Unique repetitive nucleic acid structures mirror switch regions in the human IgH locus

Ophelie Deze, Delfina Ordanoska, Delphine Rossille, Emma Miglierina, Brice Laffleur, Michel Cogné

▶ **To cite this version:**

Ophelie Deze, Delfina Ordanoska, Delphine Rossille, Emma Miglierina, Brice Laffleur, et al.. Unique repetitive nucleic acid structures mirror switch regions in the human IgH locus. Biochimie, 2023, 214, pp.167-175. 10.1016/j.biochi.2023.08.017 . hal-04212710

HAL Id: hal-04212710
https://hal.science/hal-04212710

Submitted on 29 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Biochimie 214 (2023) 167–175

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Unique repetitive nucleic acid structures mirror switch regions in the human IgH locus

Ophélie Dézé [a], Delfina Ordanoska [a], Delphine Rossille [b], Emma Miglierina [a], Brice Laffleur [a], Michel Cogné [a, b, *]

[a] Institut National de La Santé et de La Recherche Médicale, Unité Mixte de Recherche U1236, Université de Rennes, Etablissement Français Du Sang Bretagne, F-35000, Rennes, France
[b] Centre Hospitalier Universitaire de Rennes, SITI, Pôle Biologie, F-35033, Rennes, France

ARTICLE INFO

Article history:
Received 6 July 2023
Received in revised form
29 August 2023
Accepted 31 August 2023
Available online 5 September 2023

Handling Editor: J.L. Mergny

Keywords:
DNA repeats
G-quadruplex
Immunoglobulin genes

ABSTRACT

Immunoglobulin (Ig) genes carry the unique ability to be reshaped in peripheral B lymphocytes after these cells encounter a specific antigen. B cells can then further improve their affinity, acquire new functions as memory cells and eventually end up as antibody-secreting cells. Ig class switching is an important change that occurs in this context, thanks to local DNA lesions initiated by the enzyme activation-induced deaminase (AID). Several *cis*-acting elements of the Ig heavy (IgH) chain locus make it accessible to the AID-mediated lesions that promote class switch recombination (CSR). DNA repeats, with a *non-template* strand rich in G-quadruplexes (G4)-DNA, are prominent *cis*-targets of AID and define the so-called "switch" (S) regions specifically targeted for CSR. By analyzing the structure of the human IgH locus, we uncover that abundant DNA repeats, some with a putative G4-rich template strand, are additionally present in downstream portions of the IgH coding genes. These like-S (LS) regions stand as 3′ mirror-images of S regions and also show analogies to some previously reported repeats associated with the IgH locus 3' super-enhancer. A regulatory role of LS repeats is strongly suggested by their specific localization close to exons encoding the membrane form of Ig molecules, and by their conservation during mammalian evolution.

© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Antigen receptors are expressed from large multigenic loci, with the ability to undergo specific rearrangements and remodeling of their expression during lymphocyte differentiation. This involves multiple target recombination sites and regulatory elements spared among the corresponding coding genes. Accordingly, the primary sequence of these loci carries unique features that ensure the efficiency and specificity of these processes.

The exons encoding the antigen-binding "variable" (V) domains of the immunoglobulin (Ig) or T-cell receptor genes are composed of a combination of distant V, D and J segments. This VDJ assembly is supported by short and well-defined recombination signals that are precisely cut by the *recombination-activating gene* (RAG) complex [1].

In contrast to the highly precise RAG-mediated recombination, a much less precise type of rearrangement, known as class switch recombination (CSR), occurs specifically in activated B cells and results in deletions of variable length within the Ig heavy (IgH) chain topologically associated domain (TAD) [2,3]. Class switching hereby replaces the production of IgM with IgG, IgA or IgE, which have different effector properties. CSR relies on the dynamic folding of the IgH TAD into sub-TADs anchored to flanking insulator elements which bind the CTCF factor in a process known as loop extrusion [3,4]. Within such dynamic loops, CSR breakpoints can occur at multiple sites within large repetitive regions, each spanning several kilobases, located upstream of each IgH constant gene. In mammals, these so-called switch (S) regions also feature a non-template strand (NTS) with abundant G-rich sequences that form non-canonical four-stranded G4-DNA secondary structures [5]. S regions are located in the intron preceding the constant (C) IgH genes (Fig. 1) and are transcribed in a cytokine-dependent manner [6,7]. The specific structure and location of S regions contribute to

* Corresponding author. Institut National de La Santé et de La Recherche Médicale, Unité Mixte de Recherche U1236, Université de Rennes, Etablissement Français Du Sang Bretagne, F-35000, Rennes, France.
E-mail address: michel.cogne@univ-rennes.fr (M. Cogné).

https://doi.org/10.1016/j.biochi.2023.08.017
0300-9084/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

their function by forming R-loops and G-loops in their G4-rich portions, which facilitate AID targeting [8–12]. In addition, G4-rich RNA transcribed from S regions participates in the recruitment of the AID enzyme, which can then initiate local DNA breaks [12].

Although they vary in size and nucleotide composition among different IgH C genes and in different species, S regions all share a similar general organization. In human, they are composed of short (typically 5 bp long) tandem repeats such as GAGCT and GGGG/CT, which in turn are incorporated into longer (40–50 bp) tandem repeats [9,13–15]. These features thus differ from those of the large heterochromatin-associated α satellite DNA (composed of 171bp repeat units) which represents 10% of the human genome [16], the β -satellite DNA (arrays of 68-bp repeat units) [17] and the less organized pericentromeric SatIII repeats. Sequence homologies could classify S regions into two families, with the Sμ, Sα and Sε on one side, and the large Sγ family on the other side. However, CSR occurs abundantly between Sμ and Sγ and homology is not the driver of CSR [14]. Homology between S regions from different genes is only partial and rather than homologous recombination, CSR repair has been shown to involve non-homologous end-joining (NHEJ) along with some degree of alternate end-joining which then ultimately involves microhomology [18]. In parallel, human γ genes form a kind of multigene family that originated from a duplicated ancestral gene and they have maintained a significant homology in both their coding regions, Sγ regions and intervening sequences, although some aspects of their architecture have diverged and they ultimately encode antibodies with very different properties [19,20]. Once CSR has occurred, IgH genes can still yield two different types of transcripts due to alternative splicing and alternative usage of polyadenylation sites, so that mostly secreted type Ig chains are produced in plasma cells and mostly membrane-type Ig chains in lymphocytes (Fig. 1).

