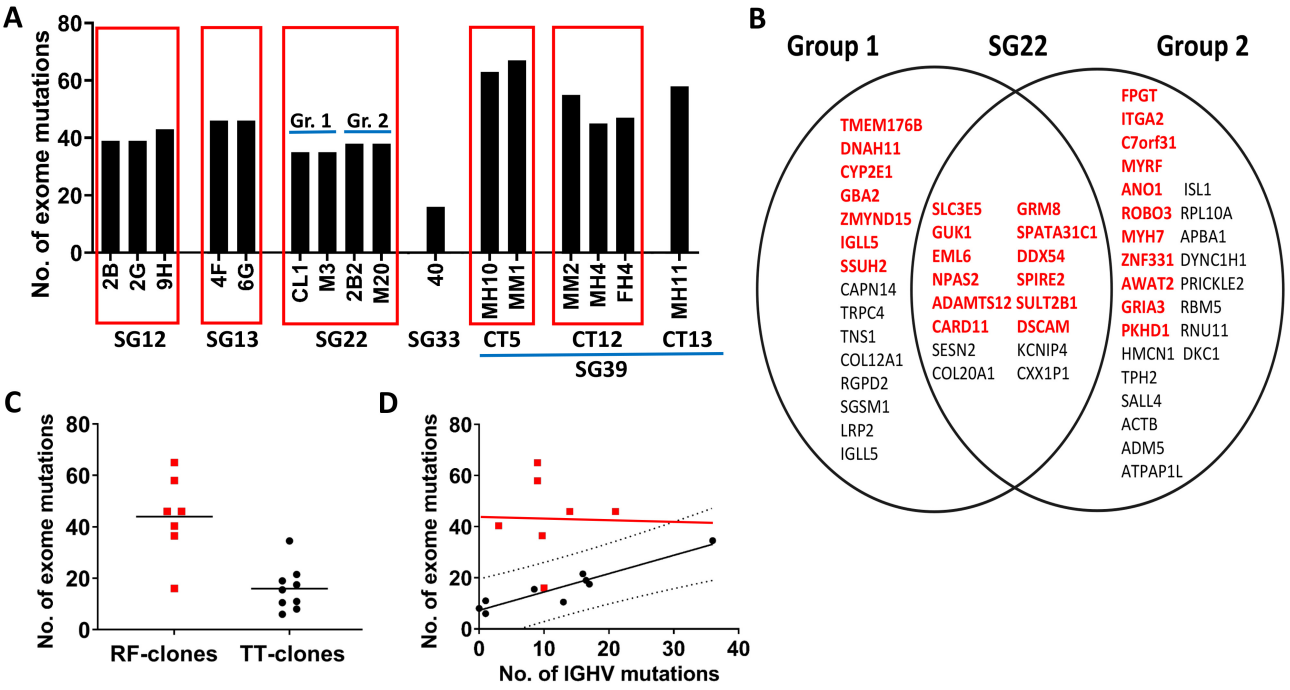

Figure 2. Elevated numbers of *de novo* genomic mutations in RF B cells compared with TT B cells. (A) Numbers of exome mutations detected in the RF B cells cloned from PMBCs (SG12, SG13, SG22, SG33) or salivary gland tissue (SG39) of SS patients. Histograms within red boxes represent individual members of an indicated RF clone. (B) Genomic mutations identified in two major subsets (each represented by two members) of clone SG22 sharing 16 mutations. Genes indicated in red and black represent non-synonymous and synonymous mutations, respectively. (C) Graph depicting a statistically significant difference in the number of exome mutations between RF clones and TT clones (Mann–Whitney test, two-tailed, $p = 0.002$). (D) Graph depicting the number of IGHV mutations (*x*-axis) versus the number of exome mutations (*y*-axis) of individual RF clones (red squares) and memory TT clones (black dots). In TT clones, a statistically significant correlation between the number of IGHV and exome mutations was observed as reported previously (Slot *et al* [6]) ($p = 0.0002$). In RF clones, no statistical correlation between the number of IGHV and exome mutations was observed ($p = 0.9258$). The dashed line represents the 99% prediction interval of the ratios of IGHV and exome mutations for the normal B-cell clones.

Table 2. Exome and IGHV mutation spectrum characteristics of RF clones versus those of TT clones.

	Exome mutations		IGHV/IGKV mutations	
	RF clones	TT clones	RF clones	TT clones
No. of sequenced clones	7	9	7	9
Total number of mutations (all clones)	346	153	125	148
Mean number of mutations per clone	44	16	18	16
No. of mutations located within 1.5 kb of the transcription initiation site	79/346 (23%)	21/153 (14%)	na	na
No. of genes for which expression data in GC B cells are available in Amazonia!	272/346 (79%)	116/153 (76%)	na	na
No. of mutated genes known to be expressed in GC B cells	92/272 (34%)	25/116 (22%)	na	na
No. of mutations in genes, expressed in GC B cells and located within 1.5 kb of the transcription initiation site	22/272 (8%)	5/116 (4%)	na	na
No. of nucleotide substitutions (all clones)	333	153	125	148
Nucleotide substitutions at G or C	230/333 (69%)	98/153 (64%)	83/125 (66%)	85/148 (57%)
G or C nucleotide substitutions at WRC	58/230 (25%)	25/98 (26%)	53/83 (64%)	52/84 (61%)
No. of substitutions at G or C in genes, for which expression data are available	181/230 (79%)	84/98 (86%)	na	na
Nucleotide substitutions at G or C located at WRC in genes expressed by GC B cells	13/181 (7%)	7/84 (8%)	53/83 (64%)	52/84 (61%)

na, not applicable.

RF B cells have non-random mutation spectra enriched for mutations in lymphoma-associated genes

We evaluated the potential biological significance of the 147 and 80 non-synonymous mutations (missense, splice site, small insertions, or deletions) identified in the seven RF clones and the nine TT clones, respectively. Recently, Cascione *et al* [25] and Moody *et al* [26], based on respective analyses of 72 and 249 MALT lymphomas of various sites, together identified a set of 57 mutated genes (supplementary material, Figure S1).

