UNIVERSITY OF LEEDS

This is a repository copy of *Insertional diversification of the immunoglobulin heavy chain in chronic lymphoproliferative disorders*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/226696/

Version: Accepted Version

Article:
Fisher, A., Shingles, J., Newton, D. orcid.org/0000-0002-0214-1486 et al. (1 more author) (2025) Insertional diversification of the immunoglobulin heavy chain in chronic lymphoproliferative disorders. British Journal of Haematology. ISSN 0007-1048

https://doi.org/10.1111/bjh.20139

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Insertional Diversification of the Immunoglobulin Heavy Chain in Chronic Lymphoproliferative Disorders

Authors

Amelia Fisher[1,2], Jane Shingles[2,3], Darren Newton[1], Reuben Tooze[1,2,3]

[1]Leeds Institute of Medical Research, University of Leeds, Leeds, UK.

[2]Leeds Teaching Hospitals NHS Trust, Leeds, UK.

[3]Leeds Haematological Malignancy Diagnostic Service, Leeds, UK.

Insertional diversification of the immunoglobulin heavy chain (IgH) switch and variable regions is an unconventional antibody diversification mechanism described as contributing to the generation of broadly neutralising antibodies against *P. falciparum* malaria; IgH *LAIR1* gene insertions can create extra protein loops, resulting in viable antibodies specific to malarial antigens(1). Insertions of genomic sequences from outside IgH into the B-cell receptor (BCR) variable or IgH isotype switch regions(2) have since been demonstrated to occur in up to 1 in 10^4 B-cells in polyclonal populations(3), but to our knowledge not within clonal populations of chronic B lymphoproliferative disorders.

Binding of IgH to antigen is central in BCR signalling, contributing to B-cell proliferation, clonal expansions and lymphomagenesis. Insertional diversification most likely occurs during VDJ gene segment recombination, somatic hypermutation and class-switch recombination, which underpin B-cell generation of diversity, affinity maturation and immunoglobulin class switching (3, 4). Insertions maintaining the open-reading frame will alter amino acid sequence, changing BCR antigenic recognition (2) impacting on BCR and reactive antigen specificity. Stereotyped BCR specificities are linked to pathogenesis and disease risk in CLL(5), and there is emerging evidence that IgH/BCR VDJ (IGHV) mutational status can affect small molecule inhibitor treatment responses(6). Therefore, the question of whether insertional events could contribute to clonal lymphoproliferative disorders and their response to treatment is important to address.

Insertional diversification requires DNA breaks allowing insertion of sequences of several hundred base pairs from outside the IgH locus(7), usually from other chromosomes(2, 3). Emerging evidence suggests that inserts occur through acceptor sites generated by enzymes RAG or AID, and chromatin structures such as RNA/DNA hybrids called R-loops(3, 8), and can occur in naïve, memory or activated B-cells(3).

We present an exploratory analysis into whether insertional diversification events occur in cLPDs, which are hypothesised to have antigenic drive(4) and could recapitulate the process seen in *LAIR1* insertion public antibody generation.

Extended methods are in supplementary file 1 (S1). Patient samples and sequencing data was analysed in accordance with ERIC recommendations using BIOMED2 primers(5, 9). Genome version Hg19 (GRCh37) was used.

Routine cLPD IGHV electrophoresis results were screened for potential insertional events, selecting samples with longer than normal amplicons or multiple bands for analysis. Raw FASTA consensus sequences were queried in IMGT V-QUEST(10). Sequence sections not aligning to VDJ segments were copied and queried again with IMGT. If they failed to align the sequence was queried using

Basic Local Alignment Tool (BLAT). This enabled identification of both biallelic VDJ rearrangements(11), and insert sequences. IGHV electrophoresis results from patients that lacked longer amplicons or additional bands were used as controls and evaluated using the same process.

The switch region is not routinely sequenced in clinical practice and a searchable consensus map was not available. Therefore, a consensus map (supplementary file 2) was developed enabling identification of 'hotspots' for insertion events from published data and *in-silico* primer design for next generation sequencing (NGS).

A range of B-cell cLPD diagnoses and controls were selected for switch region targeted NGS, table 1.