In addition to S regions, other sequences with unique structures have been reported in the IgH locus. We previously reported that the major super-enhancers that regulate the accessibility of the IgH genes to CSR, *i.e.* the 3′ regulatory regions (3′RRs) standing downstream of the IgH Cα genes (Fig. 1), also contain dense stretches of tandem repeats that resemble S regions and have been termed like-S (LS) repeats [21,22]. Although the mechanistic role of these enhancer-associated LS regions remains to be elucidated, it is noteworthy that they were found proximal to DNA double-strand breaks in some B-cells undergoing an atypical type of IgH recombination that joins the Sμ region with the 3′ super-enhancer, thereby terminating IgH locus expression, in a process thus termed “locus suicide recombination” (LSR) [21,22]. In contrast to S regions, the 3′RR-associated LS regions contain neither G-rich DNA nor G4-DNA on the NTS, but eventually on the template strand (TS).

To more comprehensively characterize the presence of unique structures and in particular DNA repeats within Ig genes, we extended our search for similarly organized elements at other positions of human Ig loci. We now show that, with the exception of the most upstream Cμ and Cδ IgH genes, S regions preceding C genes are systematically echoed by other LS repetitive elements proximal to the last exons of each human IgH gene, with structural analogies to S regions. Since evolutionary conservation is usually correlated with functionality, we also compared human IgH γ genes with those from other mammals, uncovering other cases with dyad symmetries very similar to those found in human genes.

2. Materials and methods

IgH gene sequences were analyzed using the various online available algorithms listed below.

Sequences from coding regions were first obtained from the IMGT database (https://www.imgt.org, last accessed July 5, 2023). Sequences from exons were blasted on the ENSEMBL website (https://www.ensembl. org last accessed July 5, 2023) to identify the largest possible genomic fragments including the complete IgH coding genes.

After precise annotation of exons in the genomic fragments, these sequences were searched for direct or inverted repeats using the YASS similarity search algorithm [34] at https://bioinfo.univ-lille.fr/yass/yass.php (last accessed July 4, 2023), using a minimum window of 20bp and checking for more than 75% identity with less than 10% of indels. The references and coordinates of all Ig sequences analyzed are listed in Table 1.

Sequences were also analyzed with the G4-hunter algorithm (https://bioinformatics.cruk.cam.ac.uk/G4Hunter/last accessed July 5, 2023) to identify potential G-quadruplexes using 20bp windows and a stringency score of 1.5 (1.2 for analyses of rodent sequences).

Repetitive motifs in the LSγ regions were searched both by manual alignment of sequences and by processing them with the MEME algorithm (https://meme-suite.org/meme/doc/fimo.html, last accessed July 5, 2023).

Transcription of the M1-M2 region for the different human IgH C genes in RNAseq datasets from normal human B cells was performed using the Integrative Genomics Viewer (IGV) software [35]. Reads were sorted from the TS or the NTS (*i.e.* coding strand). Analyzed strand-specific RNAseq datasets were from GSE114816 and GSE114803 [36]. Statistical analysis was performed using a paired *t*-test.

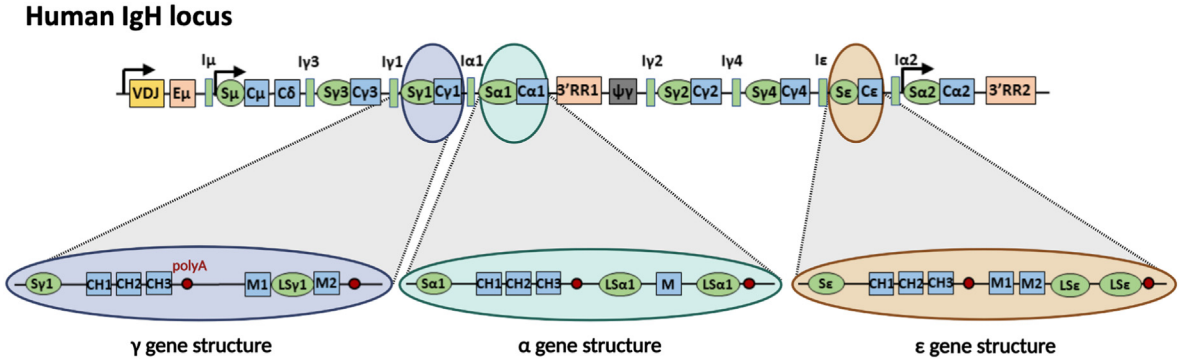

Fig. 1. Schematic of the human IgH locus, showing the sequential order of the C genes (*top*), each preceded by an S region, and the positions of the major 3′RR superenhancers promoting CSR, together with the positions of the polyadenylation sites (polyA, red circles) used alternatively in B lymphocytes and plasma cells to produce membrane of secreted IgH transcripts, respectively (*bottom*).

Table 1
References and coordinates of Ig sequences analyzed in this report.

Target sequence	Genbank reference	start	end
human IgH μ chain	NG_001019.6	1,017,345	1,029,344
human IgH δ chain	NG_001019.6	1,031,835	1,043,894
human IgH γ3 chain	NG_001019.6	1,105,125	1,117,124
human IgH γ1 chain	NG_001019.6	1,132,344	1,144,343
human IgH α1 chain	NG_001019.6	1,167,031	1,179,030
human IgH pseudo-γ chain	NG_001019.6	1,207,709	1,217,708
human IgH γ2chain	NG_001019.6	1,231,399	1,241,398
human IgH γ4 chain	NG_001019.6	1,250,435	1,260,434
human IgH ε chain	NG_001019.6	1,275,456	1,287,455
human IgH α2 chain	AL928742.3	65,358	75,357
human Ig kappa locus	NG_000834	418,771	509,506
human Ig lambda locus	GRCh38 (Chr22)	Chr22:22,875,000	Chr22:22,945,000
Lemur catta IgH γ1	BK063711.1	479,065	489,084
Gorilla gorilla IgH γ1	BK063607.1	1,131,821	1,142,002
Bos taurus γ3 γ1	OX344710.1	74,975,786	74,992,032
Canis lupus 1st IgH γ gene (genome assembly chr 8)	HG994390.1	2,662,661	2,672,273
Canis lupus 2nd IgH γ gene (genome assembly chr 8)	HG994390.1	2,698,259	2,708,255
Canis lupus 3rd IgH γ gene (genome assembly chr 8)	HG994390.1	2,736,469	2,746,485
Canis lupus 4th IgH γ gene (genome assembly chr 8)	HG994390.1	2,767,165	2,777,162
Mus musculus (strain 129S1) γ1 IgH gene	AJ851868.3	1,506,838	1,526,841
Rattus norvegicus γ1 IgH gene	mRatBN7.2 primary assembly	chr6:132,393,999	chr6:132,382,000
Gallus gallus (chicken) IgY gene	NC_052566.1 (chr 35)	189,977	219,976
Anas platyrhynchos (Duck) IgH α gene	AJ314754	38,952	20,744