Three of the seven RF clones, but none of the nine TT clones, harbored one or more non-synonymous mutations in genes also identified in this reference set of MALT lymphoma-mutated genes; SG22, SG39 CT5, and SG39 CT12 carried mutations in *CARD11*, *HIST1H1D*, *TBL1XR1*, and *CD79A*, respectively. SG39 CT12 harbored mutations in both *TBL1XR1* and *CD79A*. In the combined cohorts of Cascione *et al* [25] and Moody *et al* [26], *CARD11* mutations were detected in 5/71 (7%) and *TBL1XR1* mutations were found in 15/71 (21%) of salivary gland MALT lymphomas. Moreover, a *HIST1H1D* mutation was detected in a thyroid MALT lymphoma [25], and mutated *CD79A* was detected in two gastric MALT lymphomas [26].

Subsequently, we explored which of the RF-clone mutated genes are expressed by GC B cells and/or DLBCL using two publicly available databases, i.e. microarray expression on Amazonia! (regarding the analysis of B cell to plasma cell differentiation) and RNA-seq data of DLBCL cell lines [23,24]. This revealed that 68 of the 147 (46%) and 32 of the 80 (40%) non-synonymously mutated genes in SS RF and TT clones, respectively, are expressed by GC B cells and/or DLBCLs and thus may potentially have a biological effect (supplementary material, Table S6). Next, all of these expressed non-synonymously mutated genes were compared with genes that were found to be mutated in six WES studies on a total of 518 DLBCLs [27–32]. We arbitrarily scored a gene mutation as being lymphoma-associated when it was reported as non-synonymously mutated in at least four DLBCLs (0.8%) and expressed by GC B cells and/or DLBCL. For the SS RF and TT clones, we thus assigned 24 mutations (in 22 genes) and eight mutations (in eight genes), respectively, as lymphoma-associated (Table 3 and Figure 3). Importantly, six RF clone mutations, i.e. in *CARD11*, *IGLL5*, *ZEB2*, *TBL1XR1*, *BTG1*, and *HIST1H1D*, and only one TT clone mutation in *SGK1* were found to be prominently mutated in a higher frequency of >2.5% in the set of 518 DLBCLs (Table 3). Except for *IGLL5*, mutations in *CARD11*, *ZEB2*, *TBL1XR1*, *BTG1*, and *HIST1H1D* are known as driver mutations in lymphomas and other cancers [33–39]. The well-known lymphoma driver mutation in *CD79A* of SG39 CT12 and of two reported gastric MALT lymphomas [26] was found to be mutated in only two of the 518 DLBCLs. Of note, as an additional control, we had sequenced one B-cell clone of SG22 with unknown specificity, which showed a mutation profile comparable with that of the TT clones (supplementary material, Table S5). In Table 3 and supplementary material, Text S1, the most prominent recurring mutations and their potential biological significance are described.

Development of a diffuse large B-cell lymphoma from the SG22 RF clone

During the follow-up period of our cohort, patients SG4 and SG22 were diagnosed with a *de novo* DLBCL. We were previously unable to sequence the clonal IGHV rearrangement in the DLBCL of patient SG4; therefore, this patient was not studied further [40]. From the DLBCL of patient SG22, a fine needle biopsy was available for further analysis. The DLBCL was found to be clonally related to the SG22 RF clone identified in peripheral blood (PB) and in a salivary gland biopsy, respectively taken 11 and 26 months earlier [40]. By IG RT-PCR and sequencing of various samples microdissected from the salivary gland tissue, we had demonstrated that the SG22 RF clone was highly expanded and had somatically diversified its IGHV (Figure 4A). Among a total of 111 molecular IGHV clones, we detected a high degree of intraclonal sequence variation (range 3–15 mutations/IGHV), compatible with an ongoing IGHV somatic hypermutation process [40]. The SG22 RF clone-members isolated from PB carried an amount and pattern of IGHV mutations similar to those in the salivary gland. The evolved DLBCL, however, harbored significantly more IGHV and IGKV somatic mutations, i.e. 32 and 27 mutations, respectively (Figure 4A) [40].

Mutation analyses of the DLBCL and its early precursor SG22 RF clone

The group 1 and group 2 clone-members of the SG22 RF clone retrieved from PB together harbored 30 non-synonymous gene mutations, of which 12 mutations were shared between the group 1 and group 2 clone-members and 7 and 11 mutations are respectively group 1 and group 2 clone-member-specific (Figure 2B). Subsequently, NGS targeted at these 30 gene mutations was performed on whole DNA isolated respectively from the labial salivary gland, the DLBCL, and from nine SG22 RF clone-members of PB, four of which have been sequenced exome-wide (CL1 and M3 of group 1, and 2B2 and M20 of group 2).

All 19 exome mutations of the group 1 clone-members (12 mutations shared with group 2 and 7 mutations group 1-specific) were detected in the group 1 clone-members CL1 and M3. Similarly, all 23 exome mutations of the group 2 clone-members (12 mutations shared with group 1 and 11 mutations group 2-specific) were confirmed in the group 2 clone-members 2B2 and M20. In addition, the five additionally sequenced SG22 RF clone-members all harbored the 23 exome mutations of the group 2 clone-members, indicating that the group 2 clone-members might be the dominant SG22 RF subclone in PB.

In the salivary gland, only the 12 shared group 1 and group 2 exome mutations were detected, with varying allele frequencies. Four mutations, in *NPAS2*, *EML6*, *GUK1*, and *SPIRE2*, were dominant and detected with allele frequencies of 11–14%. The other eight shared mutated genes were detected with allele frequencies of 2–6% (Figure 4B). All the other mutations were essentially below the detection limit (VAF <0.3%). Thus, the salivary gland biopsy contained major members of the SG22 RF clone that harbors four of the 12 shared mutations as well as minor clone-members having

Table 3. Lymphoma-associated mutations identified in RF and in memory TT clones.