Diagnosis	**Number of Samples**
Chronic lymphocytic leukaemia (CLL)	6
CD5 negative lymphoproliferative disorder	7
Marginal zone lymphoma (MZL)	5
Lymphoplasmacytic lymphoma*	1
Diffuse Large B-cell Lymphoma (DLBCL)	1
CLL and MZL dual diagnosis	1
Control type	**Number of samples**
CD19 negative PBMCs**	1
HeLa Cells**	1
Total B-cells D6***	1
Class switched B-cells D6	1
Non-class switched B-cells D6	1

Table 1. Patient and control samples for switch region PCR. *Confirmed MYD88 mutated, also known as Waldenström's Macroglobulinaemia. **Non-B-cell controls. ***D6 – day 6 cultured B-cells. These samples were not linked to IGHV cases.

Bioinformatic pipeline details are in S1.

819 IGHV PCR gel results were screened, characteristics of the 64 selected samples are shown in figure 1. Extended results are in supplementary file 3 (S3). 18 control samples showed no IGHV anomalies.

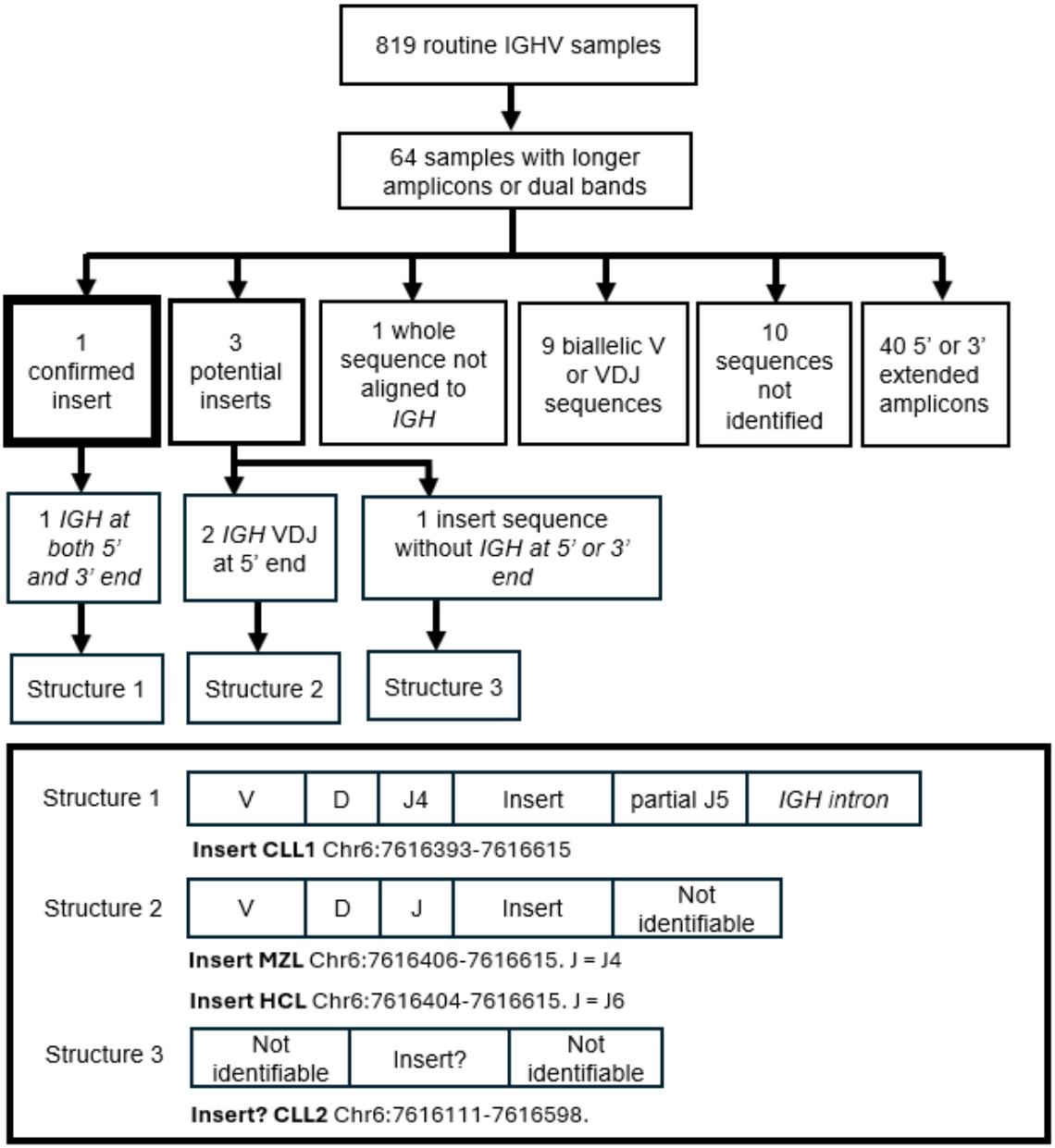

Figure 1. Identification of long amplicon sequences, outside of the reported IGHV sequence. FASTA sequences are available in S3. Structure 1 is a confirmed insert, length 222bp, with *IGH* at both the 5' and 3' ends of the insert sequences. Structures 2 and 3 are 'potential' inserts, lengths 209bp, 211bp and 487bp, with only 5' *IGH* identifiable in structure 2.

All insert-containing sequences were amplified using the Vh4-Fr1 BIOMED primer set and demonstrated the same region of chr6. BLAT of structure 1 3' sequence aligned to *IGHJ5* and the J4-J5 intron. Nucleotide Basic Local Alignment Tool (BLASTn) of the chr6 region did not demonstrate significant homology to IGH. Open reading frame analysis and VDJ usage is detailed in S3.

The Chr6 insert loci were characterised using UCSC genome browser(12), and shown to be an intergenic 'DNase hypersensitive site'.

In the switch region insertion 'hotspot' sites were defined as follows: Sγ1 region chr14:106210574-106213018 q32.33; Sμ region chr14:106323549-106326839 q32.33.

The bioinformatic pipeline confirmed 2 inserts, both in a 'non-class switched B-cells D6' control sample, Sμ primer set. The FASTA sequences are in S3.

Insert 1: length 194bp, mean bp coverage 91.4, identified as gene *ANKRD44* exon 2, Chr2q33.1. It is highly expressed in lymph node *(https://www.ncbi.nlm.nih.gov/gene/91526),* EBV transformed lymphocytes and whole blood(12).

Insert 2: length 250bp, mean coverage 51.9, intergenic, 500bp upstream of gene *ACTB* (a cytoskeleton actin protein that is expressed in B cells (13)). This aligns to CFS FRA7b position Chr7:1-7300000(14).