3. Results and discussion

Heterochromatin is known for its abundance of repetitive DNA, which is notably responsible for the binding of the HP1 factor to some G4-rich DNA repeats [23]. The recent release of complete human genome assemblies derived from long-read sequencing has confirmed the abundance of repetitive DNA, accounting for approximately 54% of the human genome, and has further suggested its association with heterochromatin [24,25]. Interestingly repetitive DNA from heterochromatin can also correlate with the position of boundaries demarcating TADs and sub-TADs and may be involved in differential genome organization in different cell types [25]. In contrast, DNA repeats are rarely present in human coding genes that define euchromatin, so the architecture of the IgH locus and S regions is quite specific. To verify the unique structure of the IgH S regions, we systematically evaluated the genomic context of the Ig C genes by building dot matrices with the YASS similarity search algorithm, using a minimal window range at 20bp and checking identities above 75% with less than 10% of indels. This was done first for the Cμ (IGHM) gene, which is the first expressed in naïve B-cells. As expected, the IGHM sequence analysis revealed the presence of dense tandem repeats (*in green*) spanning approximately 4 kb and strictly confined to the region preceding the first Cμ exon (CH1), corresponding to Sμ region (Fig. 2A, top). In parallel, the same sequence including the [Sμ - Cμ gene] sequence was analyzed for the presence of putative G4-forming sequences using the G4Hunter algorithm (using a 25bp window and the stringent score of 1.5). This confirmed the skewed composition of Sμ, with putative G4-forming sequences particularly abundant in the core Sμ and strictly carried by the non-template strand (NTS) (Fig. 2A, bottom). A short stretch of low-density DNA repeats downstream of the CH4 exon showed no enrichment in G4-DNA. However, some G4 imbalance is evident along the gene, with an upstream G4-rich NTS for Sμ preceding the coding exons, and a reciprocal imbalance downstream of CH4, where putative G4-forming sequences tend to be on the TS (Fig. 2A).

To test whether the inclusion of DNA repeats is a general feature of Ig genes, we performed the dot matrix analysis for the human IgH δ (IGHD) gene (Fig. 2B), which is also expressed in naive B-cells, and for the Ig light chain κ (IGKC) and λ (IGLC) genes (Suppl Fig. 1). As expected, this did not reveal any repetitive portion resembling either an S region. For the δ gene, this is consistent with the known physiology of the δ H chain expression, which is known to result mainly from alternate splicing of long μ-δ primary transcripts, with rare DNA breaks upstream of Cδ at non-repetitive sites [26]. No repetitive region appears near the Cκ gene or its associated Eκ and 3′κ transcriptional enhancers, except for the array of homologous Jκ segments (Suppl Fig. 1A). The same was true for the Igλ locus, except for a short (0.5 kb long) stretch of G4-rich repeats located amidst the 3′λ enhancers (Suppl Fig. 1B), which minimally resembles the repeats interspersed between the IgH 3′RR core enhancers. Extending the analysis to a few randomly chosen non-Ig genes, we also found no organized repetitive structure resembling an S region in the genomic regions encompassed by the human MYC, BCL2, KRAS and MALT1 oncogenes or the chicken ovalbumin gene (not shown).

When carrying the same analysis on the downstream IgH C genes α1 (IGHA1), α2 (IGHA2) and ε (IGHE), which are expressed after B-cell activation and in memory B-cells, a different picture emerged. All three human genes are characterized by the presence of dense tandem DNA repeats (*in green on the figure*) not only upstream of the coding region and corresponding to the S regions targeted by CSR, but also downstream of the coding unit (Fig. 3). Each of these genes appears followed by two blocks of LS repeats. The IGHA1 and IGHA2 genes feature a first LS block quite similar to an S region in terms of G4-richness on the NTS and preceding the last exon (Fig. 3A and B), an alternatively spliced membrane (M) exon which encodes the membrane-anchored form of IgA molecules when expressed as B cell receptors on the membrane of switched B-cells [27,28]. In both human α genes, a second array of dense LS repeats follows the M exon and spans about 2 kb for the α1 gene and 1 kb for the α2 gene. In both cases, the most downstream array of LS repeats is devoid of G4 structures on either strand, although its initial part interestingly carries homology to Sα, but in reverse orientation (showing as *red lines* on the dot matrix) (Fig. 3A and B). The human ε gene somehow shows a similar architecture, with two blocks of LS repeats downstream of the coding sequence, with a G4-rich NTS only for the first one. These ε LS repeats are also