Clone	Mutated gene	No. of DLBCLs reported with mutation	References	Function/pathway
SG12	*SMARCC1*	2 (0.4%)	[61,63]	SWI/SNF chromatin remodeling complexes
	HIST1H3G	7 (1.4%)	[38,39,62]	H3A1 histone core protein
	HECTD4	10 (1.9%)	None	E3 ubiquitin-protein ligase?
	GLI1	4 (0.8%)	[64]	Transcriptional activator in SHH signaling
	SYNRG	10 (1.9%)	None	Endocytosis/membrane trafficking?
SG13	*PER3*	6 (1.2%)	None	Clock gene, interacts with CLOCK/NPAS2
	HIST2H2BF	2 (0.4%)	[38,39,62]	H2B histone core protein
	N4BP2	4 (0.8%)	None	Polynuclease/nicking act., role in DNA repair?
	UBA7	4 (0.8%)	None	Activates ubiquitin
	HERC4	1 (0.2%)	[66]	E3 ubiquitin-protein ligase?
	WWOX	9 (1.7%)	[65]	Oxidoreductase, acts as TSG, role in apoptosis
	HAUS8	4 (0.8%)	None	Contributes to mitotic spindle assembly
SG22	*ZNF331*	4 (0.8%)	None	A role in transcriptional regulation
	CARD11	61 (11.8%)	[33,34]	BCR signaling, activates NF-κB
	IGLL5	44 (8.5%)	[45,54]	IG lambda-like 5, non-rearranged
SG39 CT5	*ARHGEF6*	6 (1.2%)	None	RAC1 guanine nucleotide exchange factor
	CLASP1	9 (1.7%)	None	Microtubule plus-end tracking protein
	ZEB2	15 (2.9%)	[52]	Transcriptional inhibitor of various promotors
	HIST1H1D	29 (5.6%)	[38,39,62]	H1H1 linker histone protein
	IGLL5	44 (8.5%)	[45,54]	IG lambda-like 5, non-rearranged
SG39 CT12	*TBL1XR1*	33 (6.4%)	[36]	Transcription act. of nuclear receptors
	ALDH7A1	5 (1.0%)	None	Protects cells from oxidative stress
	TAGAP	8 (1.5%)	None	GTPase-act. with a role in T-cell activation
	FLI1	3 (0.6%)	[67]	Sequence-specific transcriptional activator
	NACA	8 (1.5%)	None	Prevents wrong targeting of non-secretory polypeptide
	CD79A	2 (0.4%)	[49]	BCR signaling, required with CD79B
	IGLL5	44 (8.5%)	[45,54]	IG lambda-like 5, non-rearranged
SG39 CT13	*ZNF627*	5 (1.0%)	None	Transcriptional regulation?
	BTG1	60 (11.6%)	[32,52]	Anti-proliferative protein
TT3 A(36)	*FOCAD*	8 (1.5%)	None	Focadhesin, potential TSG in gliomas
	SGK1	64 (12.4%)	[32,45]	Ser/Thr-protein kinase, role in ion channels
TT3 N	*ZNF627*	5 (1.0%)	None	Transcriptional regulation?
TT4 G	*PI4KA*	9 (1.7%)	None	Phosphatidylinositol 4-kinase alpha
TT4 H	*KIAA1958*	4 (0.8%)	None	Nucleotide sugars from cytosol into Golgi
TT4 K	*TRAPPC9*	5 (1.0%)	None	Activates NF-κB, phosphorylation of IKK
TT4 R	*INCENP*	4 (0.8%)	None	Inner centromere protein, a role in mitosis
	TNKS1BP1	8 (1.5%)	None	182 kDa tankyrase-1 binding protein

List of all non-synonymously mutated expressed genes identified in RF clones and normal memory TT clones, which have also been found to be mutated in at least four DLBCLs in a series of 518 DLBCLs and/or for which literature is available. The genes highlighted in bold are genes that are frequently (>2.5%) mutated in DLBCL. For the underlined genes, literature on lymphomagenesis is available, but these genes have not been found to be mutated in at least four DLBCLs. In one RF clone (SG33) and in three TT clones [TT3 A(15/16), TT3 Q, and TT4 AZ], no lymphoma-associated mutations were identified. Except for the *IGLL5* and *ZNF331* mutations of SG22, all lymphoma-associated mutations were present in all clone members of the particular RF clones.

the four dominant shared mutations and varying frequencies of the other eight shared mutations, including the D357V mutation in *CARD11*, which was detected with an allele frequency of 3% (Figure 4B). Because of the importance of the *CARD11* D357V mutation in lymphoma, we confirmed the presence of this mutation also by Sanger sequencing in three out of 12 tissue samples, microdissected from the salivary gland biopsy [40].

In the DLBCL, the four dominant mutations of the salivary gland were detected with allele frequencies of 13–20%, whereas five of the eight shared mutations, including the *CARD11* mutation, were detected with very low allele frequencies (0.4–0.9%). All the other mutations, including the other three shared mutations, were not detected (VAF <0.3%) (Figure 4B). The DLBCL harbored 13 mutations in exon 1 and intron 1 of *IGLL5*, with allele frequencies of 13–32%. Remarkably, none of these 13 *IGLL5* mutations of the DLBCL tissue overlapped with the three *IGLL5* mutations found in the group 1 members of the PB-derived SG22 RF clone. Of the four gene mutations (*NPAS2*, *EML6*, *GUK1*, and *SPIRE2*), detected in both the RF clone of the salivary gland and the DLBCL, only *NPAS2* has previously been found to be non-synonymously mutated in two DLBCLs in two series of, in total, 52 patients who suffered from relapsed or refractory disease [30,31]. However, as *NPAS2* is not expressed in GC B cells or in DLBCL cell lines, a biological function for *NPAS2* in the development of lymphoma is unlikely (supplementary material, Table S5) (Figure 5).

In addition to the 30 mutated genes of our targeted sequencing panel, we also included a set of nine other genes, i.e. *TNFAIP3*, *KLF2*, *MYD88*, *BIRC3*, *TRAF3*, *TP53*, *NOTCH1*, *NOTCH2*, and *IKBKB*, known to be frequently mutated in marginal zone lymphoma and DLBCL. However, no mutations in these genes were detected in the salivary gland and DLBCL DNA.

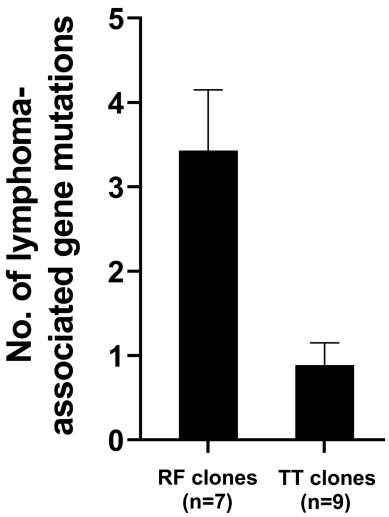

Figure 3. RF cells harbor significantly more mutations in lymphoma-associated genes than TT memory B cells. A non-synonymous gene mutation was designated as lymphoma-associated when it was also reported to be non-synonymously mutated in at least four cases of a series of 518 WES-sequenced DLBCLs. The difference in lymphoma-associated gene mutations between RF and TT clones is statistically significant according to the Mann–Whitney two-tailed test ($p = 0.0098$).