Like the Chr6 insert from VDJ gene sequencing, both switch region inserts include DNase hypersensitivity clusters.

Our results demonstrate that accurate targeting of the switch region using the consensus map and *in-silico* primer design combined with a long amplicon PCR method can enable sequence analysis of this crucial genomic region.

We also demonstrate that identification of insertion events from routine IGHV gene sequencing is possible and suggest that IGHV NGS could identify more insertions and improve sequencing depth for subclonal insert discovery(7). This could be coupled with longer PCR extension times and/or alternative polymerases to enhance longer insert amplification.

The finding of the same IGHV insert sequence from chr6 in 4 samples of 3 different diagnoses (CLL, MZL and HCL) suggest a shared insert pattern in cLPD patients. Inserts from this region were not identified in published datasets from normal donors(2, 3). IGHV and BCR data from online repositories may demonstrate inserts and would be helpful for developing NGS-based IGHV reporting guidance when it is updated by ERIC.

Inserts identified by this study map to 'DNase hypersensitive sites'(12) of accessible chromatin(15), consistent with other evidence linking insert origins to open chromatin regions(7). IGHV VDJ-insert-partial J-*IGH* as gDNA has not been described but the IGHJ loci are within the 'switch region' definition used by previous bioinformatic analysis. Inserts after VDJ and no 3' *IGH* sequence identified have also been published previously(3). Finding switch inserts in only unswitched cells may reflect that μ-μ and γ1- γ1 primer sets were used.

Insertional diversification events are identified as rare events in clonal IGHV sequences of cLPDs. Given the small numbers studied screening additional samples and NGS analysis would be interesting to evaluate insertion events alongside IGHV stereotypy and mutational status.

Acknowledgements:
HeLa cells were provided by Nicole McDermott. Control cells were prepared by Michelle Campbell. This work was undertaken on ARC3, part of the High-Performance Computing facilities at the University of Leeds, UK. Sequencing was performed at Leeds HMDS. Dr Paul Evans, Principal Clinical Scientist at Leeds HMDS provided the concept for investigating routine cLPD samples for IGHV inserts and has since retired. Dr Matthew Care, bioinformatician now at the University of York set author AF up on the ARC3 High Performance Computing system and provided initial advice for its use.