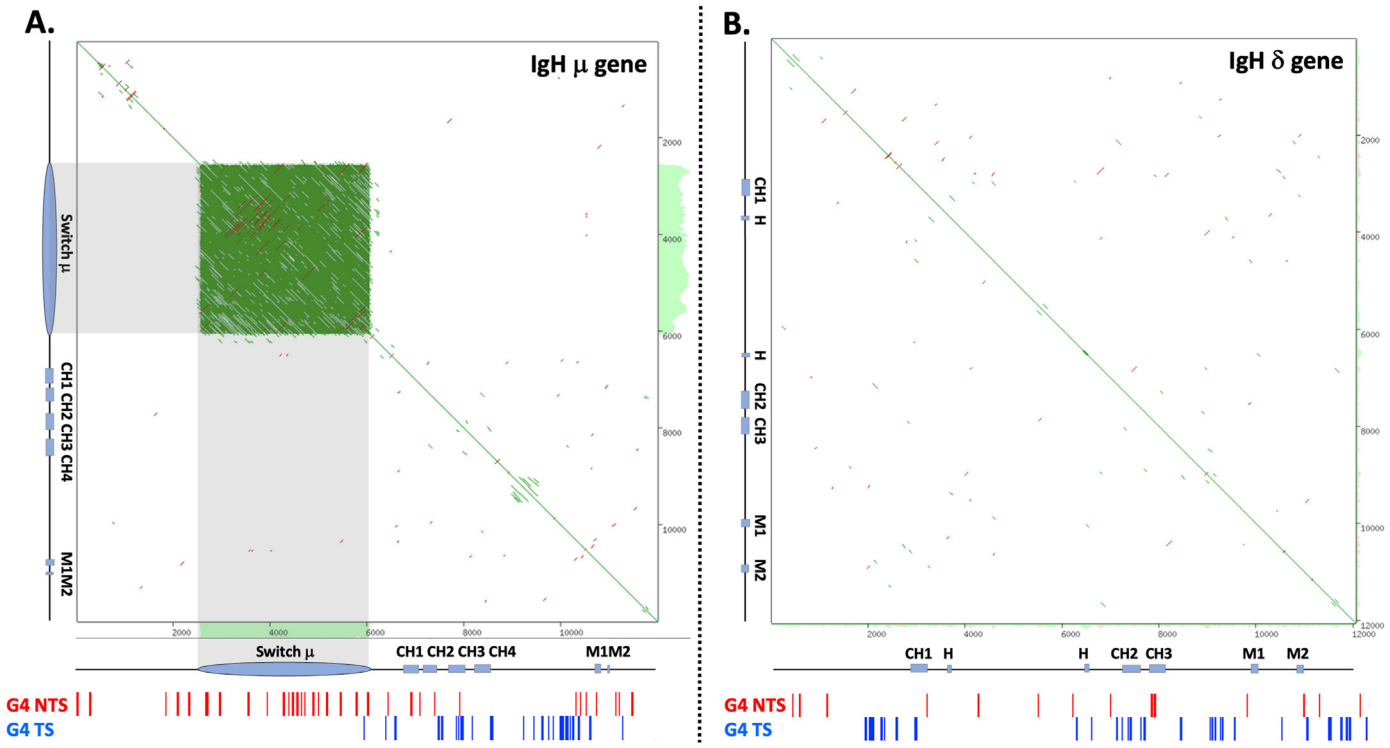

Fig. 2. A. Dot plot matrix analysis of the human IgH μ (IGHM) gene using the YASS algorithm showing short tandem repeats in green and inverted repeats in red. G4-DNA as detected by G4Hunter on the NTS strand (red) and the TS (blue) is shown at the bottom with bars located at the start position of each G4 motif with a width proportional to the motif length. B. The same analysis of DNA repeats and of G4 motifs as in A is shown for the IgH δ (IGHD) gene.

in proximity to the alternatively spliced exons encoding the membrane-anchored IgE BCR (Fig. 3C).

Human γ genes include four functional γ1, γ2, γ3 and γ4 genes, encoding the heavy chain of IgG1 through IgG4, and the Ψγ pseudogene. Analysis of their sequences (using the same criteria as above) evidenced the expected and dense G4-rich repeats which mark S regions (Fig. 4), upstream of the 4 functional coding genes, while Ψγ lacks such an S region (Suppl Fig. 2). More unexpectedly, the 3' part of all five genes revealed a unique and shared architecture, featuring very dense arrays of LS repeats that span variable distances (from 0.7 kb for γ2 and γ4, to 2 kb for γ1 and Ψγ). Strikingly, the LSγ repeats stand as mirror images of Sγ regions: while similar in length and structure to Sγ, they display an inverse imbalance in composition, with G4-richness on the template strand, together with partial homologies to Sγ in terms of inverted repeats (see *red lines* on the dot matrix alignments from Fig. 4, most notably for γ1).

Detailed analysis revealed that the LSγ repeats contained multiple occurrences of the 5bp TCCCC sequence, which in turn were embedded within higher order DNA repeats. This higher order architecture could be aligned as consisting of repeated 32–34 bp motifs (including 3 regularly spaced TCCCC motifs, as shown in Fig. 5A. Using the MEME consensus sequence identification algorithm, we accordingly obtained a 34bp consensus sequence (Fig. 5B**).** A longer (49bp) consensus motif, then consisting of 4 regularly spaced TCCCC motifs, could also be defined using the MEME algorithm (Fig. 5C). The latter motif is very similar in length and structure to an S region. This strong analogy to an S region may suggest either an evolutionary or a functional link between the Sγ and LSγ elements. In addition, since CSR strongly relies on TADs and loop extrusion, and since LSγ includes several potential binding sites for the CCCTC-binding factor CTCF which anchors loop extrusion, it is tempting to speculate that the structure of LSγ may contribute to the formation of sub-TADs during the CSR process.

Since the human γ genes arose from a common ancestral gene after several duplication events [19,20], their common organization among human isotypes does not come as a surprise and does not necessarily imply that the conserved elements have a functional role. To further evaluate whether the dyad symmetry structure of γ genes is a more general feature conserved during genome evolution, we examined γ gene sequences in other primate species for which such sequences are available: gorilla and lemur, the latter being the most distant from humans, with approximately 50 million years of divergence [29]. In both cases, the downstream parts of γ genes reveal the presence of DNA repeats standing in the M1-M2 intron (in-between those exons that encode the membrane-anchored γ chain), with clear G4 richness on the TS (Suppl Fig. 3), as in humans.

Going further back in the phylogeny, we analyzed other γ genes for which complete genomic sequences of the coding exons have been determined in non-primate mammals such as dogs, cattle and rodents, which then followed divergent branches of the phylogeny for about 80 million years [29].

The architecture of the four canine γ genes (Suppl Fig. 4) appears strikingly similar to that described above for human genes. The LSγ regions are precisely intercalated between the M1 and M2 terminal exons and stand as mirror images of the Sγ regions both in terms of length, repeat density and G4-richness concentrated at the TS. The analysis of bovine genes gives a different picture since only one of them (encoding the γ1 chain of bovine IgG1) follows an architecture similar to primate and dog genes with a long (>3 kb) LSγ1 stretch of repeats featuring G4 motifs on the TS (Suppl Fig. 5). In contrast, the γ3 gene (Suppl Fig. 5) and the γ2 gene (not shown) have only minimal LS regions within a short (≈160bp) M1-M2 intron, featuring a single stringent G4 motif on the TS together with some homology with the LSγ1 repeats. An intermediate