Discussion

In this study, we assessed the genomic mutations present in auto-reactive human B cells in patients with Sjögren's syndrome. Our analyses demonstrated that primary memory B cells expressing functional, affinity-selected RFs carry a substantial number of somatic mutations in the protein coding regions of non-IG genes exceeding the mutation frequencies found in normal memory TT-specific B cells of healthy donors (Figure 2). The mean mutational load in the seven RF clones retrieved from SS patients was almost three times higher than the mean number of mutations detected in nine TT memory B cells [6]. As in the TT B cells, the majority of the somatic non-IG mutations in the RF clones did not appear to be AID-mediated but most likely were replication errors. In fact, the mutation spectrum of TT and RF B cells was highly comparable with that of DLBCL, i.e. 64%, 69%, and 66% of the mutations are at G/C respectively, of which 26%, 25%, and 30% are located in the AID WR**C** hotspot [32].

In contrast to TT memory cells, in the RF cells no correlation was found between the mutation frequencies in the IGHV genes and those in the non-IG genes. The fact that RF cells harbored in general more exome mutations compared with TT cells and the lack of a correlation between IGHV and exome mutation numbers together suggest a scenario of RF B cells being chronically stimulated to clonally expand beyond GCs, thus without gaining AID-mediated IGHV mutations. Indeed, whereas the frequencies of exome mutations are significantly higher in the RF clones, the mean numbers of IGHV mutations in RF cells and in the TT cells were highly similar (11 versus 12, respectively).

In contrast to the mutations identified in TT B cells, which appear random and non-selected, the mutation spectrum of RF clones is biased and clearly enriched for genes that are frequently affected in malignant lymphomas and in other cancers, i.e. *CARD11*, *TBL1XR1*, *CD79A*, *ZEB2*, *BTG1*, *HIST1H3G*, *HIST2H2BF*, *HIST1H1D*, and *IGLL5* [33–39]. Recently, Singh *et al* [41] sequenced 153 lymphoma-associated genes expressed in single non-malignant RF B cells derived from four SS patients and detected mutations in *CARD11*, *KLHL6*, *TNFAIP3*, *CCND3*, *ID3*, and *BTG2*. *CARD11*, *TBL1XR1*, *KLHL6*, *CD79A*, *TNFAIP3*, and *HIST1H1D* have also been reported to be mutated in MALT lymphomas [25,26,42,43]. In particular, *TBL1XR1* and *TNFAIP3* mutations frequently occur in salivary gland- and ocular adnexal-MALT lymphomas, respectively [26,43]. The RF B-cell mutation spectrum overlapped with those of two genetically defined DLBCL subsets, MCD/C5 and BN2/C1, mostly with an activated B-cell (ABC)-like expression profile, as recently reported by Wright *et al* [44] and Chapuy *et al* [32]. *TBL1XR1*, *BTG1*, and *BTG2* are frequently (co-) mutated in MCD/C5 DLBCL, and *TNFAIP3*, *ZEB2*, *CCND3*, and *HIST1H1D* are mutated in BN2/C1 DLBCL. SG39 CT5 harbored both *ZEB2* and *HIST1H1D* mutations, and one of the RF clones of Singh *et al* [41] harbored *TNFAIP3* and *CCND3* mutations, each fitting in the BN2/C1 profile. The SG39 CT12 RF clone, with combined *TBLXR1* and *CD79A* mutation, fits in the MCD/C5 profile.

Remarkably, within this small cohort, two patients (SG4 and SG22) developed a DLBCL. Patient SG22 was diagnosed with DLBCL 26 months after the identification of the expanded precursor RF clone in a salivary gland [40]. In the salivary gland RF clone, four genomic mutations (in *NPAS2*, *EML6*, *GUK1*, and *SPIRE2*) were dominant and detected with allele frequencies of 11–14%, whereas eight mutations with lower allele frequencies of 2–6% were found. This finding is compatible with our previously reported high degree of intraclonal IGHV heterogeneity of the SG22 RF clone in the salivary gland [40]. Based on the variant allele frequencies (VAFs) of the various mutations found in the salivary gland, it is estimated that about 25% of the RF clone-members harbor the *CARD11* D357V lymphoma driver-mutation. It is remarkable that the *CARD11* D357V mutation was not detected in the DLBCL, indicating that it evolved out of the pool of 75% of the precursor cells in the salivary gland that lacked this mutation. Apparently, the RF precursor cells had accumulated other secondary genomic damage, including lymphoma driver-mutations. This may be associated with AID activity beyond the IG loci, as reflected by the strong accumulation of *IGHV/IGKV* and *IGLL5* mutations in the DLBCL. IGLL5 resembles IGLL1, also called λ5, and is a known target of AID-mediated kataegis in DLBCL [45].

Previously, we reported that although the majority of SS-associated salivary gland MALT lymphomas express RFs, this is not a reflection of the incidence of RF B cells