Data availability statement:
Data is available in the supplementary files and upon request from the corresponding author.

Funding sources:
The study was supported by an Early-Stage Research Start-Up Grant from the British Society of Haematology and NIHR Academic Clinical Fellowship awarded to author AF.

RT and DJN are supported in part by the National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre (BRC) (NIHR203331). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Conflict of interest disclosure:
All authors declare no competing interests.

Ethics approval statement: Use of anonymised samples and sequencing data is granted under REC 04/Q1205/125. Research sampling and analysis is approved under REC 07/Q1206/47.

Patient consent statement:
N/a.

Permission to reproduce material from other sources:
N/a.

Clinical trial registration:
N/a.

Author Contributions:
AF designed the switch region map and defined the 'hotspot' region. Performed all laboratory and bioinformatic methods and wrote the paper; JS produced IGVH sequencing reports in HMDS and revised the paper; DN developed the original long amplicon PCR methodology, advised on NGS library preparation and revised the paper; RT identified the research question, advised on sample selection and result interpretation, and revised the paper.

Supplementary files

1) Extended methods
2) Switch region map
3) Extended results

References

1. Tan J, Pieper K, Piccoli L, Abdi A, Perez MF, Geiger R, et al. A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. Nature. 2016;529(7584):105-9.
2. Pieper K, Tan J, Piccoli L, Foglierini M, Barbieri S, Chen Y, et al. Public antibodies to malaria antigens generated by two LAIR1 insertion modalities. Nature. 2017;548(7669):597-601.
3. Lebedin M, Foglierini M, Khorkova S, Vazquez Garcia C, Ratswohl C, Davydov AN, et al. Different classes of genomic inserts contribute to human antibody diversity. Proc Natl Acad Sci U S A. 2022;119(36):e2205470119.

4. Rosen A, Murray F, Evaldsson C, Rosenquist R. Antigens in chronic lymphocytic leukemia--implications for cell origin and leukemogenesis. Semin Cancer Biol. 2010;20(6):400-9.
5. Rosenquist R, Ghia P, Hadzidimitriou A, Sutton LA, Agathangelidis A, Baliakas P, et al. Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations. Leukemia. 2017;31(7):1477-81.
6. Munir T PA, Bloor A, Pettitt A, Patten P, Forconi F, et al. . Combination of Ibrutinib Plus Venetoclax with MRD-Driven Duration of Treatment Results in a Higher Rate of MRD Negativity in IGHV Unmutated Than Mutated CLL: Updated Interim Analysis of FLAIR Study. 64th American Society of Haematology Annual Meeting and Exposition 10/12/2022; United States2022.
7. Dale GA, Wilkins DJ, Rowley J, Scharer CD, Tipton CM, Hom J, et al. Somatic Diversification of Rearranged Antibody Gene Segments by Intra- and Interchromosomal Templated Mutagenesis. J Immunol. 2022;208(9):2141-53.
8. Lebedin M, de la Rosa K. Diversification of Antibodies: From V(D)J Recombination to Somatic Exon Shuffling. Annu Rev Cell Dev Biol. 2024;40(1):265-81.
9. Matthews C, Catherwood M, Morris TC, Alexander HD. Routine analysis of IgVH mutational status in CLL patients using BIOMED-2 standardized primers and protocols. Leuk Lymphoma. 2004;45(9):1899-904.
10. Giudicelli V, Brochet X, Lefranc MP. IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. Cold Spring Harb Protoc. 2011;2011(6):695-715.
11. Visco C, Moretta F, Falisi E, Facco M, Maura F, Novella E, et al. Double productive immunoglobulin sequence rearrangements in patients with chronic lymphocytic leukemia. Am J Hematol. 2013;88(4):277-82.
12. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002;12(6):996-1006.
13. Gu Y, Tang S, Wang Z, Cai L, Lian H, Shen Y, Zhou Y. A pan-cancer analysis of the prognostic and immunological role of beta-actin (ACTB) in human cancers. Bioengineered. 2021;12(1):6166-85.
14. Kumar R, Nagpal G, Kumar V, Usmani SS, Agrawal P, Raghava GPS. HumCFS: a database of fragile sites in human chromosomes. BMC Genomics. 2019;19(Suppl 9):985.
15. Chen Y, Chen A. Unveiling the gene regulatory landscape in diseases through the identification of DNase I-hypersensitive sites. Biomed Rep. 2019;11(3):87-97.