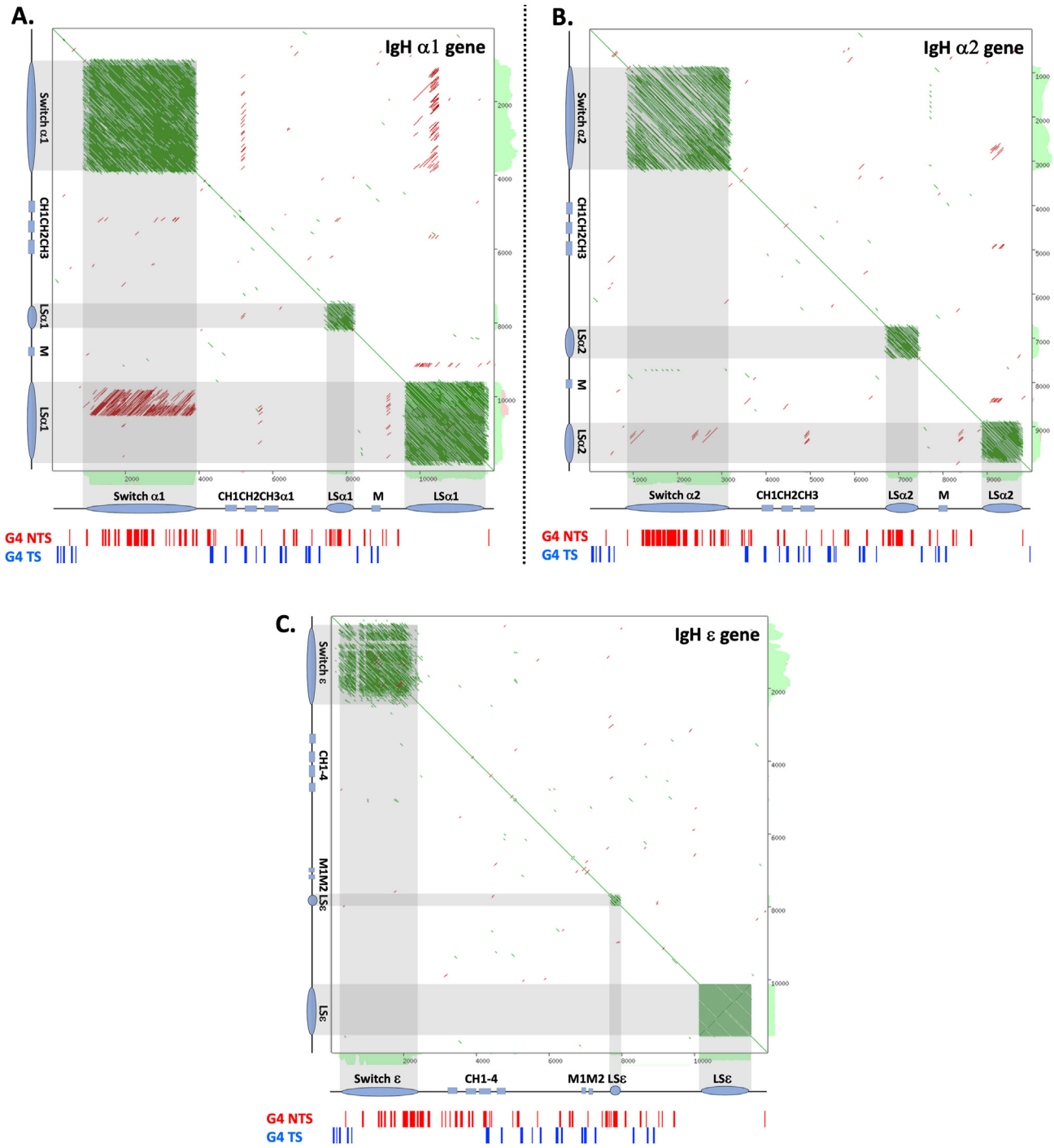

Fig. 3. A. Dot plot matrix analysis of the human IgH α1 (IGHA1) gene using the YASS algorithm showing short tandem repeats in green and inverted repeats in red. G4-DNA as detected by G4Hunter on the NTS strand (red) and the TS (blue) is shown at the bottom with bars located at the start position of each G4 motif with a width proportional to the motif length. B. The same analysis of repeats and G4 motifs as in A is shown for the IgH α2 (IGHA2) gene. C. The same analysis as in A is shown for the IgH ε (IGHE) gene.

architecture is observed in rodents, where both mouse and rat γ genes have short M1-M2 introns (culminating at 0.8 kb for their γ1 genes), but also show a rudimentary LS organization with a short stretch of DNA repeats and some G4 motifs on the TS (Suppl Fig. 6). Finally, analysis of the sequences available for avian Ig genes, *i.e.* the duck α IgH gene and the chicken ν H chain gene (constituting chicken IgY), reveals the puzzling picture of widespread DNA repeats not only upstream of the CH1 exon, probably corresponding to S regions targeted by CSR, but also in all the introns with G4-richness only seen in the chicken gene (Suppl Fig. 6).

Although it varies in terms of length and density of repeats, the above evaluated mammal γ genes share a similar dyad symmetry architecture and this phylogenetic conservation suggests that it can serve a functional role. It now remains to be explored whether such a role relates to gene expression, accessibility to transcription, replication or other processes at the DNA level, or whether it impacts the processing or the alternative splicing of primary transcripts carrying such repetitive introns. The length of S regions has