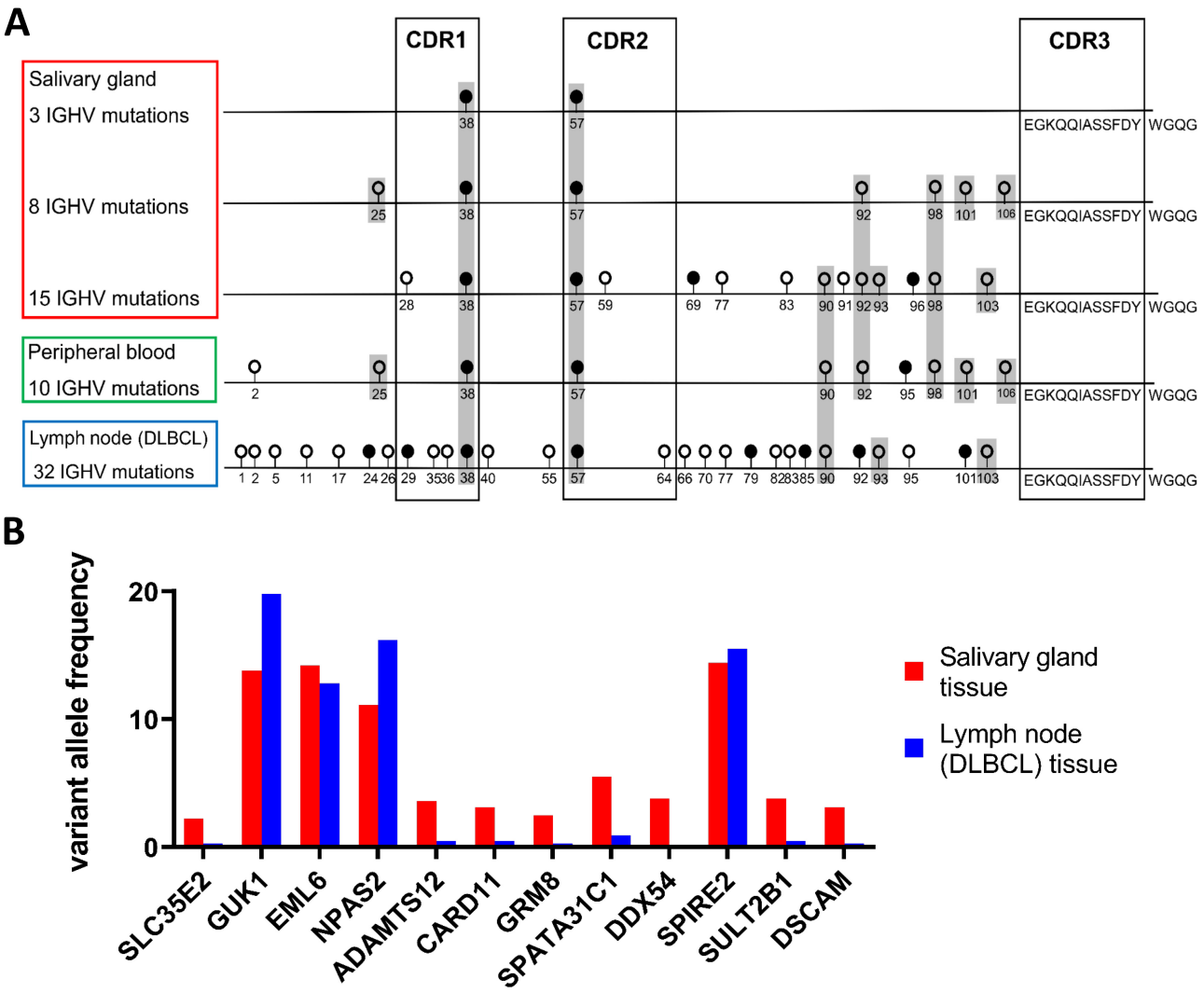

Figure 4. IGHV and non-IG mutations in the SG22 RF clone successively retrieved from salivary gland, blood, and lymph node. (A) The peripheral blood and the lymphoma samples were taken 15 and 26 months, respectively, after the initial salivary gland biopsy. For the salivary gland, three out of a total of 111 previously identified molecular IGHV clones are depicted [40]. IGHV sequences are schematically depicted. Lollipop-shaped symbols indicate somatic mutations identified by comparison with the germline IGHV1-69 gene segment. Closed and open circles represent replacement and silent mutations, respectively. Gray-shaded boxes indicate identical mutations. (B) Variant allele mutation frequencies (VAFs) of 12 non-synonymously mutated genes shared by the group 1 and group 2 clone-members of the PB SG22 RF clone as detected in the salivary gland and lymph node (DLBCL) DNA.

present in inflamed salivary glands of SS patients [12,40,46]. In fact, the frequency of RF B cells in the salivary glands of SS patients was found to be low [40,46,47]. A common feature of RF-expressing B-cell lymphomas is that the majority arise in a background of chronic inflammation caused by autoimmunity or infectious agents, such as *H. pylori* (*Hp*) or HCV [48]. In this inflammatory environment, RF B cells likely experience chronic BCR ligation by IgG present in immune complexes, e.g. IgG-SSA/SSB in SS sialadenitis, IgG-*Hp* in *Hp* gastritis, and IgG-HCV in HCV hepatitis. The immune complexes also include ligands for TLR7 or TLR9, such as SSA/SSB-associated single-stranded RNA stem-loop structures (Y-RNAs), HCV viral RNA, and *Hp* bacterial DNA, thus delivering dual RF BCR/TLR stimulation, potentiating the NF-κB pathway [48]. It is unknown why SS-related MALT lymphomas do not evolve from auto-reactive SSA/SSB-specific memory B cells, which may upon binding of SSA/SSB coupled to Y-RNAs, like RF B cells, also experience chronic dual BCR/TLR stimulation. Also, ABC DLBCLs of the MCD/C5 subset depend, like RF-expressing MALT lymphomas, on chronic dual BCR-/TLR-signals, together activating NF-κB and STAT3 [49]. In 70% of MCD/C5 DLBCLs however, sustained TLR signaling is secured by a *MYD88* (L265P) mutation, alleviating the need for physiological TLR stimulation [44,49,50]. Still, these DLBCLs are reported to be also dependent on chronic antigen-induced BCR signaling for their survival [49].

We and others have reported that RF cells in SS patients have a high risk of transforming into lymphoma and that the majority of salivary gland MALT lymphomas express high affinity RFs [10–12]. It was hypothesized that RF B cells, due to chronic BCR/TLR-induced cell replication brought about by