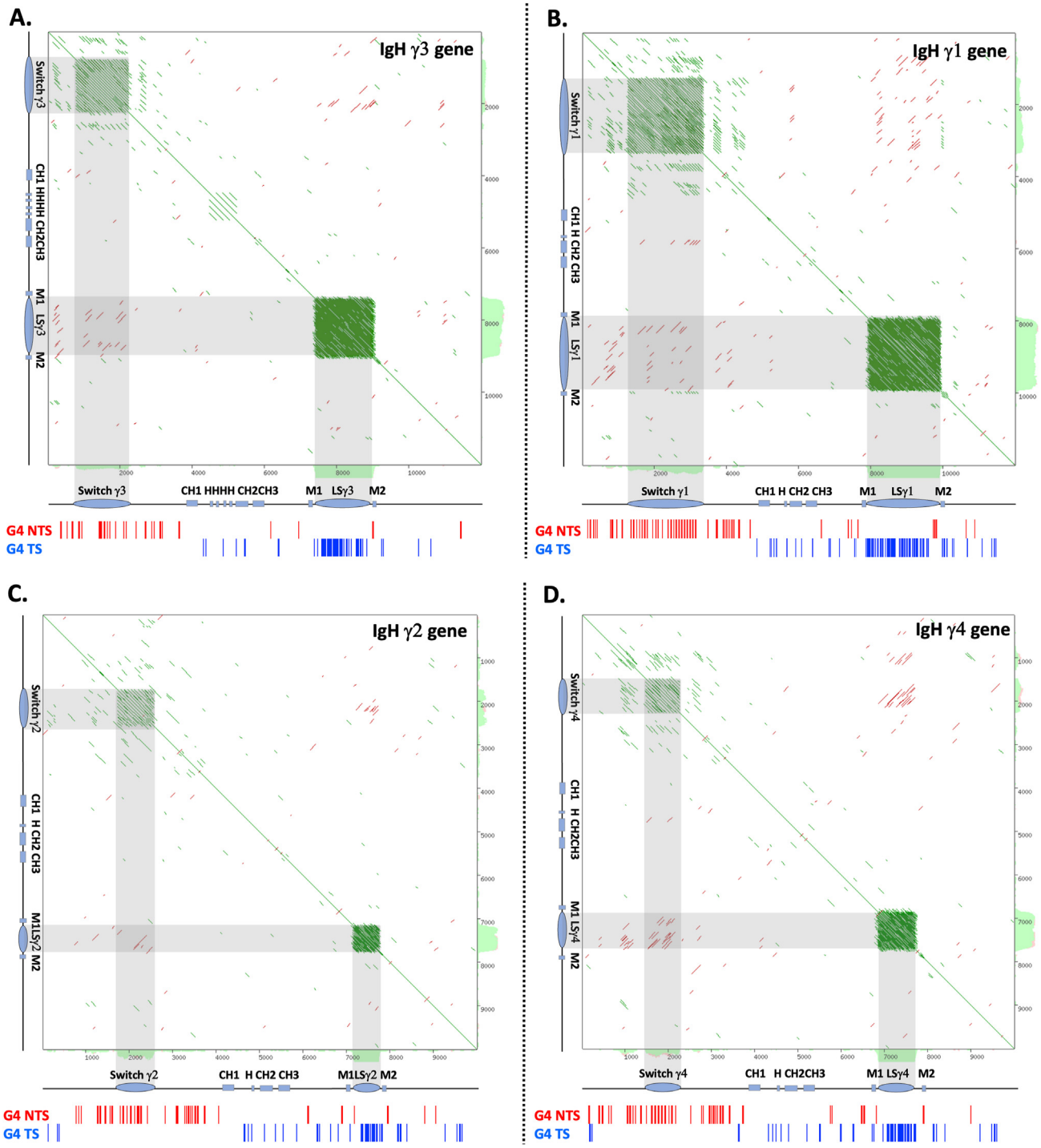

Fig. 4. Analysis of human IgH γ genes with the YASS algorithm showing short tandem repeats in green *vs* inverted repeats in red. G4-DNA as detected by G4Hunter on the NTS strand (red) and the TS (blue) is shown at the bottom with bars located at the start position of each motif with a width proportional to length. The analysis is from the most 5′ to the most 3′ gene: A. γ3 gene (IGHG3). B. γ1 gene (IGHG1). C. γ2 gene (IGHG2). D. γ4 gene (IGHG4).

been positively correlated to CSR efficiency by Alt and coll [30]; whether the length and structure of LS regions can reveal additional correlations will remain to be determined. Interestingly, the long and well-organized LS structure that we describe above for the bovine γ1 gene corresponds to the most expressed Ig gene in cattle (IgG1 representing more than half of all Ig present in bovine serum). For mammalian genes, the double symmetry of DNA repeats and G4-DNA on the TS *vs* the NTS, is a specific feature of γ genes, not found for the ε and α genes, making it possible that both aspects the DNA structures carry differential roles. G4-DNA has been characterized with multiple regulatory roles notably in promoter regions, regulating transcription [31], but also within introns where they can regulate alternate splicing [32].

Since the processing of S regions before DNA breaks involves R-loops associated with G4 on the NTS, we wondered whether antisense transcription could occur at the level of human LSγ regions

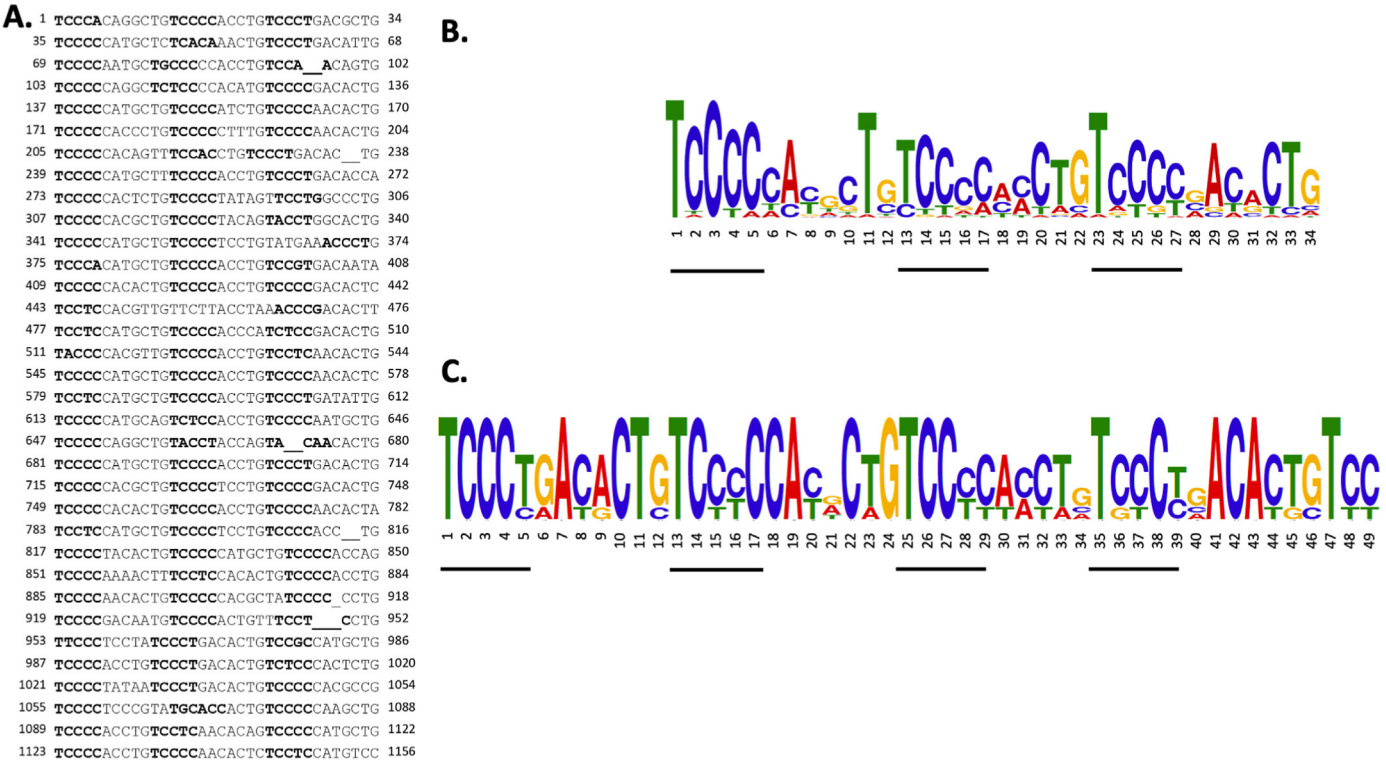

Fig. 5. A. A typical portion of the LSγ1 sequence (Chr14:105738848–105739991, as from human GRCh38 release, reverse orientation) is shown aligned according to its internal 34bp rhythm of partially homologous tandem repeats. B and C. The same LSγ1 region as in A was submitted for repeat motif using the MEME algorithm, yielding either a 34bp consensus motif as shown in (B) or a 49 bp consensus motif as shown in (C) (size of the displayed nucleotides is proportional to their occurrence).