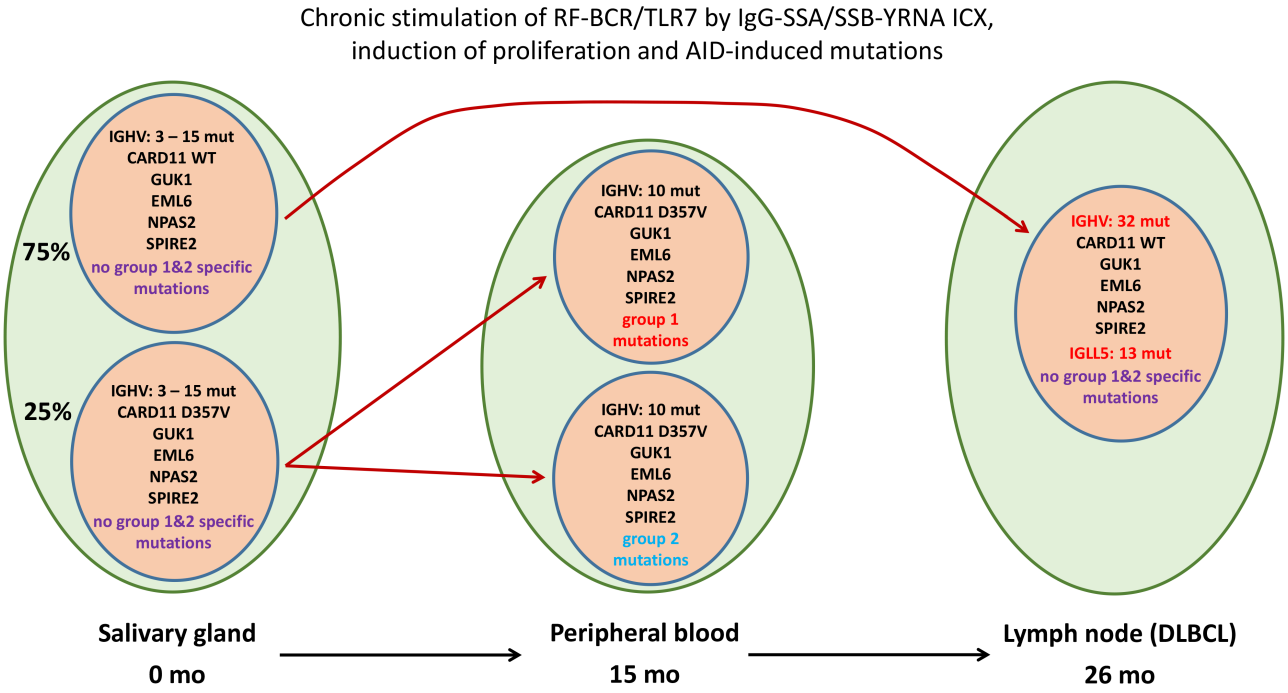

Figure 5. DLBCL development in patient SG22. In patient SG22, an RF-expressing DLBCL was diagnosed in a lymph node taken 26 months after an initial salivary gland biopsy, in which a clonally-related expanded RF clone was present. The RF clone was also detected in PB, taken 15 months after the salivary gland biopsy. Sequencing revealed that four SG22 RF clone-members from PB represented two subclones of two clone-members, each harboring identical exome mutations, named group 1 and group 2. Twelve non-synonymous mutations were shared between the group 1 and group 2 clones. Four of the shared mutations, which were also detected in the DLBCL, are depicted in the figure. The shared group 1 and group 2 *CARD11* D357V mutation was not present in the DLBCL or in about 75% of the RF clone-members of the salivary gland. Remarkably, the DLBCL had gained a very high number of *IGHV* and *IGLL5* mutations, indicating that AID activity beyond the IG loci may have played a role in the transformation process. ICX, immune complex.

IgG immune complexes, accumulate more genomic mutations than non-autoreactive B cells, including driver gene mutations, providing additional growth advantage. Our current observation that RF cells indeed carry elevated numbers of mutations, also in genes affected in MALT lymphomas and ABC-type DLBCLs, as well as the documented transformation of an expanded RF clone into a DLBCL in one patient, is compatible with this scenario and raises the question of whether RF cells are to be considered as (pre)neoplastic.

Acknowledgements

We thank Berend Hooibrink, Toni van Capel, and Kim Brandwijk from the Department of Medical Biology, Amsterdam UMC for FACS sorting and Martin de Rooij from the Department of Pathology, Amsterdam UMC for technical assistance. This research was supported by grants from the Dutch Arthritis Foundation (2015-2-310) and the Dutch Cancer Society (UVA2009-4524 and UVA2014-6824).

Author contributions statement

RJB, LMS, JEG, FGK and CJN were responsible for study conception and design. RJB, LMS, MJK, TAW and GMV performed experiments for acquisition of data. RJB, LMS, TAW, AJ, ACK, JEG, FGK and CJN analyzed and interpreted data. RJB, LMS, FGK and CJN wrote the manuscript. All the authors approved the final submitted version.

Data availability statement

The raw FAST-Q files are available from the authors, without undue reservation.

References

1. Blokzijl F, De Ligt J, Jager M, *et al*. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* 2016; **538:** 260–264.
2. Martincorena I, Roshan A, Gerstung M, *et al*. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 2015; **348:** 880–886.
3. Martincorena I, Fowler JC, Wabik A, *et al*. Somatic mutant clones colonize the human esophagus with age. *Science* 2018; **362:** 911–917.
4. Lodato MA, Rodin RE, Bohrson CL, *et al*. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* 2018; **359:** 555–559.
5. Vogelstein B, Papadopoulos N, Velculescu VE, *et al*. Cancer genome landscapes. *Science* 2013; **339:** 1546–1558.
6. Slot LM, Wormhoudt TAM, Kwakkenbos MJ, *et al*. *De novo* gene mutations in normal human memory B cells. *Leukemia* 2019; **33:** 1219–1230.

7. Machado HE, Mitchell E, Øbro NF, *et al*. Diverse mutational landscapes in human lymphocytes. *Nature* 2022; **608:** 724–732.
8. Di Noia JM, Neuberger MS. Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* 2007; **76:** 1–22.
9. Zhang L, Dong X, Lee M, *et al*. Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. *Proc Natl Acad Sci U S A* 2019; **116:** 9014–9019.
10. Bende RJ, Aarts WM, Robert GR, *et al*. Among B cell non-Hodgkin's lymphomas, MALT lymphomas express a unique antibody repertoire with frequent rheumatoid factor reactivity. *J Exp Med* 2005; **201:** 1229–1241.
11. Martin T, Weber JC, Levallois H, *et al*. Salivary gland lymphomas in patients with Sjögren's syndrome may frequently develop from rheumatoid factor B cells. *Arthritis Rheum* 2000; **43:** 908–916.
12. Bende RJ, Janssen J, Beentjes A, *et al*. Salivary gland mucosa-associated lymphoid tissue-type lymphoma from Sjögren's syndrome patients in the majority express rheumatoid factors affinity-selected for IgG. *Arthritis Rheumatol* 2020; **72:** 1330–1340.
13. Bende RJ, Janssen J, Wormhoudt TA, *et al*. Identification of a novel stereotypic IGHV4-59/IGHJ5-encoded B-cell receptor subset expressed by various B-cell lymphomas with high affinity rheumatoid factor activity. *Haematologica* 2016; **101:** e200–e203.
14. De Re V, De Vita S, Marzotto A, *et al*. Sequence analysis of the immunoglobulin antigen receptor of hepatitis C virus-associated non-Hodgkin lymphomas suggests that the malignant cells are derived from the rheumatoid factor-producing cells that occur mainly in type II cryoglobulinemia. *Blood* 2000; **96:** 3578–3584.
15. van Maldegem F, Wormhoudt TA, Mulder MM, *et al. Chlamydia psittaci*-negative ocular adnexal marginal zone B-cell lymphomas have biased V_H *4-34* immunoglobulin gene expression and proliferate in a distinct inflammatory environment. *Leukemia* 2012; **26:** 1647–1653.
16. Hoogeboom R, Wormhoudt TA, Schipperus MR, *et al*. A novel chronic lymphocytic leukemia subset expressing mutated IGHV3-7-encoded rheumatoid factor B-cell receptors that are functionally proficient. *Leukemia* 2013; **27:** 738–740.
17. Kostareli E, Gounari M, Janus A, *et al*. Antigen receptor stereotypy across B-cell lymphoproliferations: the case of IGHV4-59/IGKV3-20 receptors with rheumatoid factor activity. *Leukemia* 2012; **26:** 1127–1131.
18. Kwakkenbos MJ, Diehl SA, Yasuda E, *et al*. Generation of stable monoclonal antibody-producing B cell receptor-positive human memory B cells by genetic programming. *Nat Med* 2010; **16:** 123–128.
19. Kwakkenbos MJ, van Helden PM, Beaumont T, *et al*. Stable long-term cultures of self-renewing B cells and their applications. *Immunol Rev* 2016; **270:** 65–77.
20. 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–W508.
21. Kent WJ, Sugnet CW, Furey TS, *et al*. The human genome browser at UCSC. *Genome Res* 2002; **12:** 996–1006.
22. Jourdan M, Caraux A, Caron G, *et al*. Characterization of a transitional preplasmablast population in the process of human B cell to plasma cell differentiation. *J Immunol* 2011; **187:** 3931–3941.
23. Le Carrour T, Assou T, Tondeur S, *et al*. Amazonia!: an online resource to Google and visualize public human whole genome expression data. *The Open Bioinformatics Journal* 2010; **4:** 5–10.
24. Barretina J, Caponigro G, Stransky N, *et al*. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012; **483:** 603–607.
25. Cascione L, Rinaldi A, Bruscaggin A, *et al*. Novel insights into the genetics and epigenetics of MALT lymphoma unveiled by next generation sequencing analyses. *Haematologica* 2019; **104:** e558–e561.
26. Moody S, Thompson JS, Chuang SS, *et al*. Novel *GPR34* and *CCR6* mutation and distinct genetic profiles in MALT lymphomas of different sites. *Haematologica* 2018; **103:** 1329–1336.
27. Lohr JG, Stojanov P, Lawrence MS, *et al*. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc Natl Acad Sci U S A* 2012; **109:** 3879–3884.
28. Zhang J, Grubor V, Love CL, *et al*. Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc Natl Acad Sci U S A* 2013; **110:** 1398–1403.
29. Morin RD, Mungall K, Pleasance E, *et al*. Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. *Blood* 2013; **122:** 1256–1265.
30. Mareschal S, Dubois S, Viailly PJ, *et al*. Whole exome sequencing of relapsed/refractory patients expands the repertoire of somatic mutations in diffuse large B-cell lymphoma. *Genes Chromosomes Cancer* 2016; **55:** 251–267.
31. Morin RD, Assouline S, Alcaide M, *et al*. Genetic landscapes of relapsed and refractory diffuse large B-cell lymphomas. *Clin Cancer Res* 2016; **22:** 2290–2300.
32. Chapuy B, Stewart C, Dunford AJ, *et al*. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med* 2018; **24:** 679–690.
33. Lenz G, Davis RE, Ngo VN, *et al*. Oncogenic *CARD11* mutations in human diffuse large B cell lymphoma. *Science* 2008; **319:** 1676–1679.
34. Bedsaul JR, Carter NM, Deibel KE, *et al*. Mechanisms of regulated and dysregulated CARD11 signaling in adaptive immunity and disease. *Front Immunol* 2018; **9:** 2105.
35. Soen B, Vandamme N, Berx G, *et al*. ZEB proteins in leukemia: friends, foes, or friendly foes? *Hemasphere* 2018; **2:** e43.
36. Venturutti L, Teater M, Zhai A, *et al*. TBL1XR1 mutations drive extranodal lymphoma by inducing a pro-tumorigenic memory fate. *Cell* 2020; **182:** 297–316.
37. Tijchon E, van Emst L, Yuniati L, *et al*. Tumor suppressor BTG1 limits activation of BCL6 expression downstream of ETV6-RUNX1. *Exp Hematol* 2018; **60:** 57–62.
38. Nacev BA, Feng L, Bagert JD, *et al*. The expanding landscape of 'oncohistone' mutations in human cancers. *Nature* 2019; **567:** 473–478.
39. Yusufova N, Kloetgen A, Teater M, *et al*. Histone H1 loss drives lymphoma by disrupting 3D chromatin architecture. *Nature* 2021; **589:** 299–305.
40. Bende RJ, Slot LM, Hoogeboom R, *et al*. Stereotypic rheumatoid factors that are frequently expressed in mucosa-associated lymphoid tissue-type lymphomas are rare in the labial salivary glands of patients with Sjögren's syndrome. *Arthritis Rheumatol* 2015; **67:** 1074–1083.
41. Singh M, Jackson KJL, Wang JJ, *et al*. Lymphoma driver mutations in the pathogenic evolution of an iconic human autoantibody. *Cell* 2020; **180:** 878–894.
42. Jung H, Yoo HY, Lee SH, *et al*. The mutational landscape of ocular marginal zone lymphoma identifies frequent alterations in *TNFAIP3* followed by mutations in *TBL1XR1* and *CREBBP*. *Oncotarget* 2017; **8:** 17038–17049.
43. Bult JAA, Placa JR, Haacke EA, *et al*. Low mutational burden of extranodal marginal zone lymphoma of mucosa-associated lymphoid tissue in patients with primary Sjögren's syndrome. *Cancers (Basel)* 2022: 14: 1010.
44. Wright GW, Huang DW, Phelan JD, *et al*. A probabilistic classification tool for genetic subtypes of diffuse large B cell lymphoma with therapeutic implications. *Cancer Cell* 2020; **37:** 551–568.
45. Qian J, Wang Q, Dose M, *et al*. B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* 2014; **159:** 1524–1537.
46. Visser A, Doorenspleet ME, de Vries N, *et al*. Acquisition of N-glycosylation sites in immunoglobulin heavy chain genes during