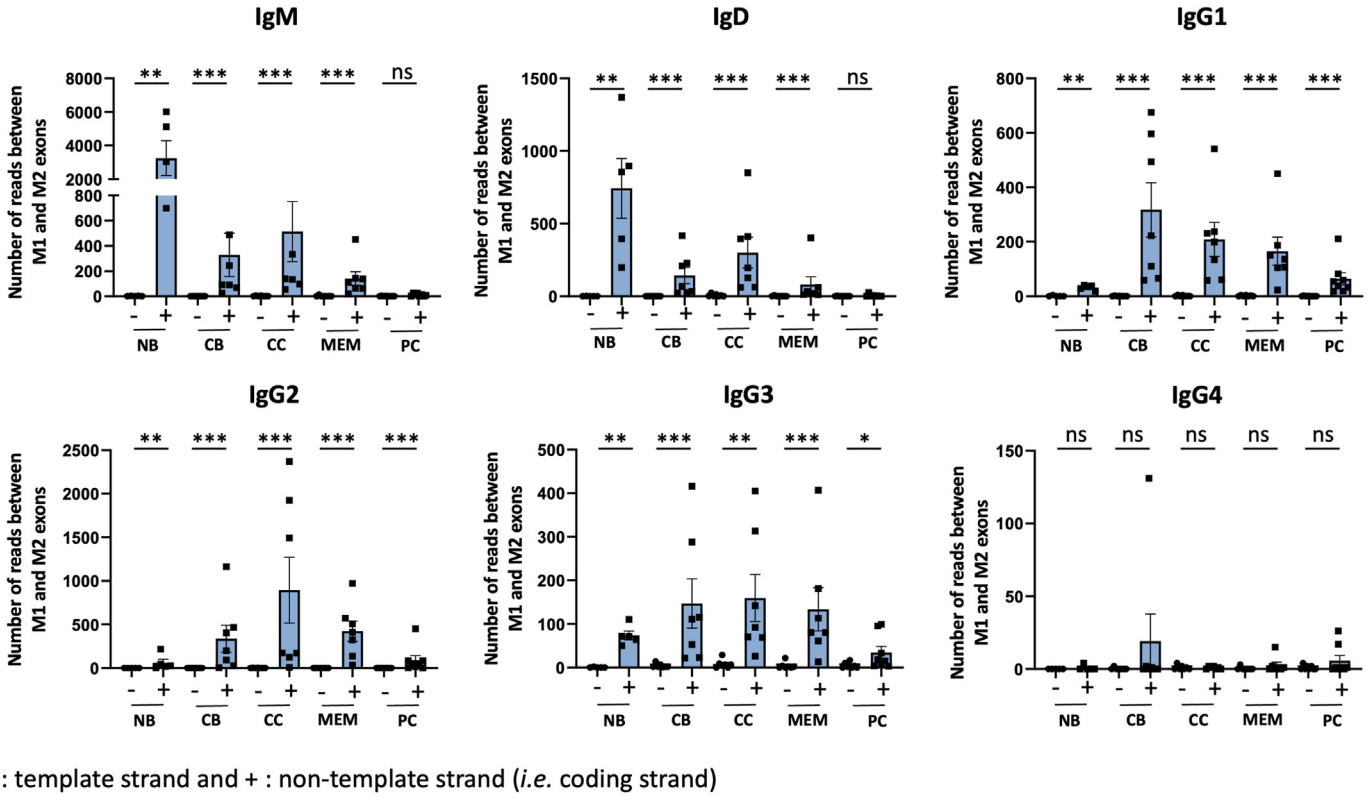

Fig. 6. Transcription of the M1M2 region was quantified for the different human IgH C genes in B cells from healthy donors: naive B cells (NB), centroblasts (CB), centrocytes (CC), memory B cells (MEM), plasma cells (PC) from either tonsil or bone marrow. Reads are either on the TS (−) or the NTS (*i.e.* coding strand) (+). Statistical analysis was performed using a paired *t*-test. RNAseq datasets analyzed are from GSE114816 and GSE114803 [36].

and also favor the occurrence of R-loops. However, by analyzing strand-specific datasets of RNAseq from human activated B cells, we observed that primary transcripts detectable in human B cells at various activation stages almost exclusively include sense transcripts (Fig. 6). Thus, our data do not support the formation of R-loops at the positions of LS repeats, while these antisense RNAs may be produced but quickly degraded by the RNA exosome complex [33], thus requiring further investigation.

4. Conclusion

In conclusion, the present study uncovers an unanticipated organization of the human IgH downstream genes, featuring mirror images of S regions (LS regions) following the main body of the coding unit, proximal to its terminal exons encoding the transmembrane part of class-switched BCRs. The conspicuous position of LS elements in symmetry with S regions suggests a role for CSR, which could be related in particular to the dynamic evolution of sub-TADs and loop extrusion, to regulate the accessibility and the folding of an adjacent S region, either promoting its accessibility or, conversely, protecting it from unwanted breaks. Because of their structural similarities to S regions, one can also wonder whether LS regions themselves could be targeted by recombination events, but there is currently no description of such DNA breaks in physiology and this would be unlikely to favor Ig production. Alternatively to a role in DNA recombination, the precise location of LS elements within introns may suggest a role in the regulation of alternative splicing, either promoting the insertion of M1 and M2 exons into IgH transcripts in lymphocytes, or negatively regulating their usage in plasma cells which predominantly produce secreted-type IgH transcripts. Finally, a role in mRNA decay is an additional possibility that needs to be explored.

Mutational strategies in the mouse genome do not seem attractive for addressing the role of structures that have been expanded in humans and in several other mammals but show a more rudimentary organization in rodents. Elucidating the potential role of this unique architecture of IgH C genes in the physiology of B lymphocyte differentiation and activation, will therefore be a future challenge.

Author contributions

OD (collection of data; analysis of data; writing of manuscript); DO (collection of data); DR (analysis of data); EM (collection of data); BL (collection of data; analysis of data; writing of manuscript); MC (conception of the work; collection of data; analysis of data; writing of manuscript).

Acknowledgments

This work was supported by grants from Agence Nationale pour la Recherche (grant AID-G4-CSR to MC), Association pour la Recherche sur le Cancer (grant ARCPJA2021060003753 to BL), and Ligue contre le Cancer.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biochi.2023.08.017.

References

[1] D. Jung, F.W. Alt, Unraveling V(D)J recombination; insights into gene regulation, Cell 116 (2004) 299–311.
[2] D.D. Dudley, J. Chaudhuri, C.H. Bassing, F.W. Alt, Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences, Adv. Immunol. 86 (2005) 43–112.
[3] X. Zhang, H.S. Yoon, A.M. Chapdelaine-Williams, N. Kyritsis, F.W. Alt, Physiological role of the 3'IgH CBEs super-anchor in antibody class switching, Proc. Natl. Acad. Sci. U.S.A. 118 (2021) e2024392118.
[4] X. Zhang, et al., Fundamental roles of chromatin loop extrusion in antibody class switching, Nature (2019), https://doi.org/10.1038/s41586-019-1723-0.
[5] O. Dézé, B. Laffleur, M. Cogné, Roles of G4-DNA and G4-RNA in class switch recombination and additional regulations in B-lymphocytes, Mol. Basel Switz. 28 (2023) 1159.
[6] S. Lutzker, P. Rothman, R. Pollock, R. Coffman, F.W. Alt, Mitogen- and IL-4-regulated expression of germ-line Ig gamma 2b transcripts: evidence for directed heavy chain class switching, Cell 53 (1988) 177–184.
[7] H. Leung, N. Maizels, Regulation and targeting of recombination in extrachromosomal substrates carrying immunoglobulin switch region sequences, Mol. Cell Biol. 14 (1994) 1450–1458.
[8] K. Yu, F. Chedin, C.-L. Hsieh, T.E. Wilson, M.R. Lieber, R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells, Nat. Immunol. 4 (2003) 442–451.
[9] J. Chaudhuri, et al., Evolution of the immunoglobulin heavy chain class switch recombination mechanism, Adv. Immunol. 94 (2007) 157–214.
[10] J. Chaudhuri, et al., Transcription-targeted DNA deamination by the AID antibody diversification enzyme, Nature 422 (2003) 726–730.
[11] W.T. Yewdell, et al., A hyper-IgM syndrome mutation in activation-induced cytidine deaminase disrupts G-quadruplex binding and genome-wide chromatin localization, Immunity 53 (2020) 952–970.e11.
[12] S. Zheng, et al., Non-coding RNA generated following lariat debranching mediates targeting of AID to DNA, Cell 161 (2015) 762–773.
[13] T.H. Rabbitts, A. Forster, W. Dunnick, D.L. Bentley, The role of gene deletion in the immunoglobulin heavy chain switch, Nature 283 (1980) 351–356.
[14] W. Dunnick, G.Z. Hertz, L. Scappino, C. Gritzmacher, DNA sequences at immunoglobulin switch region recombination sites, Nucleic Acids Res. 21 (1993) 365–372.
[15] J. Chaudhuri, F.W. Alt, Class-switch recombination: interplay of transcription, DNA deamination and DNA repair, Nat. Rev. Immunol. 4 (2004) 541–552.
[16] S.M. McNulty, B.A. Sullivan, Alpha satellite DNA biology: finding function in the recesses of the genome, Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol. 26 (2018) 115–138.
[17] J.S. Waye, H.F. Willard, Human beta satellite DNA: genomic organization and sequence definition of a class of highly repetitive tandem DNA, Proc. Natl. Acad. Sci. U.S.A. 86 (1989) 6250–6254.
[18] T. Saha, D. Sundaravinayagam, M. Di Virgilio, Charting a DNA repair roadmap for immunoglobulin class switch recombination, Trends Biochem. Sci. 46 (2021) 184–199.
[19] T.H. Rabbitts, A. Forster, C.P. Milstein, Human immunoglobulin heavy chain genes: evolutionary comparisons of C mu, C delta and C gamma genes and associated switch sequences, Nucleic Acids Res. 9 (1981) 4509–4524.
[20] J.G. Flanagan, T.H. Rabbitts, Arrangement of human immunoglobulin heavy chain constant region genes implies evolutionary duplication of a segment containing gamma, epsilon and alpha genes, Nature 300 (1982) 709–713.
[21] S. Péron, et al., AID-driven deletion causes immunoglobulin heavy chain locus suicide recombination in B cells, Science 336 (2012) 931–934.
[22] I. Dalloul, et al., Locus suicide recombination actively occurs on the functionally rearranged IgH allele in B-cells from inflamed human lymphoid tissues, PLoS Genet. 15 (2019) e1007721.
[23] R.J. Roach, et al., Heterochromatin protein 1α interacts with parallel RNA and DNA G-quadruplexes, Nucleic Acids Res. 48 (2020) 682–693.
[24] S.J. Hoyt, et al., From telomere to telomere: the transcriptional and epigenetic state of human repeat elements, Science 376 (2022) eabk3112.
[25] S.A. Haws, Z. Simandi, R.J. Barnett, J.E. Phillips-Cremins, 3D genome, on repeat: higher-order folding principles of the heterochromatinized repetitive genome, Cell 185 (2022) 2690–2707.
[26] P. Rouaud, et al., Elucidation of the enigmatic IgD class-switch recombination via germline deletion of the IgH 3' regulatory region, J. Exp. Med. 211 (2014) 975–985.
[27] M. Cogné, J.L. Preud'homme, Gene deletions force nonsecretory alpha-chain disease plasma cells to produce membrane-form alpha-chain only, J. Immunol. Baltim. Md 145 (1950) 2455–2458, 1990.
[28] I. Leduc, M. Drouet, M.C. Bodinier, A. Helal, M. Cogné, Membrane isoforms of human immunoglobulins of the A1 and A2 isotypes: structural and functional study, Immunology 90 (1997) 330–336.
[29] S. Álvarez-Carretero, et al., A species-level timeline of mammal evolution integrating phylogenomic data, Nature 602 (2022) 263–267.
[30] A.A. Zarrin, M. Tian, J. Wang, T. Borjeson, F.W. Alt, Influence of switch region

length on immunoglobulin class switch recombination, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 2466–2470.
[31] I. Georgakopoulos-Soares, et al., High-throughput characterization of the role of non-B DNA motifs on promoter function, Cell Genomics 2 (2022) 100111.
[32] I. Georgakopoulos-Soares, et al., Alternative splicing modulation by G-quadruplexes, Nat. Commun. 13 (2022) 2404.
[33] B. Laffleur, et al., Noncoding RNA processing by DIS3 regulates chromosomal architecture and somatic hypermutation in B cells, Nat. Genet. 53 (2021) 230–242.
[34] L. Noé, G. Kucherov, YASS: enhancing the sensitivity of DNA similarity search, Nucleic Acids Res. 33 (2005) W540–W543.
[35] J.T. Robinson, et al., Integrative genomics viewer, Nat. Biotechnol. 29 (2011) 24–26.
[36] X. Agirre, et al., Long non-coding RNAs discriminate the stages and gene regulatory states of human humoral immune response, Nat. Commun. 10 (2019) 821.