local expansion in parotid salivary glands of primary Sjögren patients. *Front Immunol* 2018; **9:** 491.
47. Visser A, Verstappen GM, van der Vegt B, *et al*. Repertoire analysis of B-cells located in striated ducts of salivary glands of patients with Sjögren's syndrome. *Front Immunol* 2020; **11:** 1486.
48. Bende RJ, van Maldegem F, van Noesel CJ. Chronic inflammatory disease, lymphoid tissue neogenesis and extranodal marginal zone B-cell lymphomas. *Haematologica* 2009; **94:** 1109–1123.
49. Young RM, Phelan JD, Wilson WH, *et al*. Pathogenic B-cell receptor signaling in lymphoid malignancies: new insights to improve treatment. *Immunol Rev* 2019; **291:** 190–213.
50. Phelan JD, Young RM, Webster DE, *et al*. A multiprotein supercomplex controlling oncogenic signalling in lymphoma. *Nature* 2018; **560:** 387–391.
51. Compagno M, Lim WK, Grunn A, *et al*. Mutations of multiple genes cause deregulation of NF-κB in diffuse large B-cell lymphoma. *Nature* 2009; **459:** 717–721.
52. Reddy A, Zhang J, Davis NS, *et al*. Genetic and functional drivers of diffuse large B cell lymphoma. *Cell* 2017; **171:** 481–494.
53. Rossi D, Trifonov V, Fangazio M, *et al*. The coding genome of splenic marginal zone lymphoma: activation of *NOTCH2* and other pathways regulating marginal zone development. *J Exp Med* 2012; **209:** 1537–1551.
54. Kasar S, Kim J, Improgo R, *et al*. Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nat Commun* 2015; **6:** 8866.
55. Matteucci C, Bracci M, Barba G, *et al*. Different genomic imbalances in low- and high-grade HCV-related lymphomas. *Leukemia* 2008; **22:** 219–222.
56. Bende RJ, van Maldegem F, Triesscheijn M, *et al*. Germinal centers in human lymph nodes contain reactivated memory B cells. *J Exp Med* 2007; **204:** 2655–2665.
57. Sungalee S, Mamessier E, Morgado E, *et al*. Germinal center reentries of BCL2-overexpressing B cells drive follicular lymphoma progression. *J Clin Invest* 2014; **124:** 5337–5351.
58. van Noesel CJ, Brouns GS, van Schijndel GM, *et al*. Comparison of human B cell antigen receptor complexes: membrane-expressed forms of immunoglobulin (Ig) M, IgD, and IgG are associated with structurally related heterodimers. *J Exp Med* 1992; **175:** 1511–1519.
59. Davis RE, Ngo VN, Lenz G, *et al*. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* 2010; **463:** 88–92.
60. Winkler GS. The mammalian anti-proliferative BTG/Tob protein family. *J Cell Physiol* 2010; **222:** 66–72.
61. Krysiak K, Gomez F, White BS, *et al*. Recurrent somatic mutations affecting B-cell receptor signaling pathway genes in follicular lymphoma. *Blood* 2017; **129:** 473–483.
62. Li H, Kaminski MS, Li Y, *et al*. Mutations in linker histone genes *HIST1H1 B*, *C*, *D*, and *E*; *OCT2* (*POU2F2*); *IRF8*; and *ARID1A* underlying the pathogenesis of follicular lymphoma. *Blood* 2014; **123:** 1487–1498.
63. DelBove J, Rosson G, Strobeck M, *et al*. Identification of a core member of the SWI/SNF complex, BAF155/SMARCC1, as a human tumor suppressor gene. *Epigenetics* 2011; **6:** 1444–1453.
64. Agarwal NK, Kim CH, Kunkalla K, *et al*. Active IKKβ promotes the stability of *GLI1* oncogene in diffuse large B-cell lymphoma. *Blood* 2016; **127:** 605–615.
65. Roy D, Sin SH, Damania B, *et al*. Tumor suppressor genes FHIT and WWOX are deleted in primary effusion lymphoma (PEL) cell lines. *Blood* 2011; **118:** e32–e39.
66. Zhang Z, Tong J, Tang X, *et al*. The ubiquitin ligase HERC4 mediates c-Maf ubiquitination and delays the growth of multiple myeloma xenografts in nude mice. *Blood* 2016; **127:** 1676–1686.
67. Bonetti P, Testoni M, Scandurra M, *et al*. Deregulation of ETS1 and FLI1 contributes to the pathogenesis of diffuse large B-cell lymphoma. *Blood* 2013; **122:** 2233–2241.

References 51–67 are cited only in the supplementary material.

SUPPLEMENTARY MATERIAL ONLINE

Text S1. Detailed explanatory text for Table 3

Figure S1. Comparison of all the genes harboring non-synonymous mutations, identified in seven RF-expressing B-cell clones, with a set of 57 genes that were reported to be mutated in 321 MALT lymphomas of different primary locations in two studies

Table S1. Clinical information, salivary gland focus scores, and serology of the SS patients

Table S2. Isolation of IgM-RF B-cell clones

Table S3. Overview of the basic whole-exome sequencing data

Table S4. VH-CDR3 homology of the isolated RF-producing B-cell clones, with VH-CDR3 aa sequences of Genbank

Table S5. Overview of the exome mutations, expression of the mutated genes in GC B cells and DLBCL, and AID WRC hotspots

Table S6. Overview of non-synonymous mutations, identified in RF and TT clones, in genes that are expressed by GC B cells and/or DLBCL cells